

5/04

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole Nationale Polytechnique

DEPARTEMENT : GENIE ELECTRIQUE

SPECIALITE : ELECTROTECHNIQUE



المدرسة الوطنية المتعددة التخصصات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique

Mémoire de fin d'études
En vue de l'obtention du diplôme
D'Ingénieur d'Etat

Thème :

**PREDICTION DES CARACTERISTIQUES
D'UN ISOLATEUR POLLUE PAR RESEAUX
DE NEURONES ARTIFICIELS**

Proposé et dirigé par :

M^r A.Mekhaldi

Etudié par :

M^r Benlalam Hicham

Promotion JUIN 2004

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole Nationale Polytechnique

DEPARTEMENT : GENIE ELECTRIQUE
SPECIALITE : ELECTROTECHNIQUE



المدرسة الوطنية المتعددة التقنيات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique

Mémoire de fin d'études
En vue de l'obtention du diplôme
D'Ingénieur d'Etat

Thème :

PREDICTION DES CARACTERISTIQUES
D'UN ISOLATEUR POLLUE PAR RESEAUX
DE NEURONES ARTIFICIELS

Proposé et dirigé par :

M^r A. Mekhaldi

Etudié par :

M^r Benlalam Hicham

Promotion JUIN 2004

ملخص :

في هذا العمل قمنا بدراسة إمكانية استعمال الشبكات العصبية الاصطناعية من أجل تعميم الدوال غير الخطية التي تتميز انتشار التفريغات الكهربائية في المساحات العازلة. و بالتالي انطلاقا من منحنيات تطبيقية نتحقق من فعالية الشبكات المقترحة. هناك ثلاث تشكيلات مقترحة ، ندرج كلها في نفس النوع من الشبكات العصبية الاصطناعية، يتعلق الأمر بالشبكة بدالة ذات الأساس القطري. من أجل إنجاز التعليم، استعملنا طريقة الاقتصار العشوائية. بدأنا بتعيين الأحسن من بين الطرق الثلاثة و هذا بتطبيقها على نفس المنحنيات. النتائج التي تحصلنا عليها تبين بأن الطريقة الثالثة (شبكة ذات مدخلين) هي الأكثر فعالية إذا تعلق الأمر بتعميم الدوال غير الخطية.

الكلمات المفتاحية :

تلوث العوازل، الشبكة بدالة ذات الأساس القطري، طريقة الاقتصار العشوائية

Résumé :

Dans ce travail nous avons illustré la possibilité d'utiliser les réseaux de neurones artificiels (RNA) dans l'extrapolation des fonctions non linéaires caractérisant la propagation des décharges électriques sur des surfaces isolantes polluées. En effet, à partir des courbes expérimentales nous testons l'efficacité des réseaux que nous allons élaborer. Trois configurations sont à utiliser, les trois entrent dans le même type des RNA, il s'agit du réseau à fonction de base radiale (RBF). Pour effectuer l'apprentissage, nous avons utilisé la méthode d'optimisation aléatoire (ROM). Nous commençons par déterminer la meilleure des trois méthodes en les appliquant aux mêmes caractéristiques. Les résultats que nous obtenons montrent que la troisième méthode (RBF à deux neurones d'entrée) est la plus efficace dans l'extrapolation des fonctions non linéaires.

Mots clés :

Pollution des isolateurs, réseaux de neurones, fonction de base radiale, optimisation aléatoire.

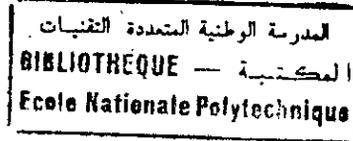
Abstract :

The purpose of this work is to illustrate the possibility to use the artificial neural networks (RNA) in the non linear functions extrapolation characterizing the electric discharges propagation on the polluted insulating surfaces. From the applied curves we test the efficiency of networks that we go elaborate indeed. Three configurations are to use, the three enter in the same type of the RNA, it is about the network to radial basis function (RBF). To do the training we used the random optimization method (ROM). We start with determining best of the three methods while applying them to the same features. Results that we get show that the third method (RBF to two neurons of entry) is most efficient in the non linear function extrapolation.

Key words:

Pollution of insulators, Neural networks, radial basis function, random optimization method

Dédicaces



Je profite de cette occasion pour témoigner à tous ceux que j'aime et que j'estime, ma profonde gratitude et la sincérité des profonds sentiments que je leurs éprouve.

Je dédie ce modeste travail tout particulièrement à ceux qui comptent le plus pour moi, ma mère et mon père, qui ont toujours été présent à mes cotés et qui m'ont toujours soutenu et encouragé dans chacune de mes entreprises. Que dieu vous garde.

A mes frères, qui ont toujours été présent quand j'avais besoin d'eux et qui m'ont beaucoup soutenu et aidé dans les moments les plus difficile de ma formation. Je leurs souhaite tout le bonheur du monde.

Au petit dernier de la famille, mon neveu Yanis. Que dieu le protège et le guide dans chacun de ses pas. Puisse t-il remplir de joies et de bonheur ses parents, ses grands-parents et toute sa famille.

A toute ma famille. Oncles, tantes, cousins et cousines. Surtout à ceux qui sont loin de nous. A ceux qui ne sont plus des nôtres, tout particulièrement mes deux grands-pères, mes deux grands-mères, Yama et Mima, mes deux tantes Zohra et Chadlia, et mon cousin Mohamed. Je leurs dis, vous êtes et vous serez toujours présent dans nos pensées et nos cœurs. Puisse dieu le tout-puissant leurs accorder sa sainte miséricorde et les accueillir dans son vaste paradis.

A tous mes amis et amies. A notre amitié, qui pour certains, malgré la distance et le temps, est restée toujours très forte. A tous ces moments de bonheur, de joies et de tristesse que nous avons passé ensemble, qui n'ont fait que nous rapprocher et renforcer notre amitié. Pour que vous restiez toujours dans mes pensées et pour que nous soyons unis à jamais.

Remerciements

Je tiens tout d'abord à témoigner ma profonde reconnaissance à mon promoteur M^r A.Mekhaldi, pour avoir bien voulu diriger ce travail. Je le remercie vivement pour tout le soutien et les facilités qu'il m'a apporté autant qu'enseignant, chef de département et encadreur.

Je voudrais exprimer toute ma gratitude et mes remerciements à tous ceux qui m'ont permis de près ou de loin de parvenir à achever ce travail. A M^{lle} L.Mokhnache, Mr A.El Hadj et M^r M.Teguar pour leurs précieux conseils et directives. A tous mes amis qui m'ont aidé et soutenus au court de la réalisation de ce travail.

J'exprime encore une fois, ma gratitude ainsi que ma reconnaissance à l'ensemble des enseignants qui ont contribué, directement ou indirectement, à notre formation et plus précisément à ceux du département Génie Electrique.

Je ne saurais terminer sans remercier vivement le président et les membres du jury pour avoir examiné ce travail et de m'avoir honorés par leur présence.

Sommaire

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

Introduction générale.....	1
Chapitre I : Introduction aux réseaux de neurones artificiels	
I.1. Introduction.....	3
I.2. Historique.....	3
I.3. Le neurone biologique.....	6
I.3.1. Structure des neurones biologiques.....	6
I.3.1.1. Le soma.....	7
I.3.1.2. Les dendrites.....	7
I.3.1.3. L'axone.....	7
I.3.2. Fonctionnement du neurone biologique.....	7
I.4. Les neurones artificiels.....	8
I.4.1. Le neurone formel.....	9
I.4.2. L'activation du neurone.....	9
I.4.2.1. Fonction discriminant.....	9
I.4.2.1.a. Fonction de base radiale (RBF).....	9
I.4.2.1.b. Fonction à base elliptique (EBF).....	10
I.4.2.2. Fonction dynamique linéaire.....	10
I.4.2.3. Fonction non dynamique non linéaire.....	10
I.5. Architecture des réseaux de neurones.....	11
I.5.1. Réseaux statiques.....	11
I.5.2. Réseaux dynamiques.....	11
I.5.3. Réseaux auto-organisés.....	11
I.6. Apprentissage des réseaux de neurones.....	11
I.6.1. Apprentissage supervisé.....	11
I.6.2. Apprentissage non supervisé.....	12
I.6.3. Le renforcement.....	12
I.6.4. Le mode hybride.....	12
I.7. Réseaux de neurones à apprentissage supervisé.....	12
I.7.1. Réseaux de neurones basés sur la décision (DBNN).....	12
I.7.1.1. Le perceptron.....	12
I.7.1.2. Adaline, Madaline.....	14
I.7.2. Réseaux multicouches statique.....	15
I.7.2.1. Réseaux à fonction de base linéaire (LBF).....	15
I.7.3. Réseaux à fonction de base radiale (RBF).....	16
I.7.3.1. Architecture et fonctionnement des réseaux RBF.....	17
I.7.3.2. Apprentissage des réseaux RBF.....	18
I.8. Algorithmes d'apprentissage des RNA à apprentissage supervisé.....	19
I.8.1. Méthode des moindres carrés (LMSE).....	19
I.8.2. Méthode de descente de gradient.....	19
I.8.3. Rétro-propagation du gradient (backpropagation).....	20
I.8.3.1. Variantes de la backpropagation.....	22
I.8.3.1.1. Backpropagation à taux d'apprentissage adaptatif.....	22
I.8.3.1.2. Backpropagation avec momentum.....	23
I.8.3.1.3. Robust backpropagation.....	23
I.8.3.1.4. Fast backpropagation.....	23
I.8.4. Méthode d'optimisation aléatoire (ROM).....	24

I.9. Réseaux de neurones à apprentissage non supervisé	25
I.9.1. Mémoires associatives.....	25
I.9.1.1. Mémoires auto-associatives.....	26
I.9.1.2. Mémoires hétéro-associatives.....	26
I.9.2. Réseaux compétitifs.....	26
I.9.3. Cartes auto-organisatrices.....	27
I.9.4. Modèle de résonance adaptative (ART).....	28
I.10. Apprentissage par renforcement	30
I.11. Problèmes d'apprentissage	31
I.12. Conclusion	32

Chapitre II : Synthèse des travaux effectués dans le domaine de la HT utilisant les RNA

Synthèse des travaux effectués dans le domaine de la HT utilisant les RNA.....	34
--	----

Chapitre III : Résultats et interprétations

III.1. Introduction	49
III.2. Architecture des réseaux	50
III.2.1. Méthode 1 : RBF avec entrée et sortie de nature différente.....	50
III.2.2. Méthode 2 : RBF avec entrée et sortie de même nature.....	50
III.2.3. Méthode 3 : RBF avec deux entrées.....	51
III.3. Résultats de la première méthode	51
III.3.1. Variation du courant de fuite en fonction du rayon de l'électrode circulaire.....	51
III.3.2. Variation du courant de fuite en fonction de la conductivité de la couche de pollution.....	54
III.3.3. Variation de la longueur d'arc électrique en fonction de la tension appliquée.....	56
III.3.4. Variation de la tension de contournement en fonction de la longueur de la couche de pollution.....	58
III.4. Résultats de la deuxième méthode	59
III.4.1. Variation du courant de fuite en fonction du rayon de l'électrode circulaire.....	59
III.4.2. Variation du courant de fuite en fonction de la conductivité de la couche de pollution.....	61
III.4.3. Variation de la longueur d'arc électrique en fonction de la tension appliquée.....	63
III.4.4. Variation de la tension de contournement en fonction de la longueur de la couche de pollution.....	65
III.5. Résultats de la troisième méthode	67
III.5.1. Variation du courant de fuite en fonction du rayon de l'électrode circulaire.....	67
III.5.2. Variation du courant de fuite en fonction de la conductivité de la couche de pollution.....	69
III.5.3. Variation de la longueur d'arc électrique en fonction de la tension appliquée.....	71

III.5.4. Variation de la tension de contournement en fonction de la longueur de la couche de pollution.....	73
III.6. Comparaison entre les trois méthodes pour la nouvelle approche.....	75
III.7. Conclusion.....	78
Conclusion Générale.....	79
Références bibliographiques.....	80

المدسة الوطنفة المفعدة الففنفاف
BIBLIOTHEQUE — المكنففة
Ecole Nationale Polytechnique

INTRODUCTION GENERALE

Dans le dernier chapitre, nous présentons les résultats obtenus, pour la prédiction de quatre caractéristiques décrivant le comportement d'un isolateur haute tension pollué, par deux réseaux différents, celui de A.El hadj et le nôtre.

Nous comparerons les résultats obtenus par notre réseau aux résultats expérimentaux ainsi qu'aux résultats A.El hadj afin de déterminer le meilleur réseau de prédiction.

Et enfin, nous terminons notre étude par une conclusion générale.



CHAPITRE I

INTRODUCTION AUX RESEAUX DE NEURONES ARTIFICIELS



1.1.Introduction :

Malgré la constante augmentation de puissance des calculateurs, malgré les approches théoriques de plus en plus sophistiquées, un certain nombre de tâches résistent encore aux algorithmes et aux méthodes classiques de traitement des signaux et des données. Ces tâches relèvent typiquement du traitement, en temps réel, de très grands flots de données souvent multidimensionnelles et arrivant à des cadences élevées. Le grand nombre des données, leur variabilité, le fait qu'elles ne répondent pas à des modèles physiques connus nous laissent souvent démunis devant des tâches de caractérisation, de reconnaissance et de prise de décision.

Or, toutes ces opérations se trouvent réalisées de manière naturelle chez les êtres vivants : un visage est reconnu en quelques dixièmes de seconde, une voix est reconnue au milieu du bruit ambiant et le discours est perçu, le contrôle simultané d'une centaine de muscles en fonction de la position et de la survenue d'un obstacle dans le champ visuel est réalisé en quelques dixièmes de secondes. Pour une machine, réaliser de telles tâches serait une véritable prouesse, et nous, nous les réalisons sans même y penser ! C'est là que réside l'attrait des neurones artificiels. Copier le cerveau restera pour longtemps encore une ambition exagérée, mais vouloir s'inspirer des architectures et des fonctions du système nerveux n'est pas un rêve inaccessible. Le développement continu des connaissances en biologie, l'apparition de nouvelles méthodes théoriques et l'incessante montée en puissance des outils de simulation nous autorisent les meilleurs espoirs.

Le concept de réseaux de neurones artificiels est donc inspiré des réseaux de neurones biologiques. Ces derniers, briques constitutives du cerveau, sont bien plus lents que des portes logiques ; pourtant, l'inférence des réseaux biologiques est bien plus rapide que ce que peut faire le plus rapide des calculateurs actuels. De fait, le cerveau compense sa faible vitesse de calcul par une énorme quantité de cellules interconnectées. Un réseau de neurones biologiques est un système non linéaire, à haut degré de parallélisme et très robuste aux pannes. De plus il peut :

1. apprendre en adaptant ses poids synaptiques en fonction de l'environnement,
2. utiliser des informations probabilistes, imprécises et "brutes",
3. généraliser à partir d'exemples (via des cas particuliers).

Les réseaux de neurones artificiels RNA (où Artificial Neural Network, ANN) sont une tentative pour simuler tout ou partie de ces caractéristiques. Cette approche diffère de la programmation classique, dans le sens où une information n'est pas stockée dans un endroit particulier, mais est répartie au niveau des différentes connexions du réseau. Les neurones ou cellules élémentaires sont alors considérés comme des processeurs élémentaires aptes à effectuer des opérations primitives telles que "sommer les entrées", "amplifier" ou "seuiller" cette valeur suivant les cas.

Il y a des centaines d'exemples nous montrant à la fois combien on peut espérer de la modélisation du système nerveux mais aussi combien il sera difficile d'imaginer et de comprendre les divers aspects des problèmes de perception. Il paraît donc naturel d'essayer de comprendre comment les systèmes biologiques sont capables de telles performances, et si possible, de s'inspirer de leurs principes pour imaginer de nouveaux algorithmes ou de nouvelles machines plus efficaces que ceux dont nous disposons actuellement. Les techniques de réseaux de neurones relèvent d'une telle approche : comprendre les principes selon lesquels

les systèmes biologiques traitent l'information et s'en inspirer pour élaborer de nouvelles techniques en sciences de l'ingénieur. C'est donc une double démarche, à la fois cognitive et synthétique où le monde biologique doit être considéré comme une source de référence et de connaissance. D'un point de vue technique, il est clair que seuls les principes seront importants.

Les caractéristiques essentielles des réseaux de neurones réels que nous conserverons dans les modèles mathématiques étudiés, concernent le grand nombre de connexions, la non-linéarité des relations entrée-sortie et la faculté de "plasticité" ou d' "adaptabilité". Ces caractéristiques, même simplifiées, leur confèrent déjà de multiples possibilités en traitement des signaux et des informations ainsi que la faculté d'apprendre à classer, à reconnaître des formes ou à réaliser des tâches complexes.

1.2. Historique :

L'apparition de l'ordinateur et le développement de nouvelles théories d'apprentissage et processus neural ont tous deux émergés dans la même période, vers la fin des années quarante. Depuis ce jour, l'ordinateur a été utilisé comme un instrument pour modéliser soit des neurones individuels où un ensemble de neurones, qui est connu sous le nom de réseau de neurones.

En vérité, il est intéressant de mentionner que les premières approches sont apparues bien avant. En 1890, W. James, psychologue américain, introduit le concept de mémoire associative et propose ce qui deviendra une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones connue plus tard sous le nom de la loi de Hebb [1].

Les prémices :

Dans la période de 1940 à 1956, les sciences et techniques de la cognition s'élaborèrent peu à peu et de l'ancienne cybernétique naquit les sciences cognitives. La volonté de fonder une science "naturelle" de l'esprit émergea quasiment au même moment à cette époque, en Europe et aux Etats-Unis. Par les efforts conjugués de chercheurs d'origines très diverses (neurosciences, intelligence artificielle, linguistique, psychologie et épistémologie principalement), un nouveau champ d'investigation pluridisciplinaire, que sont les sciences cognitives, vit donc le jour.

L'apparition des réseaux de neurones artificiels s'inscrit complètement dans cette genèse des sciences cognitives, et leur origine tient dans la volonté de modéliser, de façon mathématique, les neurones biologiques.

Les pionniers :

En 1943, J.McCulloch et W.Pitts [2] inventèrent le premier neurone formel, un produit scalaire entre un vecteur d'entée et un vecteur poids, suivi d'un élément à seuil répondant 0 ou 1 selon que le résultat du produit scalaire est supérieur ou non au seuil.

Quelques années plus tard, en 1949, c'est Hebb qui proposera une formulation du mécanisme d'apprentissage, sous la forme d'une règle de modification des connexions synaptiques (règle de Hebb [1]). Cette règle, basée sur des données biologiques, modélise le

fait que si des neurones, de part et d'autre d'une synapse (Fig.I.1), sont activés de façon synchrone et répétée, la force de la connexion synaptique va aller croissant.

Les premiers RNA :

C'est en 1958 qu'apparaît, proprement dit, le premier réseau de neurones artificiels, grâce aux travaux de F.Rosenblatt [3] qui conçoit le fameux Perceptron. Le Perceptron est inspiré du système visuel et possède une couche de neurones d'entrée "perceptive" ainsi qu'une couche de neurones de sortie "décisionnelle". Ce réseau parvient à apprendre à identifier des formes simples et à calculer certaines fonctions logiques. Il constitue donc le premier système artificiel présentant une faculté jusque là réservée aux êtres vivants : la capacité d'apprendre par l'expérience.

Malgré tout l'enthousiasme que soulève le travail de Rosenblatt, en 1969, Minsky et Papert [4] publient un ouvrage « Perceptrons » où ils démontrèrent toutes les limites de ce modèle et soulèvent particulièrement l'incapacité du Perceptron à une couche de neurones de séparer des données non linéairement séparables et que pour des opérations plus générales de classification, il fallait mettre plusieurs couches. Seulement là, on ne savait pas comment faire apprendre les bons vecteurs poids.

L'ombre :

Suite a ces limites mis au jour par Minsky et Papert, il s'en suivra alors, face à la déception, une période noire d'une quinzaine d'années dans le domaine des réseaux de neurones artificiels.

Mais les recherches ne se sont, bien sûr, pas arrêtés, au contraire des chercheurs comme Kohonen [5] et bien d'autres ont travaillé sur les prémices de la plasticité neural, apprenant progressivement à les formaliser en combinant les méthodes de l'algèbre linéaire et celle des techniques non linéaires. Il est vrai cependant que le brio des théories de Hopfield, épaulé par un engouement des physiciens à partir de 1982, a été le moteur d'un renouveau mondial sur les neurones artificiels.

Le renouveau :

Il faudra attendre le début des années 80 et le génie de Hopfield [6] pour que l'intérêt pour les RNA soit de nouveau présent. Hopfield démontre, en 1982, tout l'intérêt d'utiliser des réseaux récurrents, dits "feed-back". Les réseaux récurrents constituent alors la deuxième grande classe de réseaux de neurones, avec les réseaux type perceptron, dit "feed-forward".

En parallèle des travaux de Hopfield, Werbos [7] conçoit son algorithme de rétro-propagation de l'erreur, qui offre un mécanisme d'apprentissage, pour les réseaux multi-couches de type Perceptron (appelés MLP pour Multi-layer Perceptron), fournissant ainsi un moyen simple d'entraîner les neurones des couches cachées.

Cette découverte permet de réaliser une fonction non linéaire d'entrée/sortie sur un réseau en décomposant cette fonction en une suite d'étapes linéairements séparables.

Depuis la fin des années 80, l'intérêt pour les réseaux de neurones artificiels ne fait que croître et ceci dans de plus en plus de domaines.

1.3. Le neurone biologique :

L'élément fonctionnel de base du système nerveux est le neurone. Un neurone (cellule nerveuse) est une cellule biologiquement spécialisée pour le traitement et la transmission de l'information.

Les neurones biologiques sont caractérisés par cinq fonctions qui leur sont propre [8] :

1. recevoir des signaux en provenance de neurones voisins,
2. intégrer ces signaux,
3. engendrer un flux nerveux,
4. conduire le flux nerveux et le transmettre à un autre neurone capable de le recevoir.

1.3.1. Structures des neurones biologiques :

D'une espèce animale à une autre, les neurones présentent des différences notoires, et à l'intérieur d'une même espèce, il existe de nombreuses variétés de neurones. Les différences portent tant sur les aspects anatomiques que sur les aspects fonctionnels. Cependant, divers points communs subsistent, et sont à la base de l'archétype de la cellule nerveuse.

Le neurone est constitué principalement de trois parties qui, vis-à-vis des transferts d'information, ont un rôle fonctionnel bien défini : un corps cellulaire ou soma et de deux types de prolongements cellulaires plus ou moins ramifiés : les dendrites et l'axone (Fig.I.1).

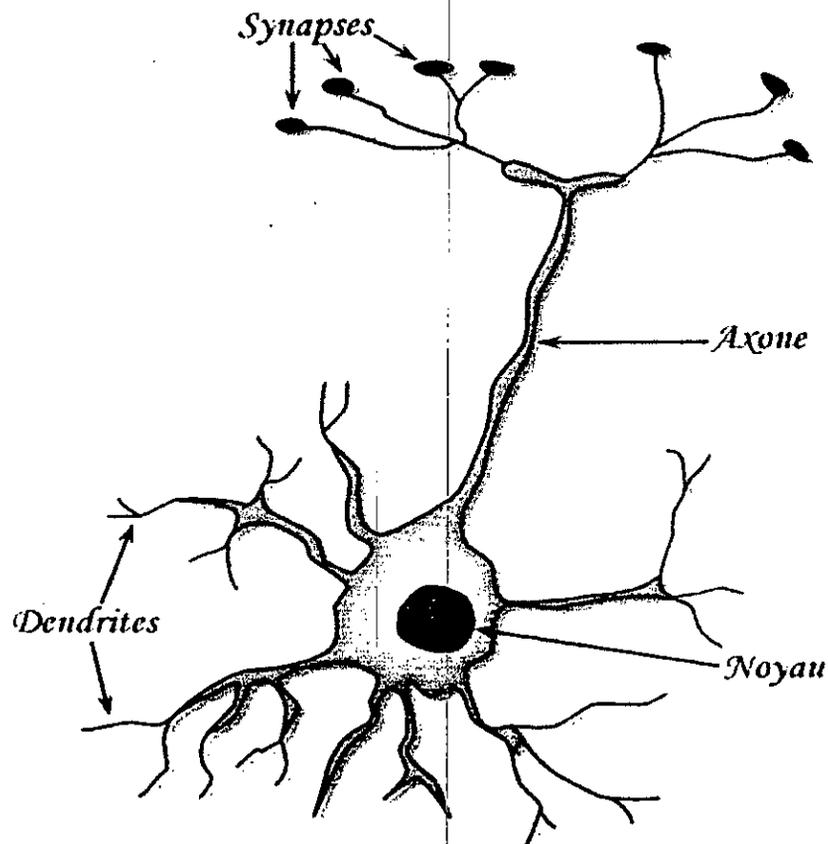


Fig.I.1. Le neurone biologique

1.3.1.1. Le soma :

Le soma possède un noyau qui contient l'information génétique de l'organisme dont le neurone est issu, et un *cytoplasme* qui contient la "machinerie de synthèse" des protéines constitutives ou régénératrices de ce neurone.

1.3.1.2. Les dendrites :

Le corps cellulaire se ramifie pour former ce que l'on nomme les dendrites. Celles-ci sont parfois si nombreuses que l'on parle alors de chevelure dendritique ou d'arborisation dendritique. Les dendrites reçoivent l'information au niveau de points de contact avec les autres neurones. Ces points de contact sont appelés synapses. L'information est ensuite acheminée vers le corps cellulaire ou soma.

1.3.1.3. L'axone :

Le neurone transmet l'information, générée dans son corps cellulaire, via l'axone. L'extrémité de l'axone peut être plus ou moins ramifiée en une arborisation terminale. Chaque branche de cette arborisation se termine en un bouton synaptique autour duquel se trouvent les synapses.

1.3.2. Fonctionnement du neurone biologique :

Le neurone reçoit des signaux (impulsions) provenant d'autres neurones au niveau des dendrites et transmet l'information via l'axone [9] (Fig.I.2).

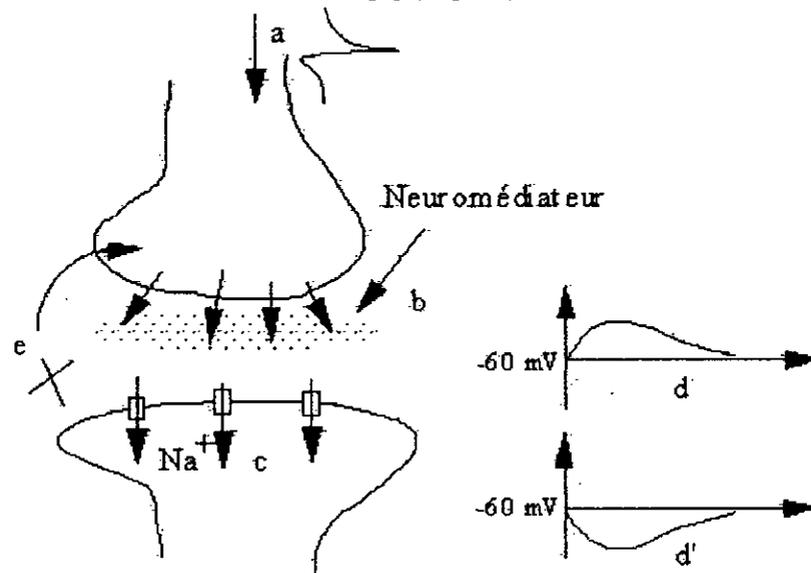


Fig.I.2. Fonctionnement au niveau synaptique

- a) Arrivée d'un potentiel d'action.
- b) Libération du neurotransmetteur dans l'espace synaptique.
- c) Ouvertures des canaux ioniques dues au neurotransmetteur.
- d) Génération d'un potentiel évoqué excitateur.
- d') Génération d'un potentiel évoqué inhibiteur.
- e) Fermeture des canaux, élimination ou capture du neurotransmetteur.

La jonction entre deux neurones est connue sous le nom de synapse. Une synapse est une structure élémentaire et une unité fonctionnelle de communication (axo-dendritique). Lorsque l'impulsion nerveuse (potentiel d'action) atteint la synapse, certaines substances chimiques (acides aminés, protéines), appelés neurotransmetteurs sont libérées dans l'espace synaptique séparant la membrane axonale (du premier neurone) de la membrane dendritique (du second neurone). Le neurotransmetteur traverse l'espace synaptique et se fixe sur des sites récepteurs post-synaptiques du neurone récepteur. Selon que la synapse est excitatrice ou inhibitrice, cette action du neurotransmetteur va potentialiser, ou au contraire, inhiber la capacité du neurone récepteur à émettre des potentiels d'action.

En fait, le fonctionnement synaptique va s'ajuster en fonction des signaux reçus, ce qui permet de dire que les synapses "apprennent" leur "comportement" à partir des activités dans lesquelles elles sont impliquées (ou à partir des signaux qui leur sont imposés ...).

On estime à environ 100 à 1 000 milliards le nombre de neurones du système nerveux humain, et on compte en moyenne de 1 000 à 100 000 synapses par neurone. Ces valeurs donnent une idée de la complexité du système nerveux. Les neurones sont donc fortement connectés entre eux. Ces connexions ne sont pas aléatoires, elles correspondent à des réseaux dont les architectures sont assez bien connues, mais dont les propriétés fonctionnelles restent encore souvent difficiles à comprendre.

1.4. Les neurones artificiels :

De nombreux travaux ont été effectués dans le cadre du développement de systèmes intelligents, dont certains inspirés par les réseaux de neurones biologiques. Des chercheurs de différents horizons ont défini et/ou construit des RNA pour résoudre des problèmes aussi variés que la reconnaissance de formes, la prévision, l'optimisation, les mémoires associatives ou le contrôle. Des approches classiques ont été proposées pour résoudre ces types de problèmes. Avec succès pour des cas où les contraintes sont clairement définies, cependant aucune solution n'est assez flexible pour s'adapter avec succès à un autre domaine.

L'évolution a offert au cerveau humain un nombre de caractéristiques; en particulier notre cerveau possède :

- une architecture massivement parallèle,
- un mode de calcul et une mémoire distribués,
- une capacité d'apprentissage,
- une capacité de généralisation,
- une capacité d'adaptation,
- une résistance aux pannes,
- une faible consommation énergétique.

Il est donc tentant d'étudier des systèmes basés sur les réseaux de neurones pour voir si certaines des caractéristiques de notre cerveau peuvent être "récupérées".

Les ordinateurs actuels dépassent les capacités humaines dans le calcul numérique et la manipulation de symboles. Cependant, les humains peuvent, sans effort, résoudre des problèmes complexes liés à la perception (reconnaissance d'un visage dans une foule) à une telle vitesse et avec une telle efficacité que n'importe quel ordinateur.

1.4.1. Le neurone formel :

Le premier neurone formel vit le jour en 1943 (Fig.I.3). Il a été modélisé par Mc.Culloch et W.Pitts en s'inspirant du neurone biologique.

Un neurone formel est une modélisation mathématique qui reprend les grands principes du fonctionnement du neurone biologique et particulièrement, la sommation des entrées. Sachant qu'au niveau biologique, les synapses n'ont pas toutes la même «valeur», les auteurs ont donc créé un algorithme qui pondère la somme de ses entrées par des poids synaptiques (coefficients de pondération).

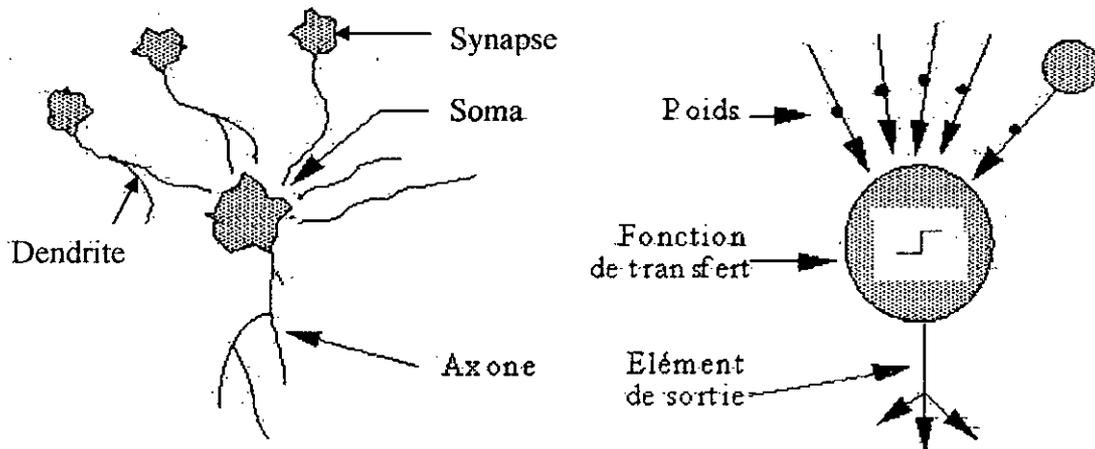


Fig.I.3. Mise en correspondance neurone biologique / neurone artificiel

1.4.2. L'activation du neurone :

Elle définit l'état interne du neurone. Elle est régit par différents types de fonctions d'activation. Les plus importantes sont citées ci-dessous.

1.4.2.1. Fonction discriminant :

Cette fonction, appelé aussi fonction de base, détermine l'activité du neurone. Dans le cas du perceptron, c'est une fonction linéaire qui est la plus utilisée.

Afin d'améliorer les performances des réseaux, d'autres fonctions discriminants non linéaires ont été élaborées. Nous citons ci-dessous les plus utilisées.

1.4.2.1.a. Fonction de base radiale (Radial Basis Function RBF) :

La forme générale de cette fonction est :

$$\phi = \frac{\|x - c\|^2}{2} \quad (I.1)$$

Où x représente les entrées et c les centres de chaque classe de l'espace d'entrée. Ce dernier sera échantillonné par cette fonction.

1.4.2.1.b. Fonction à base elliptique (Elliptic Basis Function EBF) :

Cette fonction est la généralisation de la fonction de base radiale :

$$\Phi = \sum_{k=1}^n \zeta_k (x - c)^2 + \theta \quad (\text{I.2})$$

Elle est modulée par des paramètres ζ_k et dotée d'un biais θ . Le paramètre n représente le nombre d'entrées.

1.4.2.2. Fonction dynamique linéaire :

La fonction dynamique linéaire détermine la dynamique du réseau. Elle est régie par l'équation différentielle suivante :

$$a_0 u_i(t) + a_1 u_i(t) = x_i(t) \quad (\text{I.3})$$

Où :

u_i est l'activation du $i^{\text{ème}}$ neurone,

x_i est l'entrée du système dynamique décrit par l'équation ci-dessus (I.3),

et a_i sont des paramètres.

1.4.2.3. Fonction non dynamique non linéaire :

Elle a pour rôle de borner l'activité du neurone. Le premier modèle proposé est la fonction seuil qui est une fonction binaire. L'inconvénient de cette fonction est sa non dérivabilité ce qui ne lui permet pas de faire certaines opérations, comme la différentiation qui est nécessaire pour améliorer les paramètres internes du réseau.

A cet effet, d'autres fonctions dérivables, monotones et continues pouvant effectuer le seuillage ont été mises au point. Parmi lesquelles on retrouve la sigmoïde, la tangente hyperbolique ou la log-sigmoïde.

Une autre fonction non monotone est utilisée, c'est la fonction gaussienne dont son grand avantage est son comportement local (Fig.I.4).

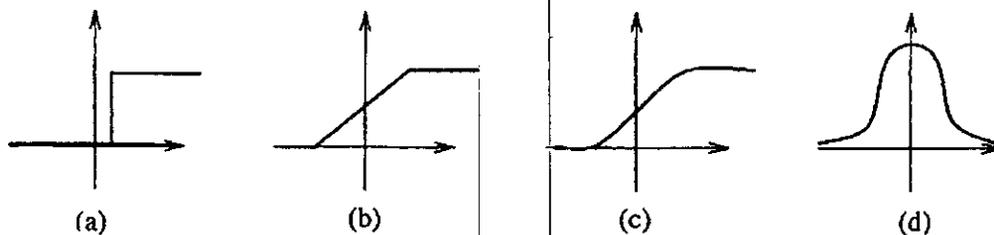


Fig.I.4. Différents types de fonctions d'activation

a : fonction à seuil , b : linéaire par morceaux, c : sigmoïde, d : fonction gaussienne

1.5. Architecture des réseaux de neurones :

On distingue trois types d'architecture.

1.5.1. Réseaux statiques :

Dans ce type d'architecture, chaque neurone reçoit ses entrées à partir de la couche précédente sans qu'il n'y ait retour d'information. Ce qui veut dire que le neurone ne communique pas avec la couche précédente, il en reçoit seulement l'information.

Ce genre de réseaux peut être utilisé pour la classification et l'approximation des fonctions.

1.5.2. Réseaux dynamiques :

Contrairement aux réseaux statiques, les réseaux dynamiques contiennent le retour d'information, ce qui leur permet de commander et d'identifier des systèmes dynamiques.

1.5.3. Réseaux auto-organisés :

Se sont des réseaux qui changent leurs structures internes pendant l'apprentissage. Les neurones se regroupent topologiquement suivant la représentation des exemples.

1.6. Apprentissage des réseaux de neurones :

L'apprentissage est vraisemblablement la propriété la plus intéressante des réseaux neuronaux.

Par définition, l'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré. L'apprentissage neuronal fait appel à des exemples de comportement

Dans le cas des réseaux de neurones artificiels, l'apprentissage se traduit par la modification des poids du réseau dans l'optique d'accorder la réponse du réseau aux exemples et à l'expérience. A l'issue de l'apprentissage, les poids sont fixés : c'est alors la phase d'utilisation.

On peut considérer qu'il existe quatre sortes d'apprentissage :

1. le mode supervisé,
2. le mode non-supervisé (ou auto-organisationnel),
3. le renforcement (certains auteurs le classent dans la catégorie des modes supervisés),
4. le mode hybride.

Pour chaque catégorie il existe de très nombreux algorithmes.

1.6.1. Apprentissage supervisé :

Dans le cas de l'apprentissage supervisé, l'adaptation intervient directement lorsque le système compare la réponse qu'il a calculée avec la réponse attendue (ou désirée) en fonction des entrées fournies. La performance de l'apprentissage est déterminée par l'intermédiaire d'un critère à optimiser.

1.6.2. Apprentissage non-supervisé :

Dans cet apprentissage, le réseau requiert seulement les entrées. Il modifie ses poids selon un critère interne.

1.6.3. Le renforcement :

Le renforcement est en fait une sorte d'apprentissage supervisé et certains auteurs le classent d'ailleurs, dans la catégorie des modes supervisés.

L'approche par renforcement essaye d'apprendre la corrélation entrées-sorties via une estimation de son erreur, c'est-à-dire du rapport échec/succès, le but étant de maximiser un index de performance (fourni) appelé signal de renforcement. Le système est capable de savoir si la réponse est correcte ou non, mais sans connaître la bonne réponse.

1.6.4. Le mode hybride :

Le mode hybride reprend en fait les deux autres approches, puisque une partie des poids va être déterminée par apprentissage supervisé et l'autre partie par apprentissage non-supervisé.

Dans ce travail on discutera uniquement du mode supervisé et non-supervisé, le renforcement étant considéré comme un mode supervisé.

1.7. Réseaux de neurones à apprentissage supervisé :

Le but de cet apprentissage est de trouver un vecteur de poids synaptiques optimal afin de ramener le réseau vers le comportement désiré imposé par un superviseur à l'aide d'échantillons aléatoires.

Il existe deux approches. La première est basée sur la validité de la décision du réseau, ce type de réseaux est appelé réseau basé sur les décisions (Decision Basis Neural Network ou DBNN). La seconde est basée sur l'optimisation d'un critère qu'il faut optimiser, appelé fonction coût.

1.7.1. Réseaux de neurones basés sur la décision :

1.7.1.1. Le perceptron :

Conçu en 1958 par Rosenblatt [3], c'est un réseau de neurones statique très simple avec une seule couche d'entrée reliée à un ou plusieurs neurones de sortie, acceptant seulement des valeurs d'entrées et de sorties binaires (Fig.1.5). L'information circule dans un seul sens, de la couche d'entrée vers la couche de sortie (feedforward).

Le réseau est capable de résoudre des opérations logiques simples comme AND ou OR.

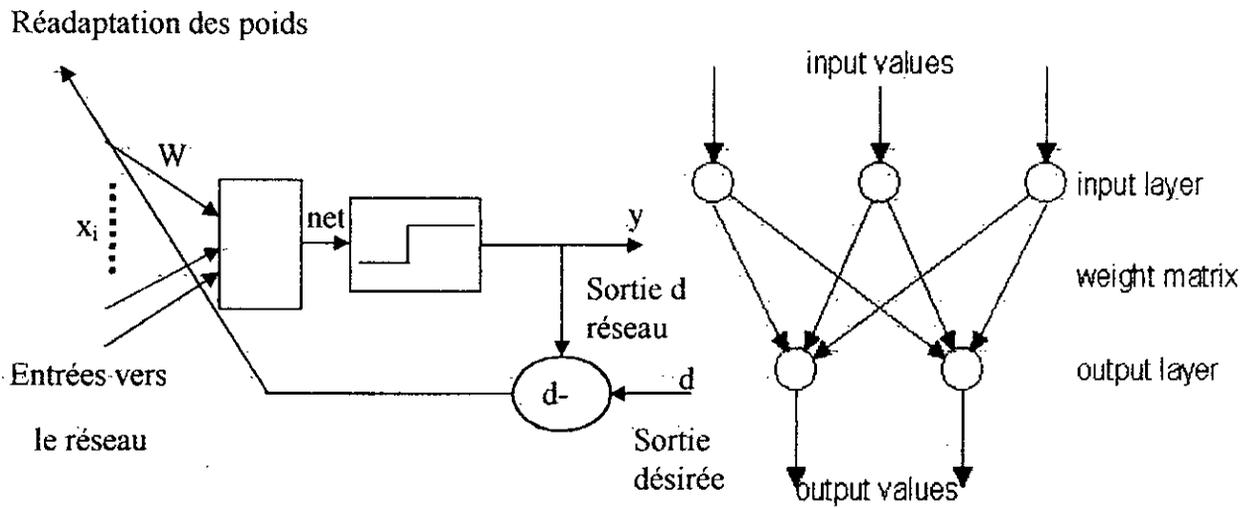


Fig.I.5. Structure du perceptron linéaire

L'activité des neurones dans ce type de réseaux est définie par l'équation linéaire suivante :

$$net_i = \sum_{j=1}^n x_j w_{ij} + \theta_i \quad i=1 \dots n \tag{I.4}$$

x_j représentent les entrées des neurones

w_{ij} le poids synaptique en provenance du $j^{ème}$ neurone de la couche d'entrée vers le $i^{ème}$ neurone de sortie.

θ_i est un paramètre qui agit sur le seuil en sortie.

La sortie est définie par la fonction seuil :

$$y = S_i(net_i) = \begin{cases} 1 & net_i > 0 \\ 0 & net_i \leq 0 \end{cases} \tag{I.5}$$

L'adaptation des poids de ces réseaux se fait comme suit :

$$w(t+1) = w(t) + \Delta w(t) \tag{I.6}$$

Tel que : $\Delta w(t) = \eta(d(t) - y(t)).x(t)$ (I.7)

Où :

η : est un paramètre qui contrôle le taux d'apprentissage.

$x(t)$: représente le vecteur présenté à l'entrée du réseau.

$y(t)$: est la sortie du réseau.

$d(t)$: est la sortie désirée.

L'entraînement du réseau se poursuit jusqu'à ce que chaque exemple soit correctement classé. Mais en 1969, Papert et Minsky [4] démontrent les limites du perceptron. Ce dernier n'est efficace que pour des problèmes linéairement séparables.

Une amélioration de ce perceptron à une ou plusieurs couches de neurones cachés a donné naissance au perceptron multicouches.

1.7.1.2. Adaline, Madaline :

En 1960 B.Widrow [10], un automaticien, développe le modèle Adaline (Adaptative Linear Element). Dans sa structure, le modèle ressemble au Perceptron, une couche d'entrée connectée à une ou plusieurs sorties (Fig.1.6). Dans le cas de plusieurs sorties, il est alors appelé Madaline (Multiple Adaptive Linear Element). Ce type de réseau est destiné à l'approximation des fonctions.

Cependant la loi d'apprentissage est différente. Celle-ci est à l'origine de l'algorithme de rétropropagation de gradient très utilisé aujourd'hui avec les perceptrons multicouches.

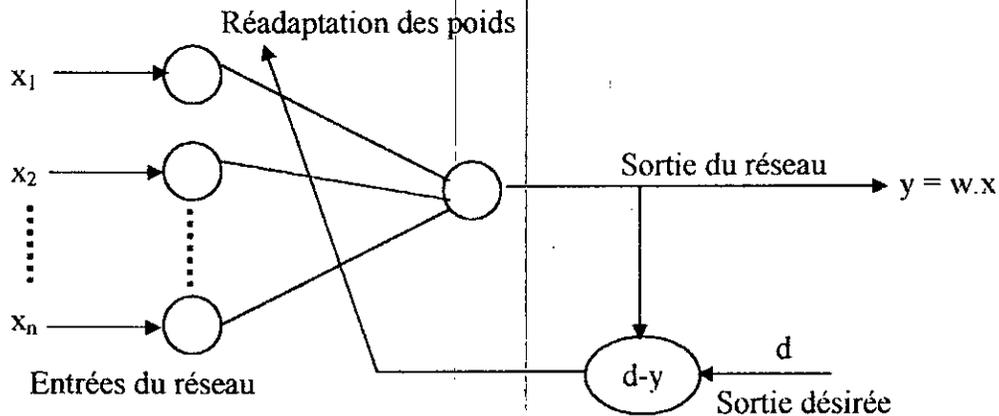


Fig.1.6. Structure du réseau ADALINE

L'apprentissage du Madaline s'effectue grâce à la minimisation d'une fonction coût calculée à partir des exemples de la base d'apprentissage et de la sortie du réseau de neurones. Cette fonction détermine l'objectif à atteindre.

A la présentation d'une entrée x_i^k , la sortie y_i^k est le produit scalaire de cette entrée avec un vecteur poids w_{ij} (Fig.1.6) :

$$y_i^k = \sum_{j=1}^n x_j w_{ij} + \theta_j \quad i=1, \dots, m \quad (1.8)$$

x_j : l'entrée du $j^{\text{ème}}$ neurone.

w_{ij} : poids synaptique du $i^{\text{ème}}$ neurone de la couche de sortie provenant du $j^{\text{ème}}$ neurone de la couche d'entrée.

θ_j : est le biais de la $j^{\text{ème}}$ sortie.

On définit l'erreur instantanée e_j^k relative à chaque sortie par la différence entre la valeur de la sortie du réseau calculée y_j^k et la sortie désirée d_j^k par :

$$e_j^k = d_j^k - y_j^k \quad (1.9)$$

Le problème réside dans l'ajustement des poids w_{ij} de telle sorte que l'écart entre la sortie désirée et la sortie calculée par le réseau soit nul pour tous les exemples. ($k = 1, \dots, n_{ex}$ avec n_{ex} le nombre d'exemples).

1.7.2. Réseaux multicouches statiques :

Un réseau multicouche est une extension du perceptron qui a une ou plusieurs sous-couches entre la couche d'entrée et la couche de sortie. Ces couches sont complètement connectées (plus exactement, une couche au niveau i , reçoit en entrée les signaux émanant de toutes les cellules de la couche $i - 1$, et propage l'information vers (toutes) les cellules de la couche $i + 1$) mais il n'existe aucune connexion intra-couche (Fig.I.7).

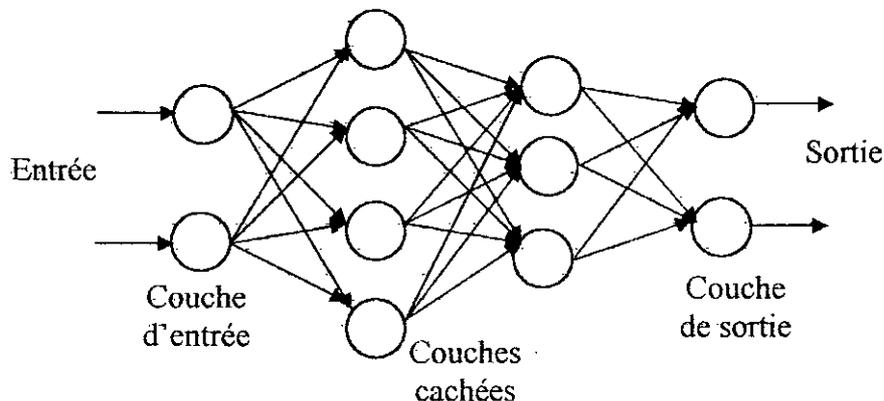


Fig.I.7. architecture d'un réseau multicouche statique

Ce type de réseau a des aptitudes à la généralisation, c'est-à-dire qu'il peut produire une réponse correcte (ou du moins presque correcte) pour une entrée qui ne lui a pas été présentée pendant la phase d'apprentissage. Le réseau peut être assimilé à une fonction d'entrée-sortie non linéaire, l'apprentissage ayant pour but de mettre au point cette fonction à partir d'un ensemble (judicieusement) choisi d'exemples.

Les réseaux multicouches ont été utilisés dans de nombreuses applications, allant du classificateur, des approximations de fonctions, jusqu'à l'identification de la parole, l'analyse de scènes ou l'identification de cibles militaires.

1.7.2.1. Réseaux à fonction de base linéaire (Linear Basis Function LBF) :

Ce réseau est basé sur l'algorithme de back-propagation, il est régi par l'équation :

$$u_i^l(t) = \sum_{j=1}^{N_{l-1}} w_{ij}^l(t) y_j^{l-1}(t) + \theta_i^l(t) \quad (1.10)$$

$$y_i^l(t) = f(u_i^l(t)) \quad (1.11)$$

Tel que :

l : le nombre de couches.

$u_i'(t)$: représente l'activité du $j^{\text{ème}}$ neurone dans la $l^{\text{ème}}$ couche.

$y_i'(t)$: la sortie du $i^{\text{ème}}$ neurone dans la $l^{\text{ème}}$ couche, en provenance du $j^{\text{ème}}$ neurone de la couche précédente.

$\theta_i'(t)$: est un paramètre appelé Biais. Il a pour objectif de donner plus de souplesse à la fonction de sortie du neurone en approximation de fonctions. Il est différent pour chaque neurone.

w_{ij} : sont les poids synaptiques de la $l^{\text{ème}}$ couche reliant les $i^{\text{ème}}$ et $j^{\text{ème}}$ neurones.

La fonction d'activation $f(u)$ est monotone non décroissante. Elle a pour rôle de limiter l'activité du neurone. Pour les neurones de la couche d'entrée, cette fonction n'est autre que l'identité. Ainsi dans la couche d'entrée, chaque neurone fait passer une composante du vecteur entrée vers sa sortie sans aucun traitement, c'est à dire $y_i'(t) = x_i(t)$ où $x_i(t)$ représente la $i^{\text{ème}}$ entrée du réseau à l'instant t .

1.7.3. Réseaux à fonction de base radiale (Radial Basis Function RBF) :

Ce sont des réseaux à une seule couche cachée de n unités qui ne réagissent significativement qu'à une partie restreinte dans l'espace d'entrée suivant une fonction d'activation de type gaussien.

Les RBF sont utilisés pour la classification et l'approximation de fonctions, et sont capables d'approximer avec un degré de précision quelconque n'importe quelle fonction continue non linéaire à travers une combinaison linéaire de fonctions non linéaires à base radiale. Le fonctionnement de ces réseaux repose sur le principe des estimateurs à noyau, et la généralisation se fait du monovarié vers le multivarié [9].

Un estimateur à noyau considère des fonctions de R^+ vers R et sont de la forme :

$$\phi_i(x) = \phi(\|x - \xi_i\|) \quad (\text{I.12})$$

Où

x : représente les entrées du réseau.

ξ_i : représente le nœud d'interpolation pour $i = 1, \dots, n$.

L'idée principale a été introduite par M.J.D.Powell [11], toute fonction $f(x)$ d'une variable x de R^n peut être approchée par une interpolation composée par la somme de p fonctions noyaux de forme fixée $\phi(\|x - \xi_i\|)$.

$$f(x) = \sum_{i=1}^p \lambda_j \phi(\|x - \xi_i\|) \quad (\text{I.13})$$

λ_j sont les paramètres que l'on déterminera à partir des exemples connus x_i, y_i en résolvant le système :

$$f(x) = y_j \quad i = 1, n \quad (\text{I.14})$$

$\phi(\|x - \xi_i\|)$ est une fonction assurant la continuité aux nœuds et la dérivabilité d'ordre supérieur en ces points.

1.7.3.1. Architecture et fonctionnement des réseaux RBF :

Le fonctionnement des réseaux RBF est basé sur l'idée des estimateurs à noyaux. En considérant le cas multivariable, la sortie délivrée par le réseau est de la forme :

$$f(x) = \sum_{i=1}^k w_j \Phi(\|x - c_i\|) \quad j=1, m \tag{I.15}$$

c_i sont appelés centres des classes. Se sont des valeurs que nous attribuons à chaque neurone de la couche cachée,

w_j sont les poids synaptiques reliant ces neurones aux sorties,

m est le nombre de sorties du réseau,

k est le nombre de centres.

La fonction noyau ϕ , qui doit toujours assurer la dérivabilité aux points de jonction, est radialement symétrique. Elle doit produire des réponses localisées dont les valeurs ne sont significatives que dans un certain intervalle de l'espace des entrées.

Parmi les fonctions noyaux qui sont utilisées dans les réseaux RBF, on peut trouver les formes suivantes :

Formes cubiques : $\phi(r) = r^3$

Formes multi-quadratiques : $\phi(r) = (r^2 + k^2)^{-1/2}$

Formes gaussiennes : $\phi(r) = \exp\left(-r^2/\beta^2\right)$

Où r est un réel quelconque, remplaçant le discriminant des neurones.

β et k sont des constantes.

Parmi ces fonctions, la fonction gaussienne est la plus utilisée.

La figure.I.8 représente un réseau RBF à trois couches. La première couche reçoit les entrées injectées au réseau, la couche cachée est constituée de neurones, dont le discriminant est à fonction de base radiale. Chacun de ces neurones est doté d'un vecteur c_i appelé centre, les valeurs de ces vecteurs représentent chacune le barycentre de la classe qui lui correspond.

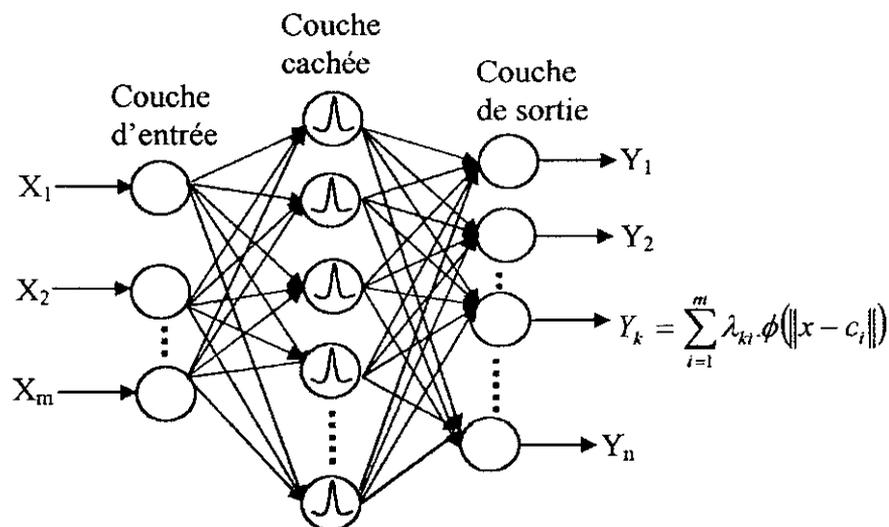


Fig.I.8. Architecture d'un réseau RBF

Lorsqu'un vecteur X est présenté à l'entrée du réseau, chaque neurone calcule, à travers son discriminant, la distance euclidienne d_i de cet exemple par rapport aux centres de la classe qui le représente :

$$d_i = \|X - c_i\| \quad (I.16)$$

La sortie de chaque neurone de cette couche est donnée comme suit :

$$s_i = \exp\left(\frac{(X - c_i)^T (X - c_i)}{2\sigma_i^2}\right) \quad (I.17)$$

Où σ_i est un paramètre qui représente une mesure de la dispersion des données associées à chaque nœud. Il est souvent égal à la distance moyenne entre le centre de la classe et les exemples d'entraînement y faisant partie.

L'utilisation de la fonction gaussienne rend le réseau très puissant. En utilisant cette fonction à caractéristique locale, chaque neurone ne réagit d'une manière significative qu'à une partie restreinte de l'espace d'entrée. En partageant l'espace d'entrée, les aptitudes de ce réseau en approximation se trouvent significativement améliorées par rapport aux réseaux LBF dont les sorties sigmoïdes ne possèdent pas cette propriété de calcul local.

La classification étant effectuée, la couche de sortie, constituée de neurones linéaires, effectue l'approximation définie par la somme pondérée suivante :

$$y_j(X) = \sum_{i=1}^m w_{ij} \phi(\|X - c_i\|) \quad j = 1, \dots, m \quad (I.18)$$

Avec m le nombre de neurones de sorties et k le nombre de neurones cachés.

1.7.3.2. Apprentissage des réseaux RBF :

L'entraînement d'un réseau RBF comprend deux étapes [12] :

- l'apprentissage de la couche cachée : sélection des centres.
- l'apprentissage de la couche de sortie : détermination des poids synaptiques de cette couche.

Dans le cas idéal, un centre doit être placé pour chaque exemple d'entraînement. Ainsi, pour la détermination des valeurs de ces centres, on aura une matrice carrée constituant autant d'équations que d'inconnues. Dans le cas réel, avec un nombre important d'exemples, ceci est impossible. Les centres doivent donc être choisis de sorte à effectuer un échantillonnage représentatif de l'espace des exemples d'entraînement.

Quant aux poids synaptiques, leur détermination est obtenue par n'importe quel algorithme d'optimisation linéaire, du fait que la sortie est linéaire.

Dans ces méthodes d'entraînement, il existe trois approches différentes qui peuvent être utilisées :

- apprentissage supervisé utilisant les algorithmes d'entraînement basé sur l'optimisation ;
- apprentissage non supervisé basé sur un algorithme de regroupement ;
- apprentissage supervisé de regroupement linéaire.

1.8. Algorithmes d'apprentissage des RNA à apprentissage supervisé :

1.8.1. Méthode des moindres carrés (Least Mean Square Error) :

Cette méthode consiste à minimiser l'erreur quadratique définie par [13] :

$$(e^k)^2 = (D^k - W^T X^k)^2 \quad (I.19)$$

Avec,

D^k est le vecteur de sorties désirées,

W^T est la matrice transposée de la matrice poids w ,

X est le vecteur des entrées.

Après l'entraînement avec plusieurs exemples, l'erreur moyenne à minimiser doit avoir un minimum, ce qui revient à résoudre l'équation:

$$\frac{\partial}{\partial w} E[(e^k)^2] = 0 \quad (I.20)$$

Or :

$$E[(e^k)^2] = E(D^k)^2 - 2E(D^k X^k)W^T + WE(X^{kT} X^k)W^T \quad (I.21)$$

En posant R_{aut} la matrice moyenne des auto-corrélations entre les exemples d'entrées et P_{int} la matrice des inter-corrélations entre les entrées et leurs sorties désirées:

$$R_{aut} = E(X^{kT} X^k)$$

$$P_{int} = E(D^k X^k)$$

On aura

$$\frac{\partial}{\partial w} E[(y^k)^2] = -2P + 2wR \quad (I.22)$$

D'où l'on obtient le vecteur poids optimal :

$$W^* = P_{int} R_{aut}^{-1} \quad (I.23)$$

1.8.2. Méthode de descente de gradient :

Elle est destinée à la détermination des adaptations qu'il faut apporter aux vecteurs poids synaptiques à chaque présentation d'un exemple d'entrée, afin de diminuer la variation de l'erreur, jusqu'à l'obtention de la solution optimale. Pour cela, il suffit de parcourir la surface d'erreur dans le sens inverse du gradient, après avoir évalué la variation de l'erreur en calculant le gradient de la fonction objective par rapport aux poids w .

L'erreur quadratique instantanée est :

$$\zeta^k = \frac{1}{2}(e^k)^2 = \frac{1}{2}(d^k - y^k)^2 \quad (I.24)$$

Afin de progresser dans le sens opposé du gradient on a:

$$\Delta w^k = -\eta \frac{\partial \zeta^k}{\partial w^k} = e^k x^k \tag{1.25}$$

A partir des équations (1.24) et (1.25), on obtient:

$$w_j^{k+1} = w_j^k + \eta_k (d_i^k - y_i^k) x_j^k \tag{1.26}$$

Où:

η_k : est le taux d'apprentissage dont le rôle est de réguler la vitesse de convergence et contrôler la stabilité du processus.

Ce réseau étant un approximateur linéaire, l'augmentation du nombre de couches n'est pas utile. L'amélioration de ses performances nécessite l'introduction des fonctions non linéaires qui le rendra sensible à l'augmentation du nombre de couches.

Le pas ne doit pas être choisi trop grand sinon le système risque d'avoir un comportement oscillatoire, alors que s'il est trop petit, l'algorithme aura une convergence très lente. Les valeurs habituelles de pas sont comprises entre 0.1 et 0.7 [9].

1.8.3. Rétro-propagation du gradient (backpropagation) :

Les premières approches datent de la fin des années 60, début des années 70. L'algorithme de rétro-propagation à plusieurs variantes, appelé l'algorithme du retour dynamique, a été découvert en 1974 par Werbos.

En 1986, Rumelhart, Hinton et Williams [14] groupe d'informaticiens, exploitent le pouvoir et le potentiel de la rétro-propagation, ceci a poussé la communauté scientifique à s'y intéresser de plus en plus, ce qui a fait de lui l'algorithme connexionniste le plus utilisé qu'on désigne par "backpropagation". L'erreur se rétropropage à travers les couches du réseau (figure I.9).

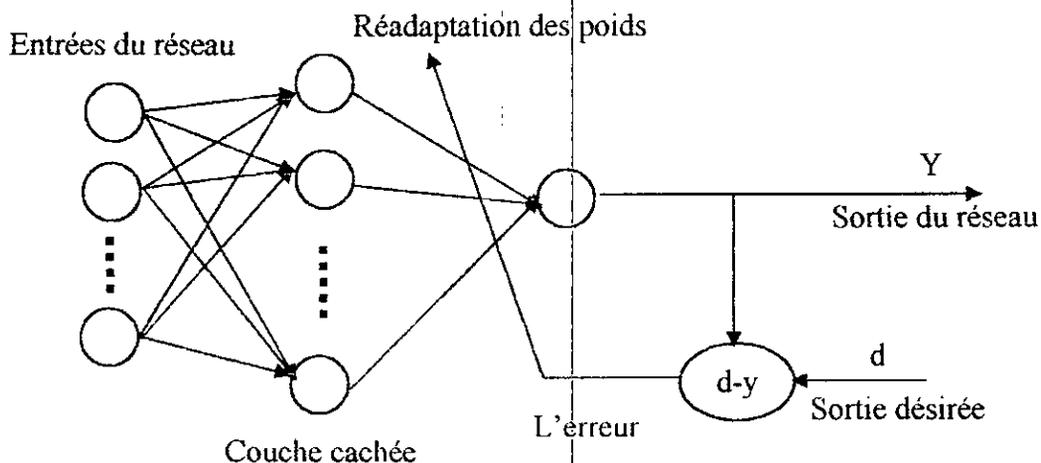


Fig.I.9. Principe de l'entraînement du réseau par rétropropagation du gradient

L'algorithme de base de la backpropagation est celui de la descente du gradient. Contrairement au Madaline, dans l'algorithme de la backpropagation, les poids directement liés à la sortie ont un lien direct avec l'erreur à minimiser. Ceci introduit des calculs supplémentaires à effectuer.

Au début, il faut initialiser les poids synaptiques à de petites valeurs comprises entre -1 et +1. Ensuite il faut :

1. Présenter un nouvel exemple x^p à l'entrée du réseau.
2. Calculer la sortie y_i^p du réseau et l'erreur quadratique E_p à la sortie, correspondant à l'exemple x^p , connaissant la sortie désirée d_i^p :

$$E_p = \sum_{i=1}^{n_L} (d_i^p - y_i^p)^2 \quad (I.27)$$

n_L est le nombre de neurones dans la couche de sortie.

3. Calculer les valeurs de réadaptation des poids :

$$[\Delta w_{ij}^L]_p = \eta \cdot f(u_i^L(t)) \delta_i^L y_j^{L-1} \quad (I.28)$$

Où:

η : est le paramètre qui contrôle la vitesse de l'apprentissage appelé taux d'apprentissage.

$f(u_i^L(t))$ est la sortie du $i^{\text{ème}}$ neurone de la $L^{\text{ème}}$ couche.

Δw_{ij}^L est la valeur rajoutée aux poids.

$$\text{Avec : } \delta_i^L = \begin{cases} (d_i^p - y_i^p) & \text{pour les neurones de la couche de sortie} \\ \sum_{k=1}^{n_{L+1}} \delta_{ki}^{L+1} w_{ki}^{L+1} & \text{pour les autres couches} \end{cases} \quad (I.29)$$

4.
 - s'il s'agit d'un data learning, réadapter les poids
 - poser:

$$\Delta w_{ij}^L(t) = \sum_{p=1}^{p=M} [\Delta w_{ij}^L]_p \quad (I.30)$$

- aller à 5.

- s'il s'agit d'un block learning:
 - Si $p \neq M$: (M étant le nombre d'exemples d'entraînement), retourner à 1.
 - Si non:

$$\Delta w_{ij}^L(t) = [\Delta w_{ij}^L]_p \quad (I.31)$$

5. Réadapter les poids.

$$w_{ij}^L(t+1) = w_{ij}^L(t) + \Delta w_{ij}^L(t) \quad (I.32)$$

6. Répéter de 1 à 5 autant de fois que nécessaire jusqu'à la convergence vers le seuil d'erreur fixé.

Comme le montre l'algorithme ci-dessus, on peut opérer de deux manières différentes, Data Adaptive Learning ou Block Adaptive Learning [9]. Pour la première méthode, la réadaptation des poids se fait au passage de chaque exemple, ce qui rend le processus sensible à chaque exemple individuellement, d'où sa sensibilité aux bruits que peuvent contenir ces entrées durant l'entraînement. Cette méthode n'est utilisée que pour un apprentissage en temps réel. Pour la deuxième méthode, le réajustement ne se fait qu'après le passage de tous les exemples d'entraînement, donc suivant la moyenne de tous les exemples, ce qui rend le processus beaucoup moins sensible aux bruits pouvant se présenter. Cette méthode ne nécessite pas un apprentissage en temps réel.

1.8.3.1. Variantes de la backpropagation :

1.8.3.1.1. Backpropagation à taux d'apprentissage adaptatif :

L'algorithme de la backpropagation réclame une variation infinitésimale des poids à chaque itération. Cette variation est contrôlée par le taux d'apprentissage. Un facteur trop petit assure la convergence mais très lentement. Alors qu'un facteur plus grand génère des changements plus importants dans les valeurs des poids synaptiques, pouvant accélérer ainsi l'apprentissage, mais engendrant aussi une instabilité dans la recherche du minimum global, ce qui compromet la convergence. D'où l'importance du choix de ce paramètre dans l'algorithme d'apprentissage.

Une technique très simple est utilisée afin d'accélérer la convergence. Elle consiste à mettre en œuvre un algorithme qui contrôle l'erreur d'entraînement à chaque étape.

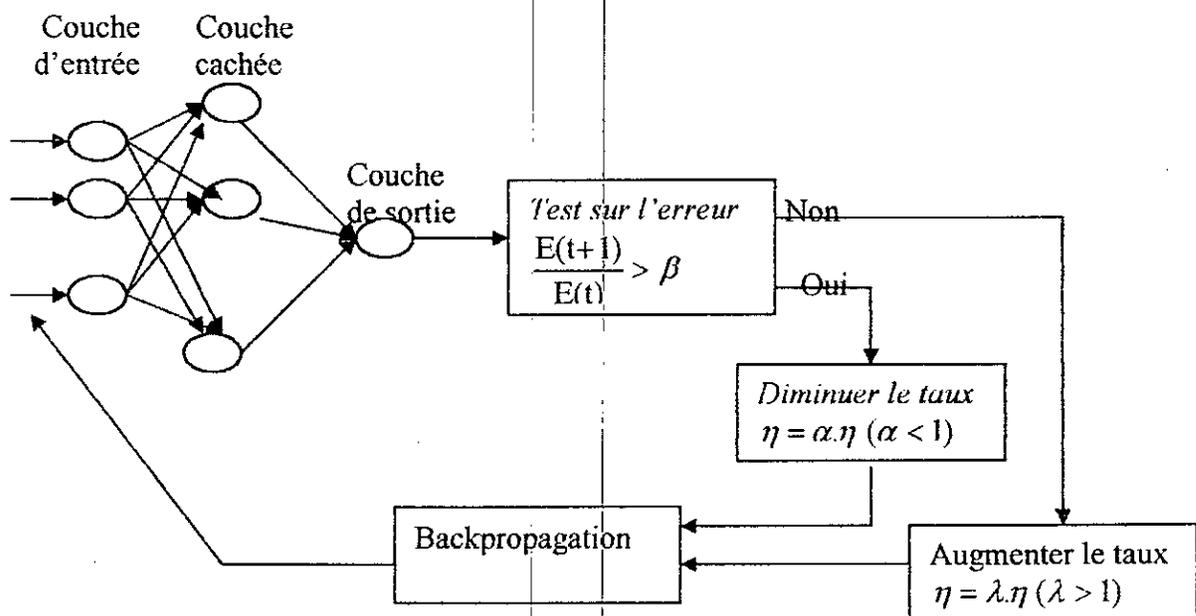


Fig.I.10. Schéma de principe d'adaptation du taux d'apprentissage

Pour un block learning, à chaque fois que l'erreur actuelle $E(t+1)$ dépasse la précédente $E(t)$ par un seuil précédemment fixé (β), due à une augmentation des oscillations d'où risque de divergence, on rejette alors les poids générés, on revient au point précédent et on diminue le taux d'apprentissage. Si par contre l'erreur diminue, les poids générés sont retenus et le taux d'apprentissage est augmenté. De cette manière, on essaye de à chaque étape d'avancer le plus rapidement possible vers l'optimum tout en évitant la divergence de

l'algorithme. Cette méthode donne une convergence plus rapide avec une bonne précision d'entraînement.

1.8.3.1.2. Backpropagation avec momentum :

La backpropagation est très utilisée avec le Momentum. C'est une méthode importante et efficace. D.E.Rumelhart a proposé une solution très efficace pour accélérer la convergence. Elle consiste à utiliser les changements précédents des poids pour la réadaptation des poids actuels.

L'équation d'adaptation devient donc :

$$w_{ij}^l(t+1) = w_{ij}^l(t) + \alpha \cdot \Delta w_{ij}^l(t) + \alpha_m \cdot \Delta w_{ij}^l(t-1) \quad (I.33)$$

Le terme ajouté à l'équation (I.32) est appelé Momentum, en analogie avec la mécanique classique, où un objet en mouvement garde l'impulsion acquise grâce à la quantité de mouvement qui lui a été communiquée précédemment pour accélérer son mouvement.

Le paramètre α_m est utilisé pour pondérer l'effet de ce terme. Sa valeur est généralement comprise entre 0.1 et 0.8. L'adaptation de sa valeur donne des résultats très intéressants. L'utilisation de cette méthode permet de faire sortir les poids des minima locaux, afin de chercher d'autres optimums, ce qui donne beaucoup de chances d'aboutir à un minimum global.

1.8.3.1.3. Robust backpropagation :

Cette variante est insensible aux perturbations affectant les échantillons grâce à des estimations robustes. Parmi les lois de densité les plus utilisées dans le domaine de la statistique, on cite, la loi de gauss ou celle de Cauchy, ou la densité de probabilité basée sur la fonction logistique.

$$S(E_p) = \tanh\left(\frac{E_p}{2}\right) \quad (I.34)$$

Ou sur la fonction de Cauchy

$$S(E_p) = \frac{2 \cdot E_p}{1 + E_p^2} \quad (I.35)$$

1.8.3.1.4. Fast backpropagation :

Dans la backpropagation classique, dès qu'on arrive à l'étape de réduire la fonction erreur, la convergence de l'algorithme devient de plus en plus lente. Ce qui a poussé N.B.Karayanis et A.N.Venetsanopoulos [15] à proposer l'application de l'algorithme de backpropagation avec une fonction objective dite généralisée, pour accélérer cette convergence. Le critère proposé est décrit par:

$$\begin{aligned} G(\lambda) &= \lambda \cdot E_1 + (1-\lambda)E_2 \\ &= \lambda \sum_{k=1}^m \sum_{i=1}^{nl} \frac{1}{2} (e_i^k)^2 + (1-\lambda) \sum_{k=1}^m \sum_{i=1}^{nl} \theta(e_i^k) \end{aligned} \quad (I.36)$$

Pour un réseau LBF, on prend:

$$O(d_i^k - y_i^k) = d_i^k (d_i^k - y_i^k) \quad (I.37)$$

Où :

y_i^k : représente la sortie du $i^{\text{ème}}$ neurone de la couche de sortie du réseau pour le $k^{\text{ème}}$ exemple en entrée.

Ainsi, à chaque itération, on a la réadaptation des poids.

$$W_{ij}^{k+1} = W_{ij}^k - \eta \frac{\delta G(\lambda)}{\delta W_{ij}^k} \quad (I.38)$$

Au début de l'apprentissage $\lambda=1$, puis on le diminue jusqu'à ce qu'il s'annule à la fin de l'apprentissage en fonction de l'erreur suivant la formule:

$$\lambda = \lambda(E) = \exp\left(\frac{-m}{E^n}\right) \quad (I.39)$$

Où:

m : est un nombre positif réel.

n : est un nombre positif entier.

1.8.4. Méthode d'optimisation aléatoire ROM (Random Optimization Method) :

Cette méthode est basée sur la génération des séquences aléatoires afin de déplacer les valeurs des poids dans l'espace à la recherche de l'optimum. En effet, partant du principe de minimiser un critère, cette méthode ne dépend nullement de la nature mathématique du système sur lequel elle est appliquée. Son grand avantage est qu'elle ne nécessite pas la détermination du gradient du critère à minimiser d'une part et d'autre part, elle permet d'échapper aux minimums locaux.

Elle a été proposée en 1965 par J.Mathias, ensuite développée en 1981 par F.J.Sollis et J.B.Wells [16], qui ont démontré la convergence vers un minimum global. L'algorithme régissant cette méthode est comme suit:

1. Initialiser aléatoirement tous les poids w du réseau. Initialiser la variance v et la moyenne de la séquence aléatoire $b = 0$.

2. Générer une nouvelle séquence aléatoire $\xi(k)$ de moyenne $b(k)$ et variance $v(k)$.

3. Calculer les erreurs $E = \sum_{p=1}^m E_p$ à la sortie du réseau pour chaque cas:

$$E(w(k)) \rightarrow E_1$$

$$E(w(k) + \xi(k)) \rightarrow E_2$$

$$E(w(k) - \xi(k)) \rightarrow E_3$$

4.

- Si $E2 < E1$ alors:

$$\begin{cases} w(k+1) = w(k) + \xi(k) \\ b(k+1) = 0.2b(k) + 0.4\xi(k) \end{cases} \quad (I.40)$$

- Si $E2 > E1$ et $E3 < E1$ alors:

$$\begin{cases} w(k+1) = w(k) - \xi(k) \\ b(k+1) = 0.2b(k) - 0.4\xi(k) \end{cases} \quad (I.41)$$

- Sinon,

$$\begin{cases} w(k+1) = w(k) \\ b(k+1) = 0.5b(k) \end{cases} \quad (I.42)$$

5. Poser $k=k+1$ et refaire les étapes de 2 à 4 jusqu'à l'obtention de l'erreur désirée.

Contrairement à la méthode de backpropagation où le calcul du Jacobien est nécessaire, le grand avantage de cette méthode est son indépendance du critère à minimiser. Son utilisation dépend seulement de la variance du vecteur gaussien. Une séquence avec une grande variance génère de grands changements dans l'erreur mais jamais des oscillations. Il est intéressant de commencer l'entraînement avec une variance importante et de la faire baisser au fur et à mesure que l'entraînement avance.

1.9. Réseaux de neurones à apprentissage non supervisé :

Comme dans le cas de l'apprentissage supervisé, l'apprentissage non supervisé des réseaux de neurones consiste à modifier les poids des connexions des neurones.

Cet apprentissage est basé sur la loi, ou postulat, de Hebb établit à partir d'observations d'expériences de neurobiologie : si des neurones, de part et d'autre d'une synapse, sont activés de manière synchrone et répétée, la force de la connexion synaptique va aller croissant. Mathématiquement, cette règle peut s'exprimer de la façon suivante :

$$w_{ij}(t+1) = w_{ij}(t) + \eta \cdot y_j(t) \cdot x_i(t) \quad (I.43)$$

Où $x_i(t)$ et $y_j(t)$ sont les sorties, au temps t des neurones i et j dont le poids de connexion (entre i et j) vaut $w_{ij}(t)$ et η est le coefficient d'apprentissage, où x_i est l'entrée de la synapse. L'une des propriétés remarquable de cette règle est qu'elle exprime que l'apprentissage se fait localement, c'est-à-dire que la modification de w_{ij} ne dépend que de l'activité des cellules i et j .

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

1.9.1. Mémoires associatives :

Les mémoires associatives ont été proposées par plusieurs auteurs dès 1977 dont T. Kohonen. Dans mémoire associative, le terme "mémoire" fait référence à la fonction de stockage de l'information et le terme "associative" au mode d'adressage. L'expression "mémoire adressable par son contenu" est aussi souvent employée.

Une mémoire associative agit en deux temps : stockage et récupération. Durant la phase de stockage, les poids du réseau sont déterminés de telle sorte que les attracteurs du réseau mémorisent un ensemble de p patrons de dimension n $\{x_1, \dots, x_p\}$. On peut utiliser une généralisation de la règle de Hebb pour positionner les poids w_{ij} . Dans la phase de récupération, la donnée sert d'état initial au réseau qui évoluera en accord avec sa dynamique. La réponse est obtenue lorsque le réseau se stabilise.

L'information mémorisée ne peut être obtenue à une adresse précise, le seul moyen d'accès est de fournir une information. Il existe deux types de mémoire, auto-assocriative et hétéro-assocriative.

1.9.1.1. Mémoires auto-assocriatives :

Dans le cas des mémoires auto-assocriatives, il faut fournir tout ou partie de l'information mémorisée. Ces mémoires sont donc principalement utilisées pour la reconstruction de données : l'opérateur fournit une information partielle que le système complète. Des expérimentations dans ce sens ont été faites avec l'annuaire électronique où l'utilisateur tape le maximum d'informations relatives à sa demande, que le système complète et corrige.

1.9.1.2. Mémoires hétéro-assocriatives :

Les mémoires hétéro-assocriatives se différencient des précédentes en rendant une information différente. Par exemple, si la clef d'entrée est une image de visage, le système répond par le nom de la personne correspondante et vis versa.

1.9.2. Réseaux compétitifs :

À la différence de la règle de Hebb (dans laquelle plusieurs neurones peuvent être activés en sortie), cet apprentissage n'active qu'un seul neurone à la sortie (Fig.I.10). On parle de "winner-take-all" (WTA, qu'on peut traduire par 'tout au vainqueur') : un tel phénomène a été mis en évidence dans le cas de réseau biologique.

Chaque cellule de sortie est connectée aux cellules d'entrée, ainsi qu'à toutes les cellules voisines de la couche de sortie (connexion inhibitrice) et à elle-même (connexion excitatrice).

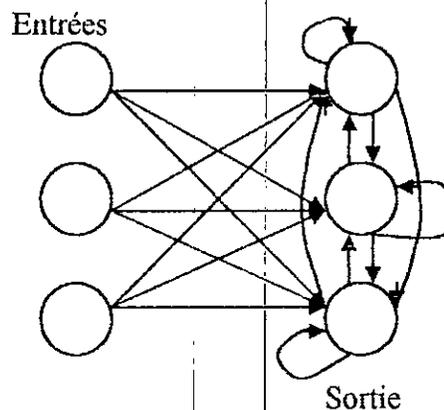


Fig.I.10

Le résultat de la compétition est de choisir la cellule i_0 ayant la plus grande (ou la plus petite) entrée, c'est-à-dire que l'on a :

$$w_{i_0}x \geq w_i x, \forall i \quad (1.44)$$

Ou bien,

$$\forall i, \|w_i - x\| \leq \|w_{i_0} - x\| \quad (1.45)$$

Ce qui, dans le cas où les vecteurs de poids sont normalisés, est équivalent.

1.9.3. Cartes auto-organisatrices :

Les cartes auto-organisatrices sont depuis longtemps connues (1977), mais ce n'est que très récemment, dans les années 90, que des applications les utilisent : carte phonétique, diagnostic de pannes, compression d'images, robotique, etc.

Ce modèle appartient à la classe des réseaux à compétition. C'est l'algorithme le plus connu, appelé aussi SOM pour Self-Organization Maps ou carte topologique de Kohonen [5].

Un réseau de Kohonen est constitué d'une couche compétitive unique de neurones de sortie dont la compétition est basée sur le principe WTA. L'apprentissage dans la couche est non supervisé et est caractérisé par une généralisation au voisinage du neurone élu. L'algorithme de Kohonen représente une simplification algorithmique de la loi d'activation compétitive dans une couche de neurones totalement interconnectés.

Les neurones de la couche (appelée couche topologique) sont disposés géométriquement selon une topologie bien définie dans un réseau de dimension p . ceci signifie que pour $p=2$, le réseau a une forme carrée (un réseau hexagonal est quelquefois utilisé) et devient une grille où chaque neurone possède quatre voisins (excepté pour ceux situés sur les bords). Dans la plupart des applications, les neurones d'une carte de Kohonen sont disposés sur une grille deux dimensions 2D (Fig.I.11).

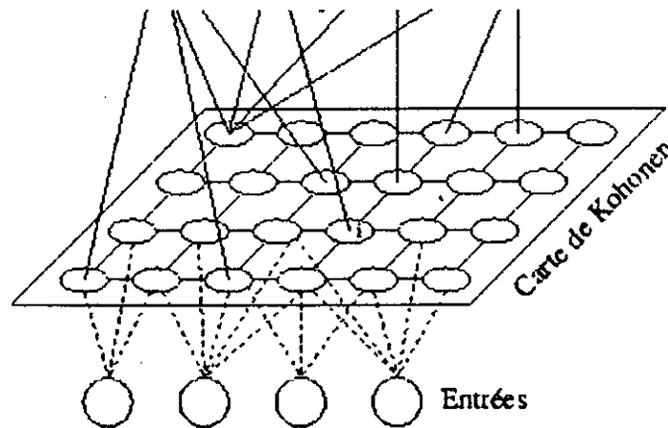


Fig.I.11. Réseau de Kohonen à 2D

Le poids associé aux liaisons latérales entre neurones est fonction de la distance entre le neurone source et le neurone cible. La valeur du poids est donnée par une fonction "chapeau mexicain" (Difference Of Gaussians - DOG - voir Fig.I.12).

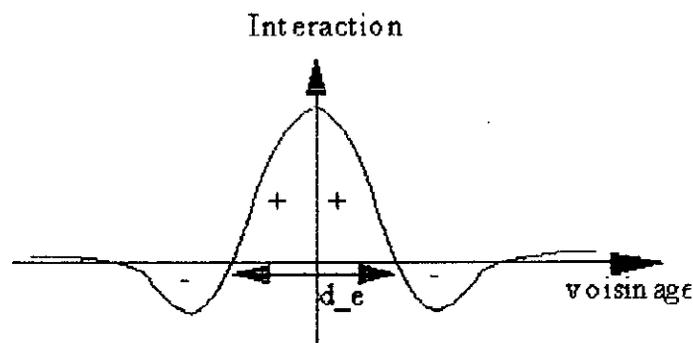


Fig.I.12. Influence d'un neurone sur ses voisins en fonction de l'éloignement.
+ : excitatrice ($w > 0$), - : inhibitrice ($w < 0$), d_e : taille du voisinage excitateur

Cette courbe s'interprète de la manière suivante :

- Dans la zone latérale proche du neurone gagnant (50 à 100 μm), les neurones qui lui sont connectés ont une action excitatrice.
- Dans un voisinage plus lointain (200 à 500 μm), leur action est inhibitrice.
- L'action des neurones les plus lointains est négligeable.

La loi de modification des poids des connexions (poids synaptiques) est dérivée de celle de Hebb. Dans le cas où les exemples d'entrées sont des vecteurs à deux composantes, l'algorithme d'apprentissage est le suivant :

1. Initialisation des poids à des valeurs aléatoires.
2. Présentation d'une entrée $E1 = (e1, e2)$.
3. Calcul de la distance de chacun des neurones par rapport à $e1$ et $e2$:

$$x_j = |w_{j1} - e1| + |w_{j2} - e2| \quad (I.46)$$

4. Sélection du neurone le plus proche : $\text{Min}(x) = x_i$
5. Modification des poids pour le neurone choisi (i) et ses 4 plus proches voisins (k).
 μ et β sont deux paramètres correspondant au pas de modification des poids. μ pour le neurone choisi et β pour ceux du voisinage.

$$w_{i1} = w_{i1} + \mu \cdot (e1 - w_{i1})$$

$$w_{i2} = w_{i2} + \mu \cdot (e2 - w_{i2})$$

$$w_{k1} = w_{k1} + \beta \cdot (e1 - w_{k1})$$

$$w_{k2} = w_{k2} + \beta \cdot (e2 - w_{k2})$$

- 6) Tant que les performances sont insuffisantes : Retour à l'étape 2 et sélection de l'exemple suivant dans la base d'apprentissage.

1.9.4. Modèle de résonance adaptative (Adaptive Resonance Theory ART) :

Le dilemme "stabilité-plasticité" est une notion fondamentale dans le cadre des réseaux à compétition. Le problème d'un système trop plastique est qu'il peut apprendre des informations même si elles ne sont pas pertinentes ou oublier des informations apprises. A l'inverse, un système trop rigide empêche tout apprentissage. La résolution de ce dilemme entre plasticité et stabilité est donc centrale à toute recherche visant à comprendre ou créer un système capable d'apprendre de manière autonome. Comment apprenons-nous de nouvelles choses (plasticité) tout en gardant une stabilité garante d'une connaissance ni supprimée ni "abîmée".

L'ART, un modèle de réseau de neurones à architecture évolutive, développé en 1987 par Carpenter et Grossberg, essaye de résoudre ce dilemme. Avec le réseau ART, la phase d'apprentissage et la phase d'exécution sont réalisées simultanément. Le réseau en phase de test, s'adapte à des entrées inconnues en construisant de nouvelles classes (ajout de neurones) tout en dégradant au minimum les informations déjà mémorisées.

Le réseau possède un réservoir de cellules de sortie qui ne sont utilisées que si nécessaire. Une cellule sera dite recrutée (committed) ou libre (uncommitted) si elle est utilisée ou non utilisée. L'algorithme d'apprentissage met à jour les vecteurs prototypes stockés uniquement s'ils sont suffisamment proches du patron fourni en entrée au réseau. Un patron et un prototype "résonnent" lorsqu'ils sont voisins. La notion de similarité est contrôlée via un paramètre de vigilance, ρ , compris strictement entre zéro et un, qui détermine ainsi le nombre de catégories reconnues. Lorsqu'un patron n'est pas assez proche des vecteurs prototypes déjà présents dans le réseau, une nouvelle catégorie est créée et une cellule libre y

est assignée avec comme vecteur prototype le patron correspondant. S'il n'existe plus de cellules libres dans le réservoir, le patron ne provoque pas de réponse du réseau.

Il existe plusieurs versions de réseaux (ART1, ART2, ART3). Le réseau ART1 est un réseau à entrées binaires. Afin d'illustrer cette approche, on prend exemple du ART1.

Le réseau ART1 est formé d'une couche d'entrée qui est aussi la couche de sortie et d'une couche cachée. Il n'y a pas de connexion entre les neurones d'entrées. Par contre, la couche cachée est une couche d'activation compétitive, tous les neurones sont reliés les uns aux autres par des connexions inhibitrices de poids fixes. Chaque neurone de la couche d'entrée est relié à tous les neurones de la couche cachée et, réciproquement, chaque neurone de la couche cachée est relié à tous les neurones de la couche de sortie (Fig.I.13). A chaque connexion est associé un poids.

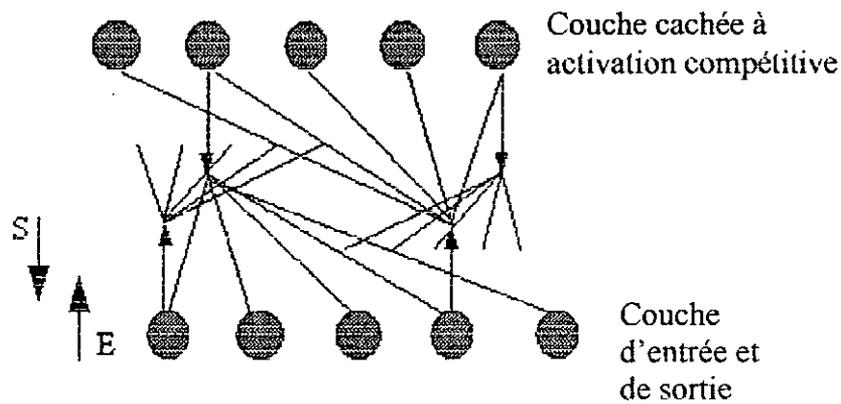


Fig.I.13. Architecture du réseau ART1

La Fig.I.14 montre un vecteur d'entrée E soumis au réseau. A cette entrée correspond, après compétition entre les neurones de la couche cachée, un unique neurone j gagnant. Ce neurone gagnant est considéré par le réseau comme le plus représentatif du vecteur d'entrée E . Le neurone j génère en retour sur la couche de sortie un vecteur S binaire (seuillage). S est ensuite comparé au vecteur d'entrée E .

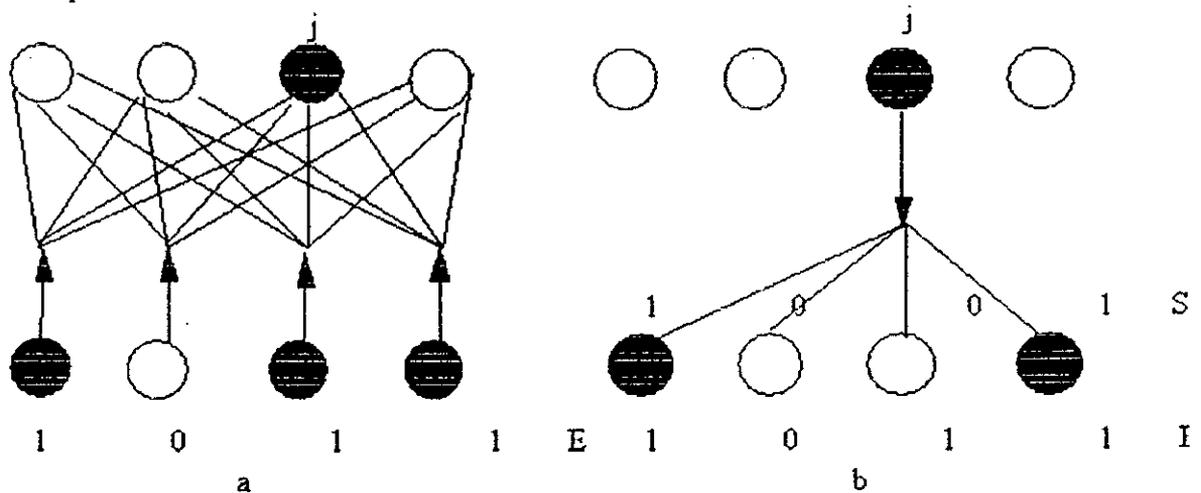


Fig.I.14. Fonctionnement du réseau ART1.

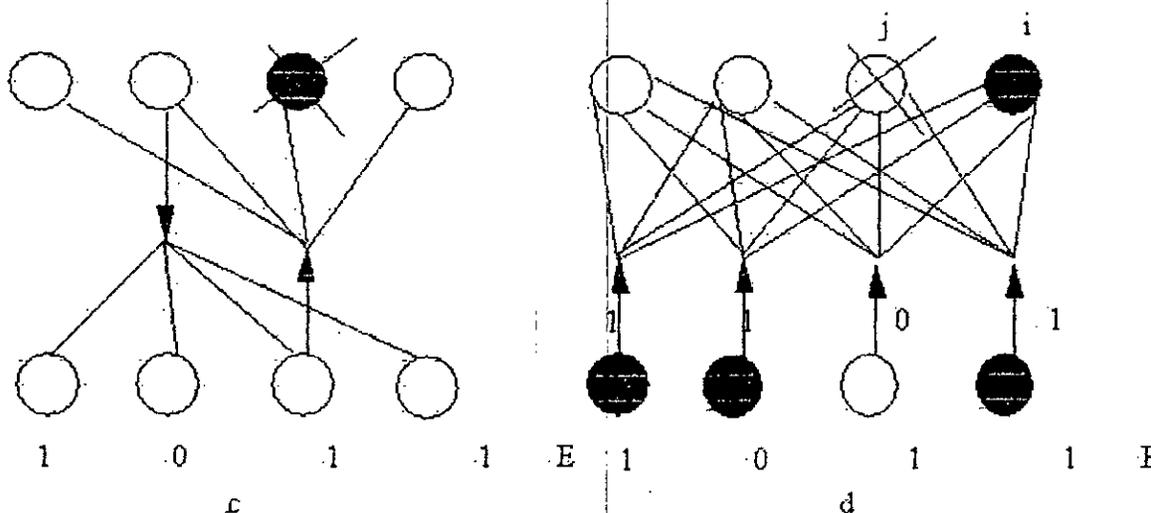


Fig.I.14. Fonctionnement du réseau ART1.

- Présentation du vecteur d'entrée E , un neurone gagnant j est sélectionné.
- Tentative d'unification entre S (retour du neurone j) et E .
- Echec : suppression du neurone gagnant, présentation de E .
- Unification : le neurone i est un représentant de la classe du vecteur d'entrée E .

Si la différence est inférieure à un seuil fixé pour le réseau, le neurone gagnant est considéré comme représentant de la classe du vecteur d'entrée. Dans ce cas, la modification des poids des connexions du neurone gagnant a pour effet de consolider ses liens d'activation avec l'entrée E ; en fait l'adéquation entre ce vecteur d'entrée et cette classe est améliorée. Dans le cas contraire, le processus reprend avec les neurones de la couche cachée moins le neurone gagnant de l'étape précédente. Si tous les neurones cachés sont passés en revue sans qu'aucun ne corresponde à E , un nouveau neurone caché est ajouté, qui est initialisé comme représentant de la classe du vecteur d'entrée E .

Le réseau ART1 a une architecture en deux couches qui interagissent entre elles. Le réseau se distingue aussi par deux caractéristiques : sa flexibilité et sa stabilité pour des entrées arbitraires. Il est capable de s'adapter à des entrées non familières en construisant de nouvelles catégories ou classes (flexibilité, plasticité) mais aussi d'adapter les classes déjà apprises tout en dégradant peu les informations déjà mémorisées (stabilité). Le problème posé par ces deux notions antagonistes (flexibilité-stabilité) est résolu par le principe de l'architecture évolutive.

I.10. Apprentissage par renforcement :

L'algorithme d'apprentissage par renforcement, appelé aussi apprentissage par pénalité/récompense, est de type qualitatif par opposition aux apprentissages dits supervisé ou non supervisé. Il n'est pas nécessaire de disposer pour les exemples d'apprentissage des sorties désirées, seulement d'une appréciation "globale" du comportement du réseau pour chacun des exemples traités. Cet algorithme s'applique à toutes structures de réseaux. La seule condition est de disposer de neurones de sortie stochastiques (binaires). La réponse du réseau

de neurones est ainsi fonction des entrées et, aussi, des neurones de sortie. On introduit donc à ce niveau une part d'aléatoire dans le comportement du système. Si la réponse fournie par le système est considérée comme bonne, l'algorithme tend à favoriser l'apparition de ce comportement en réduisant l'aléatoire. Dans le cas où la réponse du système globale est considérée comme mauvaise, on cherche à éviter l'apparition ultérieure de ce comportement. Ce processus est itéré jusqu'à l'obtention du comportement désiré pour l'ensemble du système

L'algorithme est de la forme suivante :

1. Les poids sont initialisés à de petites valeurs aléatoires qui placent les probabilités des neurones de sortie autour de 0.5.
2. Une entrée $E_i = (e_1, \dots, e_n)$ est présentée,
3. Une sortie correspondante possible x_i est calculée pour chaque neurone,
4. La sortie globale produite est analysée de façon à générer un signal de retour r , positif ou négatif, et une sortie cible (désirée) est choisie :

$$d_i = x_i \text{ si } r = +1 \text{ (récompense)}$$

$$d_i = -x_i \text{ si } r = -1 \text{ (pénalité)}$$

5. La modification des poids est réalisée par la classique méthode du gradient :

$$\sum w_{ij} = \mu \cdot r \cdot \text{erreur}_i \cdot x_j$$

En général, μ dépend de r et est pris 10 à 100 fois plus grand (μ^+) pour $r = +1$ que pour $r = -1$ (μ^-).

6. Tant que la sortie du réseau n'a pas produit une séquence satisfaisante suffisamment longue, retour à 2.

1.11. Problèmes d'apprentissage :

L'apprentissage d'un RNA passe par la définition de plusieurs paramètres ou choix s'agissant de la règle d'apprentissage ou du réseau. Chacun d'entre eux a une influence positive ou négative sur l'apprentissage lui-même. D'où la difficulté de choisir ces paramètres. Il existe un certain nombre de problèmes qu'on peut rencontrer lors de l'apprentissage. Parmi eux [17] :

- Insuffisance de la règle d'apprentissage, rien ne garantit que la règle d'apprentissage soit capable de tirer profit de plein potentiel du réseau.
- Minima locaux. La majorité des apprentissages supervisés procèdent par descente du gradient. Or, ce type de technique peut conduire à une solution sous optimale. En pratique, ce type de problème est particulièrement difficile à éviter, notamment parce que la surface d'erreur n'est généralement pas connue. Sachant que les poids initiaux du réseau conditionnent en partie les résultats de l'apprentissage, plusieurs chercheurs relancent l'apprentissage à partir du choix des poids initiaux différents pour se préserver de ce problème.
- Mauvais choix des paramètres initiaux, comme par exemple le nombre de neurones cachés du réseau. Cette valeur est particulièrement importante parce qu'elle détermine le nombre de variables libres (c'est à dire de poids) du réseau, et parce qu'elle détermine les capacités de calcul du réseau. Ainsi, un nombre insuffisant de neurones cachés peut compromettre la capacité du réseau à résoudre la tâche donnée. Inversement, trop de

neurones cachés permettent au réseau d'apprendre les particularités de corpus d'apprentissage (d'apprendre les corpus par cœur), au détriment des performances de généralisation.

Les paramètres propres à la règle d'apprentissage sont souvent aussi difficiles à choisir que ceux du réseau. Par exemple, la plupart des règles d'apprentissage comprennent un pas d'apprentissage, valeur qui affecte la taille du vecteur de correction. Une fois de plus, le bon choix de ce paramètre peut relever de l'exploit : un pas trop petit empêche le réseau de se stabiliser sur une solution et rend la convergence lente, tandis qu'un pas trop grand augmente inutilement la quantité de calculs à effectuer et peut empêcher le réseau de converger.

- Surapprentissage. Quand l'apprentissage d'un réseau est prolongé, son choix de poids reflète de trop près les particularités de corpus d'apprentissage, au détriment de la tâche réelle.
- Mauvais échantillonnage. Les corpus d'apprentissage ne reflètent pas toujours adéquatement la tâche, soit par mauvais échantillonnage, soit tout simplement parce qu'il est trop petit. Il en résulte que le réseau généralise mal et se comporte incorrectement quand il est mis en situation réelle.
- Oubli et interférence. Certains patrons peuvent interférer avec l'apprentissage d'autres patrons. Cela est particulièrement vrai lorsque le corpus d'apprentissage est grand, ou que le comportement à apprendre comprend plusieurs cas exceptionnels. Ces derniers sont alors difficiles à apprendre par le réseau et peuvent nécessiter un traitement particulier.
- Mauvais codage. Il se peut que la technique de codage choisie accule ou rende inutilement obscure l'information nécessaire à la résolution du problème.

1.12. Conclusion :

Dans ce premier chapitre, nous avons fait une introduction aux réseaux de neurones artificiels. Nous avons aussi parlé des réseaux de neurones biologiques, dont s'inspirent les RNA. Nous avons donné quelques définitions et notions de bases : architecture, principe de fonctionnement et apprentissage.

Nous avons présenté différents types de réseaux de neurones classés en deux grandes familles, selon le type d'apprentissage, supervisé et non-supervisé.

Les réseaux RBF, réseaux supervisés, sont de meilleurs approximateurs que les réseaux LBF, réseaux supervisés aussi, qui souffrent de l'impossibilité de déterminer la dimension permettant d'obtenir un meilleur résultat et de l'oubli.

Dans les réseaux non-supervisés, nous avons vu les réseaux compétitifs comme le 'Winner Take All' et des réseaux à cartes auto-organisatrices.

Quant aux méthodes d'apprentissage, La méthode de backpropagation constitue l'algorithme d'apprentissage supervisé le plus utilisé. Toutefois, cette méthode converge trop lentement et dépend étroitement de la base d'apprentissage. D'autre part, cet algorithme se bloque souvent dans des minimums locaux. L'algorithme de descente de gradient a l'avantage de son indépendance par rapport au nombre de couches cachées, mais il présente l'inconvénient du choix des taux d'apprentissage, des poids, des centres et des variances. S'ils sont trop petits, le temps d'exécution va être trop lent. Par contre, s'ils sont trop grands les paramètres à ajuster oscillent autour de l'optimum.

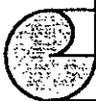
L'algorithme d'optimisation aléatoire, ROM, ne dépend pas du système sur lequel il est utilisé. De plus sa technique de recherche permet d'atteindre un minimum global.

Pour conclure, les récents développements dans le domaine des RNA ont suscité à la fois enthousiasmes et critiques. Pour de nombreuses tâches, aucune approche n'est vraiment supérieure. Le choix de la meilleure technique doit en fait être guidé par la nature du problème à résoudre. Il faut chercher à mieux comprendre les capacités, les hypothèses d'utilisation et les domaines d'applications des différentes approches afin d'utiliser au mieux leur complémentarité en vue d'obtenir des systèmes plus intelligents. Il semble qu'une approche de ce type permettrait une synergie dont l'intérêt serait de combiner les avantages des différentes techniques à même de résoudre un problème. Il est intéressant d'essayer de combiner des systèmes multi-composantes, qui nécessitent une bonne technologie d'intégration. Il est clair que la coopération et la communication entre chercheurs du domaine des RNA et d'autres disciplines permettra, ou permettrait, d'éviter des travaux redondants mais, et plus important encore, stimulera et bénéficiera à chacune des disciplines.



CHAPITRE II

SYNTHESE DES TRAVAUX
EFFECTUES DANS LE
DOMAINE DE LA HT
UTILISANT LES RNA



Dans ce chapitre, nous présentons une synthèse bibliographique, dans un ordre chronologique, de quelques travaux utilisant les réseaux de neurones dans le domaine de la haute tension. Ces réseaux ont été utilisés pour l'identification, la classification, la modélisation, le contrôle où encore la prévention des problèmes rencontrés dans les réseaux électriques. Ce n'est que ces dix dernières années que les RNA ont fait leur apparition et ont été appliqué dans la haute tension.

1994 A.A.Mazroua et al. [18,19] ont publié dans IEEE deux travaux sur la reconnaissance des tailles d'impulsions de décharges partielles 'DP' produites par différentes sources de décharges en utilisant un réseau compétitif : classificateur du plus proche voisin (nearest neighbour classifier), un réseau LVQ (Learning Vector Quantization) et un réseau multicouches utilisant l'algorithme d'apprentissage de la backpropagation. Les trois RNA ont présenté les mêmes performances pour toutes les données présentées sauf pour le cas des petites cavités. Le réseau LVQ a montré les meilleures capacités de reconnaître la différence entre les DP produites par une seule cavité et celles produites par une arborescence électrique.

1995 K.Engel et D.Pier [20] ont étudié l'influence du développement des décharges partielles sur l'identification du type de défaut en utilisant les réseaux de neurones artificiels. Le changement d'activité des DP détermine les données mesurées et leur classification. Dans cette étude, la classification des DP est effectuée en utilisant la carte de Kohonen.

La mesure des DP contribue à la qualité de contrôle. Le but étant d'avoir une information supplémentaire sur la source de la DP et la dégradation de l'isolant. L'identification et la classification des types de DP sont basées sur les quantités mesurables de ces dernières. Le changement d'activité des DP est essentiellement dû à la formation des surfaces conductrices et à la dégradation de l'isolation. D'autres paramètres peuvent influencer ce changement, tel que la pression, les décharges dans les petites cavités de gaz par de fortes interactions entre le gaz et son environnement. Les différents paramètres qui influent sur ces décharges sont présentés par la figure ci-dessous :

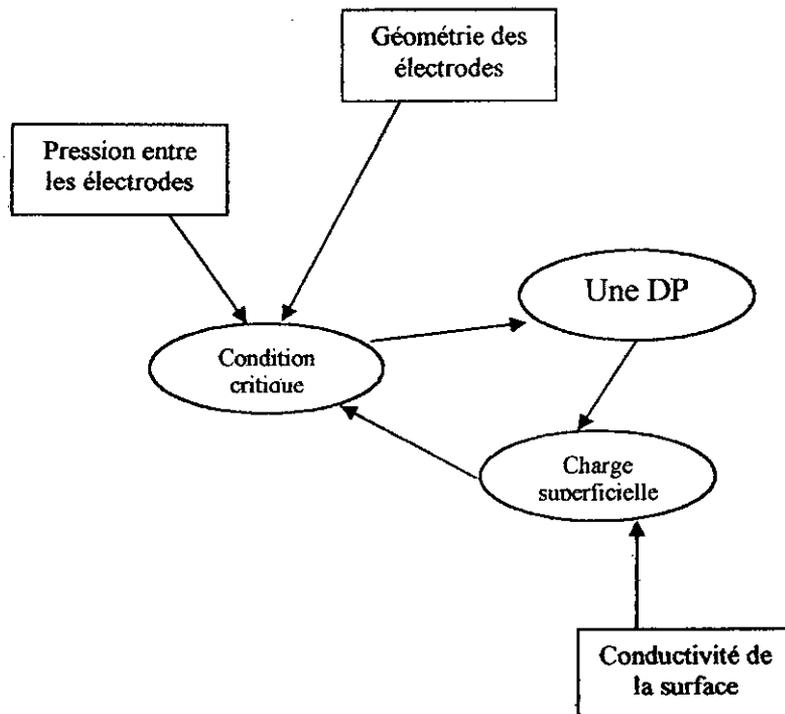


Fig.II.1. Influence des différents paramètres sur l'activité des DP

Les défauts des DP examinés sont divisés en trois groupes. Le premier groupe contient les activités des DP à court terme, lui-même regroupe huit types de défauts. Le second comprend l'activité des DP dans une ligne moyenne tension. Le troisième groupe représente l'activité des DP pour une électrode sphérique en Polyéthylène traduit par une présence de surface conductrice et une variation de pression.

Un micro-ordinateur enregistre l'activité des DP. Les paramètres liés aux DP mesurés sont : l'impulsion apparente, le déphasage par rapport à la fréquence du réseau, la polarité de la DP, la fréquence d'apparition et la distance entre impulsions. Ces paramètres constituent les composantes des vecteurs d'apprentissage du réseau de neurones.

Les résultats obtenus montrent que l'activité des DP est due à une création d'une surface conductrice et à une variation de pression. Les études séparées de chaque paramètre prouvent leurs influences sur la classification en utilisant le modèle de Kohonen.

Dans la même année, P.S.Ghosh, S.Chakravorti et N.Chatteyee [21] ont modélisé un arc électrique d'un isolant pollué en utilisant un réseau de neurones artificiels entraîné par l'algorithme d'apprentissage de type « Resilient Propagation Algorithm (RPROP) ». Le but de cette étude est de modéliser correctement l'effet de la résistance de la couche polluante (R_p), la longueur de la couche polluante (L) et la largeur de la surface de la couche polluante (W), sur la tension de contournement de l'isolant pollué (V_c).

L'algorithme utilisé, le RPROP, qui est une amélioration de l'algorithme de backpropagation qui converge très lentement, est un nouvel algorithme d'apprentissage qui prend en considération la topologie locale de la fonction erreur E . L'adaptation des poids ou leur modification dépend des signes des dérivées de E et non pas de leurs amplitudes. L'adaptation des poids est basée sur ce qu'on appelle la règle d'apprentissage de Manhattan:

$$\Delta W_y(t) = \begin{cases} -\Delta & \text{si } \frac{\delta E(t)}{\delta W_y} > 0 \\ +\Delta & \text{si } \frac{\delta E(t)}{\delta W_y} < 0 \\ 0 & \text{else} \end{cases} \quad (II.1)$$

Où Δ est la valeur d'adaptation, E l'erreur moyenne et W_{ij} les poids des connections.

Les poids évoluent durant l'apprentissage d'une valeur Δ_{ij} donné par :

$$\Delta_{ij}(t) = \begin{cases} n^+ \cdot \Delta_{ij}(t-1) & \text{si } \frac{\delta E(t-1)}{\delta W_{ij}} \frac{\delta E(t)}{\delta W_{ij}} > 0 \\ n^- \cdot \Delta_{ij}(t-1) & \text{si } \frac{\delta E(t-1)}{\delta W_{ij}} \frac{\delta E(t)}{\delta W_{ij}} < 0 \\ \Delta_{ij}(t-1) & \text{else} \end{cases} \quad (II.2)$$

Où $0 < n^- < 1 < n^+$.

A chaque fois que la dérivée partielle du poids W_{ij} change de signe cela veut dire que la dernière valeur d'adaptation est très grande et l'algorithme tombe dans un minimum local, la valeur d'adaptation Δ_{ij} est donc diminuée par le facteur n^- . Si les dérivées retiennent leurs signes, la valeur d'adaptation est légèrement augmentée pour accélérer la convergence dans les régions superficielles. Cependant, il faut faire une petite modification dans la formule

(II.1) dans une phase de transition. Si les dérivées partielles changent de signes, ce qui mène à un saut dans un minimum, l'étape d'adaptation précédente devient :

On suppose que les dérivées changent de signes une fois par étape. Pour éviter un

$$\Delta W_{ij}(t) = -\Delta W_{ij}(t-1) \quad \text{si} \quad \frac{\delta E(t-1)}{\delta W_{ij}} \frac{\delta E(t)}{\delta W_{ij}} < 0 \quad (II.3)$$

$$\frac{\delta E(t-1)}{\delta W_{ij}} = 0 \quad (II.4)$$

double calcul, on ne doit pas adapter une valeur déjà ajustée dans l'étape précédente. En pratique, ceci est obtenu en posant :

Les valeurs d'adaptation et les poids changent à chaque fois qu'un vecteur d'apprentissage est donné. De plus, il faut noter que toutes les valeurs d'adaptation sont initialisées à Δ_0 et sont limitées à $\Delta_{\max} = 50$.

Dans cette étude, P.S.Ghosh a modélisé la tension de contournement de l'isolant pollué $V_c = f(R_p, L, W)$ en utilisant les RNA avec l'algorithme d'apprentissage RPROP. Cette approche a donné de meilleurs résultats que d'autres approches plus classiques tels que les approches empiriques et analytiques.

Dans un autre travail, M.E.Ghourab [22] a étudié l'application des RNA pour identifier l'effet de la moisissure, due à l'humidité, sur le spectre de polarisation (SP) de l'isolation d'un transformateur. L'algorithme utilisé est celui de la backpropagation.

Les transformateurs sont très utilisés dans les réseaux électriques d'où l'importance de leur bon fonctionnement qui est directement lié à leur isolation. L'isolant le plus utilisé dans les transformateurs remplis d'huile est la cellulose. Elle a de bonnes caractéristiques mécaniques et électriques d'une part, et un bon rapport qualité/coût d'autre part. L'humidité tend à réduire et à dégrader les caractéristiques de l'isolant. A la longue, une rupture de l'isolant peut se produire. La durée de vie d'un isolant humide est 20 fois plus faible que celle d'un isolant sec avec 2% d'eau.

L'isolation huile/papier peut être caractérisée par plusieurs paramètres tels que la solidité, le degré de polarisation, l'analyse du gaz dissous, la distribution moléculaire du gaz et les mesures des propriétés électriques.

Dans cette étude, le RNA est appliqué pour la distinction des différents SP. L'algorithme de la backpropagation est utilisé pour faire l'apprentissage du réseau des informations sur les SP de l'isolation avec l'huile/papier pour différentes valeurs d'humidité.

Les tests expérimentaux basés sur le retour de la tension sont choisis pour déterminer le SP des échantillons d'huile/papier contenant des valeurs connues d'humidité. Ces tests sont appliqués avec les pourcentages d'humidité suivants : 0.5, 1, 2, 3 et 4% sous une température de 50°C. Les mesures obtenues sont normalisées par la valeur maximale de l'humidité. Le modèle d'entrée du RNA est le vecteur X de composantes $(x_1, x_2, x_3, \dots, x_m)$. Chaque modèle est lié à une certaine partie du SP.

Les valeurs désirées d'apprentissage correspondantes aux différentes valeurs d'humidité sont présentées dans le tableau II.1.

Une fois l'apprentissage terminé, le réseau devient capable de faire une généralisation et donc de donner des sorties à des entrées inconnues

Vecteurs d'apprentissage			Humidité correspondante
0	0	0	0.5%
0	0	1	1%
0	1	0	2%
0	1	1	3%
1	0	0	4%

Tableau.II.1. Vecteurs d'apprentissage correspondants à différentes valeurs d'humidité

La méthode de mesure par retour de tension est appliquée pour un groupe de six transformateurs de différentes durées de service. Quand le RNA peut reconnaître correctement un vecteur d'entrée, ce vecteur est exploré et discriminé par le réseau de neurones. Les vecteurs d'apprentissage du réseau sont distingués en cinq groupes. Chaque groupe appartient à un état de matériaux isolants.

L'augmentation de la quantité d'eau, existant dans le transformateur, implique celle de la densité d'hydrogène menant à l'augmentation des protons dans le matériel, ce qui provoque à son tour l'augmentation de la conductivité et une polarisation. Ce processus conduit au changement des valeurs maximales et de la forme du SP en fonction de l'humidité. Dans ce cas, une condition donnée de l'humidité tend soit à renforcer le champ local, soit à augmenter le courant de fuite, ce qui peut causer la formation de surface de basse densité ou celle des bulles. L'initialisation des décharges locales ou d'arcs dans les zones à basse densité peut conduire au claquage. De plus, une valeur élevée d'humidité peut causer une panne comme résultat à l'augmentation des pertes diélectriques sous une température élevée.

Les résultats des tests ont montré que l'utilisation du perceptron multicouches comme type de RNA est capable de distinguer la différence entre les SP dans les différentes conditions de service. En plus de la précision et de l'exactitude d'apprentissage, le RNA est aussi rapide pour la distinction en temps réel.

1997 T.Kexioug et Z.Deheng [23] ont utilisé les RNA pour la reconnaissance des modèles de décharges partielles tridimensionnelles. Ils ont utilisé l'influence du nombre de neurones des couches d'entrées et cachées afin de choisir la meilleure structure du réseau. La capacité des réseaux de neurones à reconnaître les DP est analysée par une comparaison entre les résultats analytiques et expérimentaux.

Une autre étude, celle de L.Tang et M.R.Raghvveer [24], a été faite sur la classification du vieillissement des transformateurs de puissance dans des conditions normales, de surtension et d'arc électrique, en utilisant les RNA. Ceci est obtenu par l'analyse des gaz dissous.

Pour améliorer les performances des appareils de puissance, il est important de détecter et de classer les défauts en fonction du temps et du rendement. C'est dans ce but que l'analyse des gaz dissous (AGD) est largement utilisée dans le monde entier comme un test de routine pour détecter d'éventuels défauts dans les huiles des transformateurs. Les gaz dissous dans les huiles de transformateurs sont produits par la détérioration de l'huile, souvent due aux DP, aux arcs électriques et aux surtensions.

Le RNA assume parfaitement les tâches de reconnaissance et de manipulation des perturbations. Il est capable de classer différents types de défauts quand il est entraîné par les résultats obtenus par la méthode de AGD.

Toujours en 1997, H.Okubo [25] a étudié l'optimisation du champ électrique dans le cas d'une électrode haute tension (HT) en utilisant les réseaux de neurones. L'algorithme d'apprentissage qu'il a utilisé est celui de la backpropagation.

L'augmentation de la demande en énergie électrique nécessite l'introduction d'un matériel HT plus sûr tout en réduisant au minimum son volume. D'où l'importance de la conception de l'isolation. L'optimisation du champ électrique est devenue l'un des critères de la bonne conception de l'isolation des appareils de HT. Les techniques d'optimisation du champ nous permettent d'avoir, non seulement une distribution connue du champ, mais aussi, une distribution uniforme.

Pour l'optimisation du champ électrique, les RNA ont été appliqués avec la méthode de la backpropagation. Cette méthode a montré une apparente efficacité et applicabilité dans les problèmes de reconnaissance et d'optimisation. Comme pour tout apprentissage, le principe de l'algorithme de backpropagation est de diminuer l'écart entre la sortie désirée et celle fournie par le réseau de neurones. L'erreur qu'on doit minimiser est donnée par la formule suivante :

$$E_{rr} = \frac{\sum_{k=1}^{P_n} \sum_{p=1}^{O_n} |T_p(k) - O_p(k)|}{P_n O_n} \times 100(\%) \quad (II.5)$$

$$0 < E_{rr} < 100(\%)$$

Dans cette étude, le RNA doit apprendre la relation entre la valeur du champ électrique, le contour et la distance de l'intervalle inter-électrode. Le réseau responsable de cette première étape est nommé RN1. L'objectif de cette étude est d'avoir une configuration optimale du contour de l'électrode, d'où la nécessité d'un autre réseau de neurones chargé d'apprendre la relation entre le contour et la distance d'intervalle et le contour optimal, c'est le réseau RN2. A la fin de l'apprentissage, par l'introduction de la distribution désirée du champ électrique à l'entrée du réseau, on peut obtenir une distance d'intervalle et une configuration du contour de l'électrode. La figure II.2 illustre le principe de l'optimisation proposé avec la conception décrite ci-dessus. Cette figure présente la procédure de base de la méthode d'optimisation du champ électrique basée sur les réseaux de neurones artificiels. Il a été remarqué que cette technique est répartie en trois étapes principales.

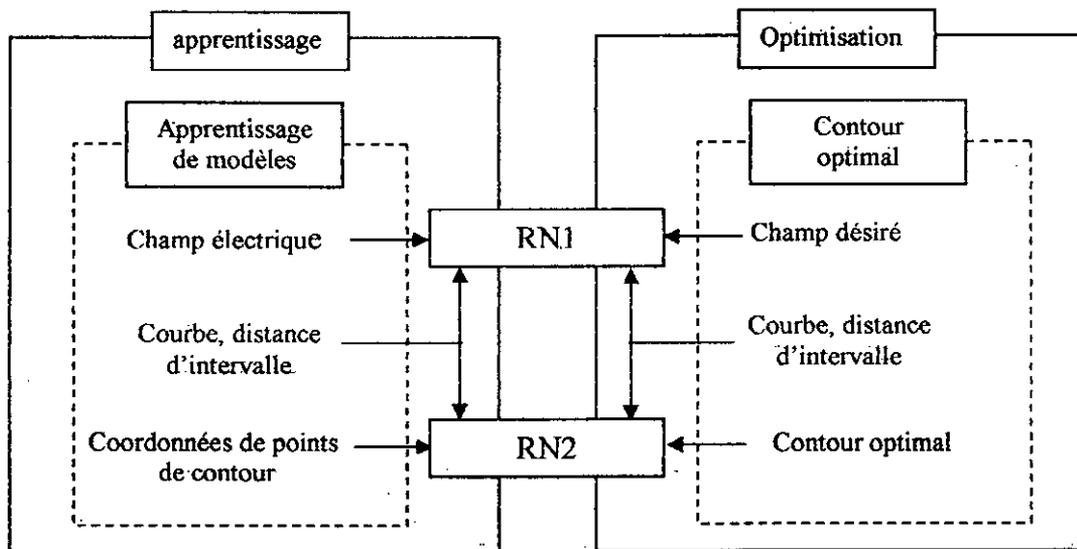


Fig.II.2. Principe de la méthode d'optimisation du champ électrique.

1. Génération de données d'apprentissage :

Dans cette première étape, la base de données de l'apprentissage est à générer. Dans un premier temps, des modèles d'apprentissage avec une simple géométrie, comme dans le cas d'un arc ou d'une droite, sont disposés dans la zone d'optimisation. On dispose d'un certain nombre de modèles d'apprentissage de ce type. Ensuite, un nombre de points de contours est disposé dans chaque modèle d'apprentissage dans un intervalle uniforme. Pour chaque point, il faut calculer le champ, la distance d'intervalle, les coordonnées, et sont par la suite utilisés comme données d'apprentissage.

2. Apprentissage du réseau de neurones :

Il faut faire l'apprentissage des deux réseaux de neurones. Le réseau RN1 apprend la relation entre le champ et le contour de l'électrode, tandis que le réseau RN2 apprend la relation entre le contour et les coordonnées de points. Une fois l'apprentissage de RN1 est terminé, la relation entre le champ et la courbe est obtenue. Après on passe à l'apprentissage de RN2.

3. Recherche du contour optimal :

En faisant introduire la distribution désirée du champ à RN1, on peut obtenir la courbe et la distance d'intervalle du contour optimal. En présentant ces deux derniers paramètres comme entrées pour RN2, on obtient la configuration du contour optimal.

L'introduction de la méthode des réseaux de neurones a prouvé une efficacité remarquable pour la technique d'optimisation du champ électrique. En plus, elle a permis d'aboutir à une distribution du champ électrique uniforme.

Dans une autre étude, S.Häppe, H.G.Kranz [26] ont développé un système de suppression en temps réel des signaux d'impulsions des DP en appliquant les RNA. Les pulsations sonores et les taux de répétition stochastique peuvent être particulièrement nuisibles et causent une mauvaise interprétation des activités des DP. Tous les systèmes de mesure des DP connus sont affectés par ce problème. Donc la suppression des pulsations non souhaitables conduit à la bonne évaluation des DP.

Le système neural utilisé à cet effet est le « Neuro TEK II ». Cette utilisation du RNA, comme une partie du processeur, permet la suppression en temps réel des formes des pulsations nuisibles pendant une large bande de mesure de DP ($f=50$ MHz). Afin de déterminer les poids des liaisons de l'algorithme du RNA utilisé, les pulsations des DP de différentes sources sont étudiées. Ces valeurs forment la base de donnée pour l'apprentissage du RNA. Dans cette phase, le RNA auto-organisé « Kohonen Feature Map » classifie les signaux inconnus (les DP et les pulsations). Pour son adaptation, on applique l'algorithme de backpropagation. Chaque échantillon de temps de résolution du signal d'entrée représente m éléments du vecteur d'entrée X . Le vecteur d'activation A représente les classes de sorties évaluées et il est fonction des entrées X . L'information dans ce cas circule uniquement dans un sens. Le calcul mathématique du vecteur sortie A est réalisé par l'algorithme donné par l'équation:

$$A = f_2 (W_2 \cdot f_1 (W_1 \cdot X + b_1) + b_2) \quad (II.6)$$

Où f_1 , f_2 sont les fonctions d'évaluation, W est la matrice poids, b_1 , b_2 sont les vecteurs de liaison.

Le signal d'entrée est attribué à l'élément du vecteur d'activation, les pulsations indésirables marquées peuvent être rejetées. Si un élément est inférieur à 50% du signal d'activation, il sera considéré comme signal inconnu et il ne peut pas être rejeté. Donc, le processus de mesure des DP est capable d'évaluer tous les signaux relevés. Après la dernière phase d'apprentissage, les poids w et les liaisons b du RNA sont appliqués au neuroprocesseur pour le temps critique de la première phase d'évaluation des pulsations de DP.

Suivant le nombre des pulsations du signal d'entrée, on peut classifier plus de 500 pulsations en 40 ms.

Le système de reconnaissance des décharges « Neuro TEK II » est constitué par les éléments suivants:

- La sonde électronique avec l'accouplement des unités.
- Des interfaces avec des raccordements transitoires.
- Le système de mesure des larges bandes digitales.
- Le système neuroprocesseur pour une première classification.

Un raccordement transitoire est placé dans le système « Neuro TEK II » afin de permettre un emplacement automatique des amplificateurs et des valeurs de déclenchement.

La base de données du RNA est générée et lue à l'intérieur du neuroprocesseur. Ensuite, le système « Neuro TEK II » peut évaluer chaque impulsion déclenchée comme mesure en temps réel.

Les résultats des mesures effectuées en laboratoire montrent la suppression des pulsations indésirables. Le « Neuro TEK II » est donc capable de reconnaître les pulsations de DP et les perturbations avec une bonne identification de distance. Il est capable de classifier plus de 25000 pulsations par seconde. Après la suppression des perturbations, seules les données de DP sont stockées dans un fichier de données normalisées.

1999 S.Chowdhury et M.S.Naidu [27], ont appliqué les réseaux de neurones dans des études de champ non uniforme pour le calcul des tensions de claquage du SF₆. Dans la pratique, le champ électrique dans le gaz est déformé par des champs non uniformes créés par les particules métalliques présentes dans le gaz et les surfaces saillies au niveau des électrodes. L'algorithme d'apprentissage est le PPLN (Projection Pursuit Learning Network).

Ce genre de réseau se construit au fur et à mesure que l'apprentissage se fait. L'algorithme commence par un neurone caché. En ajoutant un autre neurone, l'apprentissage ne se fait pas pour tout le réseau, mais seulement pour le nouveau neurone. Ce réseau a de bonnes capacités de généralisation.

Pour les différentes formes d'ondes de tensions appliquées au réseau et en variant les paramètres expérimentaux, les caractéristiques obtenues par l'application du RNA sont semblables à celles obtenues expérimentalement avec une erreur maximale de 5%.

Dans le domaine des décharges partielles, Y.Tian [28] a utilisé les RNA pour l'identification des DP à partir de la mesure des émissions acoustiques (EA). EA est basée sur la détection des ondes produites par la déformation brusque dans le matériel en question. Ces ondes se transmettent de la source vers les détecteurs qui les convertissent en signaux électriques. Cette approche n'est pas touchée par le phénomène d'interférence électrique. En plus de définir l'emplacement de la décharge, elle peut déterminer le type de DP dans les câbles de haute tension.

Une étude préliminaire sur l'identification de deux modèles de DP est présentée, décharges internes et décharges superficielles, en utilisant trois configurations de RNA ; la rétro-propagation, le modèle de Kohonen Self-Organising Map (SORG) et le Learning Vector Quantization (LVQ). Les trois modèles sont constitués de 3 couches, une couche d'entrée, une couche cachée et une couche de sortie. Le processus d'identification nécessite deux étapes, l'apprentissage puis d'identification. Différents types d'entrées sont utilisés comme signaux d'entrées des réseaux de neurones, il s'agit du domaine du temps, spectre du signal, caractéristiques $\varphi-q-n$ (angle produit par la décharge-quantité de décharge-taux d'impulsion) et les opérateurs statistiques. Pour les deux derniers cas, on peut utiliser l'amplitude pour décrire le signal de l'EA, mais leur utilisation serait très complexe car les ondes acoustiques s'atténuent pendant leur processus de propagation.

En considérant que les différents modèles de DP sont liés à différentes caractéristiques de signaux et de spectres, il est possible de les identifier en utilisant les signaux de l'émission acoustique et leur spectre de fréquence comme informations d'apprentissage des RNA. Les taux d'identification pour chaque réseau utilisant comme entrée les données relatives au temps et au spectre de fréquences ont été déterminées. Les exemples d'apprentissage contiennent 10 ensembles de signaux de l'EA pour les décharges internes, et 10 ensembles de signaux pour les décharges superficielles.

La couche cachée du réseau utilisant la rétro-propagation est constituée de 10 éléments d'apprentissage, dans le cas du réseau utilisant LVQ nous avons 4 éléments d'apprentissage et 2 pour le réseau SORG. La comparaison des taux d'identifications, obtenus par les RNA entraînés par les signaux ou les spectres, est donnée par le tableau II.2.

Types d'entrées	BP %	LVQ %	SORG %
Entrées: signal	78	80	60
Entrées: spectre	95	100	100

Tableau.II.2. Taux d'identification avec différents types d'entrées du réseau de neurones

Le tableau II.3. représente l'influence des taux d'identification en fonction des différents nombres de points d'apprentissage.

Nombre de points d'apprentissage	BP %	LVQ %	SORG %
20000 points	78	80	60
10000 points	83	75	60
5000 points	70	75	60
2000 points	95	88	63

Tableau.II.3. Taux d'identification avec différents nombres de points d'apprentissage du réseau de neurones

La variation du taux d'identification en fonction du nombre de neurones cachés du réseau entraîné par la rétro-propagation est présentée dans le tableau.II.4.

Neurones cachés	6	10	20	30	40	50	60
Taux (%)	98	95	98	85	90	100	93

Tableau.II.4. Taux d'identification avec différents nombres de points d'apprentissage du réseau de neurones

L'influence du critère de convergence de l'algorithme d'apprentissage de la rétro-propagation sur le taux de l'identification est présentée dans le tableau.II.5.

Critère de convergence	0.01	0.05	0.1	1	4
Taux (%)	95	95	95	95	98

Tableau.II.5. Taux d'identification en fonction du critère de convergence pour l'algorithme de la rétro-propagation

On refait le même travail en utilisant l'algorithme d'apprentissage du LVQ. Les résultats sont donnés par le tableau.II.6.

Neurones cachés	2	4	8	14	20	30	40
Taux (%)	85	83	83	85	85	85	83

Tableau.II.6. Taux d'identification en fonction du nombre de neurones cachés pour l'algorithme LVQ

L'influence du nombre de cycles d'apprentissage sur le taux d'identification des deux derniers algorithmes d'apprentissage (LVQ et SORG) est présentée par les tableaux II.7. et II.8.

Cycles d'apprentissage	200	400	1000	2000	3000	4000
Taux (%)	80	83	83	83	83	83

Tableau.II.7. Taux d'identification en fonction du nombre des cycles d'apprentissage pour l'algorithme LVQ

Cycles d'apprentissage	200	400	1000	2000	3000	4000
Taux (%)	75	83	80	75	75	75

Tableau.II.8. Taux d'identification en fonction du nombre des cycles d'apprentissage pour l'algorithme SORGM

A la fin de cette étude, une comparaison entre quatre procédures d'apprentissage a été présentée, la première avec 10 ensembles de signaux d'EA (5 pour les décharges internes et 5 pour les décharges superficielles), la seconde avec 20 ensembles de signaux d'EA (10 pour les décharges internes et 10 pour les décharges superficielles), puis avec 30 et 40 ensembles de signaux d'EA. Les résultats sont présentés dans le tableau.II.9.

Nombres d'ensembles d'apprentissage	BP %	LVQ %	SORG %
10 ensembles d'apprentissage	95	70	80
20 ensembles d'apprentissage	95	83	83
30 ensembles d'apprentissage	95	84	84
40 ensembles d'apprentissage	95	85	83

Tableau.II.9. Taux d'identification avec différents ensembles d'entrée d'apprentissage du réseau de neurones

Dans le cas des trois méthodes utilisées, les résultats montrent l'efficacité de l'utilisation du spectre d'énergie pour distinguer entre les différents types de décharges partielles. Il a été constaté que le taux d'identification est lié au nombre de neurones dans la couche cachée du réseau, au nombre de cycles d'apprentissage et à la valeur du critère de convergence. Il a été constaté aussi que le temps d'apprentissage dépend du type du réseau de neurones utilisé, de l'algorithme et du critère d'arrêt.

Cette étude a montré l'efficacité de l'utilisation des réseaux de neurones artificiels pour l'identification des DP, en particulier les méthodes LVQ et SORG.

Dans la même année, un autre rapport, de I.Nunes Da Silva [29], utilise les réseaux de neurones pour la détermination de l'intensité du flux électrique et de la valeur critique de la tension disruptive dans les sous stations, en tenant compte de plusieurs facteurs atmosphériques et géométriques, comme la pression, la température, l'humidité, distance entre phases, hauteurs des jeux de barres et la forme d'onde.

Des exemples de simulation et de test sont présentés pour valider l'approche proposée. Les résultats obtenus expérimentalement et numériquement permettent de proposer de nouvelles règles pour la spécification des sous stations.

Cette étude a trois objectifs. Le premier est de prouver encore une fois que les RNA peuvent traiter les problèmes d'identification des systèmes HT. Le second objectif est de trouver une méthode perspective pour le dimensionnement des sous stations de 138 kV. Le troisième quant à lui, est de présenter une analyse comparative entre l'approche proposée et les techniques conventionnelles ou classiques.

Déjà dans les années 60 et 70, on s'est intéressé à l'étude de la distance entre les barres de conduction, en utilisant les tests d'impulsions atmosphériques, sous des conditions sèches ou humides. A la fin des années 80, d'autres études ont été réalisées sur la conversion de la tension nominale des sous stations à des niveaux plus élevés. Les raisons de ces changements sont dues à l'augmentation de la charge, à l'obtention de systèmes plus économiques et à la réduction de la surface occupée par ces sous stations.

On cherche toujours la spécification des distances entre les jeux de barres afin d'optimiser le volume occupé par les substances

Dans le laboratoire de l'université de Sao Paulo, la tension critique disruptive ($U_{50\%}$) a été déterminée expérimentalement pour plusieurs distances entre jeux de barres. Nous avons $U_{10\%} = U_{50\%}(1 - 1.3\sigma)$ où σ est l'écart type (3%). Les conditions atmosphériques réelles (température, pression, humidité) liées à chaque valeur de tension ont été déterminées pour être utilisées pour l'apprentissage des RNA.

Le choix des RNA dans cette étude est dû à leur rapidité de l'analyse ou le dimensionnement des traitements essentiels. Dans ce travail, les RNA sont utilisés pour

représenter les relations entre les variables associées avec le processus de spécification des sous stations.

La figure ci-dessous montre le montage des jeux de barres en configuration parallèle.

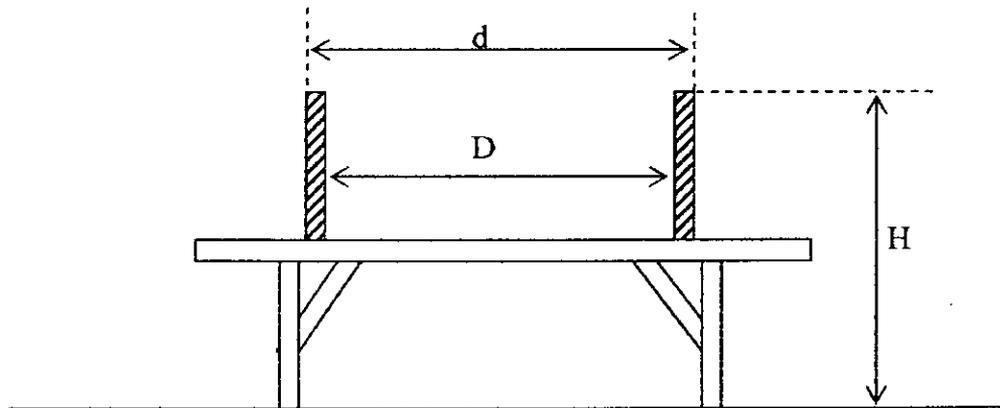


Fig. II.3. Configuration expérimentale

Où :

d : la distance entre phases,

D : la distance de l'arc,

H: la hauteur des jeux de barres par rapport à la terre.

Le processus d'identification se compose de deux étapes : la sélection des modèles et l'estimation des paramètres. Le RNA utilisé dans cette étude est le réseau multicouche statique et l'algorithme d'apprentissage utilisé pour calculer les poids du réseau correspond à l'estimation des paramètres. Il faut noter que l'apprentissage est un problème d'optimisation typique. La structure de base contient une seule couche cachée, elle apparaît assez puissante pour faire une organisation arbitraire entre les variables.

L'entraînement du RNA se fait par des présentations successives des exemples de données de type entrée/sortie. Pendant l'apprentissage, les données sont propagées en avance dans le réseau. Ce dernier va ajuster ses poids internes pour minimiser la fonction coût (l'erreur entre la sortie produite par le réseau et la sortie désirée) en utilisant l'algorithme de la rétro-propagation. La fonction à minimiser est :

$$E_{moy} = \frac{1}{N} \sum_{k=1}^N E(k) \quad (II.7)$$

Où :

N est le nombre de points d'apprentissage,

E(k) est la somme des erreurs dans tous les neurones de sortie correspondant au k^{ème} exemple.

Pour une configuration optimale des poids, E(k) est minimisée en fonction de la variation des poids, d'où il faut vérifier que :

$$\frac{\partial E(k)}{\partial w_{ij}} = 0 \quad (II.8)$$

Où :

w_{ij} est le poids de la connexion reliant le neurone i de la couche L au neurone j de la couche $(L-1)$.

Finalement, les poids du réseau sont adaptés en utilisant la relation suivante:

$$w_{ij}^L(k+1) = w_{ij}^L(k) - \eta \frac{\partial E(k)}{\partial w_{ij}} \quad (\text{II.10})$$

Où :

η est une constante qui détermine le taux d'apprentissage de l'algorithme de la rétro-propagation.

L'architecture générale du réseau de neurones utilisé est présentée par la figure ci-dessous.

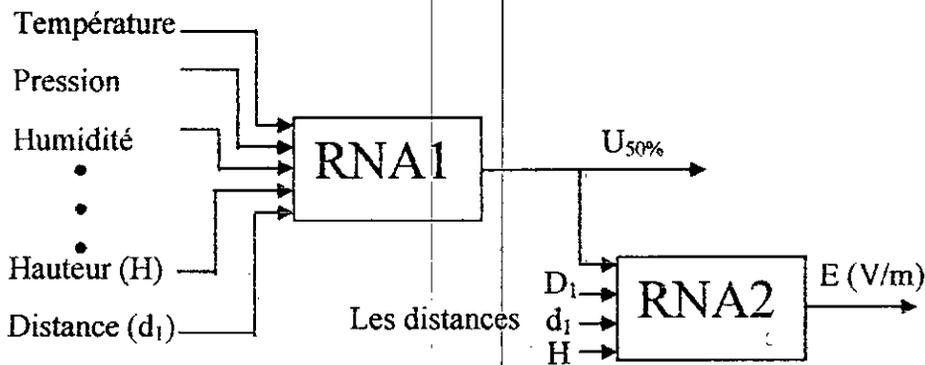


Fig.II.4. Architecture du réseau utilisé

Cette architecture est composée de deux réseaux de neurones.

1. Le premier réseau calcule la tension disruptive critique ($U_{50\%}$). Les données d'apprentissage du RNA1 sont obtenues directement des valeurs expérimentales mesurées au laboratoire. On remarque que ce réseau a pris en compte plusieurs facteurs atmosphériques et géométriques.

2. Le deuxième réseau calcule l'intensité du champ électrique entre les jeux de barres. Pour ce réseau, les données d'entraînement sont obtenues par la méthode des éléments finis. On remarque dans celui-ci que la sortie du RNA1 est utilisée comme un paramètre d'entrée pour le RNA2, en plus, bien sûr, des paramètres atmosphériques et géométriques qui sont eux aussi pris en compte pour le calcul de l'intensité du champ électrique.

Afin d'exploiter les performances de la méthode proposée, deux configurations ont été proposées, la première considère des jeux de barres avec $d = 1,75\text{cm}$, $H = 4\text{m}$, sous une tension de 800 kV. La deuxième considère $d = 2\text{m}$, $H = 4\text{m}$ sous une tension de 1000 kV. Les valeurs du champ électrique, obtenues pour les deux configurations sont respectivement 9,7.105 V/m pour la première configuration et de 1,19.105 V/m pour la seconde. Ces deux résultats n'ont pas dépassé la valeur maximale $3 \cdot 10^6$ V/m admissible dans l'air, donc y'a pas de risque d'avoir des décharges électriques.

Cette étude a présenté une nouvelle méthode pour la conception des sous stations en utilisant les réseaux de neurones artificiels. Les résultats de simulation peuvent être utiles dans l'élaboration de nouveaux critères plus adéquats pour la conception des sous stations.

Les réseaux de neurones sont utilisés dans l'identification dans le domaine de la HT. Leur apprentissage est réalisé en utilisant les données expérimentales. Une fois l'apprentissage effectué, le réseau est capable de généraliser ce qu'il a appris pour de nouvelles entrées inconnues, qui n'ont pu être effectuées expérimentalement, ce qui permet d'éviter de les faire au laboratoire et donc réduire le temps des expérimentations et leurs coût.

2000 À la conférence internationale IEEE sur les Isolations Electriques et les Phénomènes de Décharges (CEIDP'2000), X.H.Jin [30] a présenté un article sur l'utilisation des réseaux de neurones dans la haute tension pour classer les défauts provoqués par les décharges partielles ayant lieu dans l'isolation d'un transformateur. Les deux réseaux de neurones utilisés sont la backpropagation et le LVQ. Ils ont tous les deux prouvé leur capacité de généralisation.

A l'Ecole Nationale Polytechnique d'Alger (ENP), il y'a eu, depuis quelques années, des travaux effectués sur l'utilisation des réseaux de neurones artificiels dans le domaine de la haute tension.

2001 A.El Hadj [17] dans sa thèse de magister sous la direction de A.Mekhaldi et M.Teguar, a présenté un travail sur l'utilisation des RNA dans le domaine de la pollution des isolateurs. En se basant sur le pas d'apprentissage d'une part et du domaine d'apprentissage d'autre part, il a déterminé les paramètres qui conduisent à une bonne prédiction. Pour montrer l'efficacité de ses programmes, les résultats ont été comparés à des courbes réelles.

Cette étude est liée à quelques caractéristiques décrivant le comportement d'un isolateur sous pollution, il s'agit de la variation du courant de fuite en fonction de la conductivité de la couche de pollution, du rayon de l'électrode circulaire et la variation de la longueur d'arc électrique en fonction de la tension appliquée.

Les réseaux de neurones ont montré leurs limites dans le cas des propriétés dont la forme est convexe. Pour éviter ce problème, il a été considéré au en premier la prédiction de la fonction inverse de la propriété. Autrement dit, au lieu de prédire Y en fonction de X, on fait l'inverse c'est à dire X en fonction de Y puis à la fin on présentera la forme réelle de la propriété Y en fonction de X. C'est le cas pour les courbes de la variation du courant de fuite en fonction de la conductivité de la couche de pollution et la variation du courant de fuite en fonction du rayon d'électrode circulaire. Pour la variation de la longueur d'arc électrique en fonction de la tension appliquée le problème des fonctions convexes ne se pose pas.

Le réseau de neurones utilisé dans cette étude est le RBF. Trois méthodes différentes ont été élaborées. La première est la RBF avec une entrée et une sortie de natures différentes (X_i, Y_i). La deuxième est la RBF avec une entrée et une sortie de même nature (Y_i, Y_{i+1}). La troisième méthode est la RBF avec deux entrées et une sortie (X_i, Y_i, Y_{i+1}).

L'algorithme utilisé dans les deux méthodes est la méthode d'optimisation aléatoire (ROM).

Il a été constaté que la première méthode est meilleure que la seconde mais la troisième méthode présente la meilleure qualité de prédiction par rapport aux deux autres.

2004 Toujours à l'ENP, en, L.Mokhnache [31] présenta sa thèse de doctorat, sous la direction de A.Boubakeur, sur l'application des réseaux de neurones dans le diagnostic et la prédiction des isolations de haute tension.

Dans un premier temps, elle a présenté les mécanismes diélectriques dans les matériaux isolants solides, liquides et gazeux. Respectivement, les isolants choisis pour étude sont le PVC utilisé dans les câbles, la BORAK22, huile de transformateur utilisée par

SONELGAZ, et les intervalles d'air pointe-barrière-plan. Ces matériaux ont fait l'objet de travaux de recherche au laboratoire de haute tension de l'ENP et dont les résultats constituent une base de données pour les réseaux de neurones.

Ce travail décrit les caractéristiques de vieillissement thermique du PVC ; les caractéristiques de vieillissement et propriétés de la BORAK22 ; et enfin, les caractéristiques de la tenue diélectrique dans les systèmes pointe-barrière-plan avec différentes configurations (distance inter-électrodes, distance pointe-barrière, largeur de la barrière, matériau de la barrière,...).

Dans un deuxième temps, elle a présenté les différentes applications des réseaux de neurones dans le diagnostic et la classification de l'huile de transformateur, la BORK22. un réseau de neurones multicouches fut utilisé, tel que la couche d'entrée est composée de six neurones, la(les) couche(s) cachée(s) contient(nent) un nombre variable de neurones et la sortie a quatre neurones. La sortie produit des nombres réels entre 0 et 1 indiquant la probabilité de chaque décision qui doit être prise après l'analyse des tests.

Plusieurs algorithmes ont été appliqués ; la backpropagation (BP) avec ses différentes variantes : BP avec momentum, BP avec momentum et adaptation du taux d'apprentissage (MADPTA), méthode de Levenberg-Marquardt (L.M) ; la Régularisation bayésienne (RB).

Les paramètres qui influent sur l'apprentissage des réseaux de neurones du diagnostic sont le nombre de couches cachées, le nombre de neurones dans la couche cachée, le taux d'apprentissage, le seuil de l'erreur et la fonction d'activation. Mis à part ces paramètres, l'étude met en évidence l'influence de la base de données.

En plus des différents algorithmes, différentes architectures ont été utilisées dont les réseaux compétitifs, les cartes auto-organisées de Kohonen et les réseaux bayésiens.

Cette étude a encore une fois démontré que les réseaux de neurones supervisés et non supervisés sont de bons systèmes pour le diagnostic et la classification. De plus, les réseaux non supervisés ont montré une grande souplesse dans la classification et l'apprentissage de ces réseaux ne présente pas beaucoup de complexité comme dans certains réseaux supervisés. Quant aux réseaux de Levenberg-Marquardt et les réseaux bayésiens, ils présentent une bonne robustesse pour le diagnostic. Cette étude a montré aussi, que l'architecture du réseau de neurones (nombre de couches cachées, nombre de neurones) est un paramètre important décidant de la qualité de l'apprentissage plus que les paramètres d'apprentissage (taux d'apprentissage,...).

Dans un dernier temps, ce travail traite de l'application des réseaux supervisés et non supervisés dans la prédiction du vieillissement thermique de l'huile de transformateur, du PVC utilisé dans les câbles et la rigidité diélectrique des intervalles d'air pointe-barrière-plan. Les RNA supervisés utilisés sont des réseaux à base radiale de fonction gaussienne (Radial Basis Function Gaussian RBF) entraînés par trois algorithmes : la backpropagation (BP), la méthode d'optimisation aléatoire (ROM) et la ROM couplée avec la méthode de K-means clustering. Le RNA non supervisé utilisé est celui de Kohonen (Self-Organization Maps SOM), appelé aussi la carte de Kohonen. Il se compose d'une seule couche de neurones fonctionnant en mode compétitif auto-organisé.

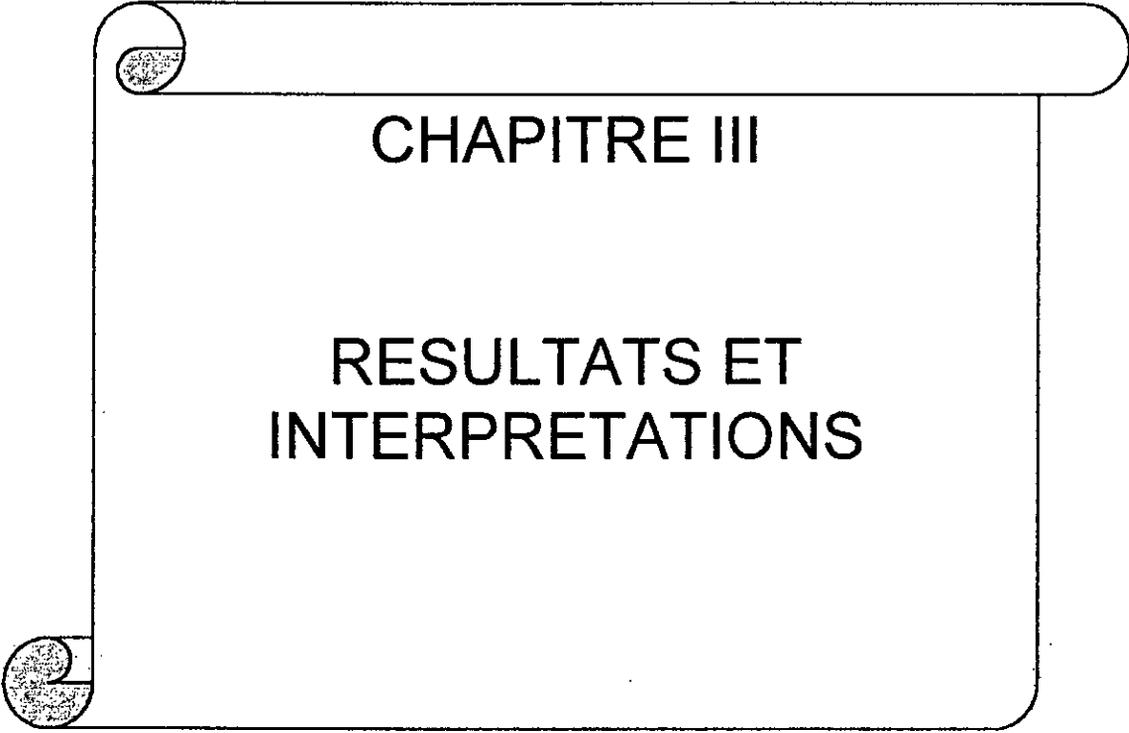
En général, ces différents réseaux ont donné des résultats satisfaisants

Pour conclure ce chapitre, on peut dire qu'une fois encore, l'aptitude des réseaux de neurones à l'estimation des fonctions non linéaires compliquées, leur permet d'être des modèles d'identification et des estimateurs dans plusieurs domaines y compris le domaine de la haute tension.

Un réseau de neurones artificiels est un système dynamique contenant des éléments non linéaires, bien interconnectés montrant une capacité formidable de calcul. Les principaux avantages d'utilisation des RNA dans le domaine de la HT sont :

- la possibilité d'apprentissage et donc de généralisation ;

- facilité d'implémentation ;
- la capacité d'organiser des systèmes complexes sans connaître les modèles mathématiques associés ;
- La possibilité de réduire le temps passé au laboratoire pour effectuer les tests et leurs coûts.



CHAPITRE III

RESULTATS ET INTERPRETATIONS

III.1. Introduction :

Dans ce chapitre nous avons utilisé une nouvelle approche pour la prédiction de certaines caractéristiques électriques par un réseau RBF. Cette approche consiste à ajouter des neurones à la couche cachée du réseau jusqu'à ce que l'erreur spécifiée soit atteinte.

Les étapes de 2 à 5 sont répétées jusqu'à atteindre ce but (fig.III.1) :

1. déclarations des différentes variables,
2. le réseau est simulé,
3. recherche de l'entrée qui correspond à l'erreur la plus élevée,
4. un neurone à fonction radiale est ajouté au réseau avec un poids égal à cette entrée,
5. réadaptation des poids pour minimiser l'erreur.

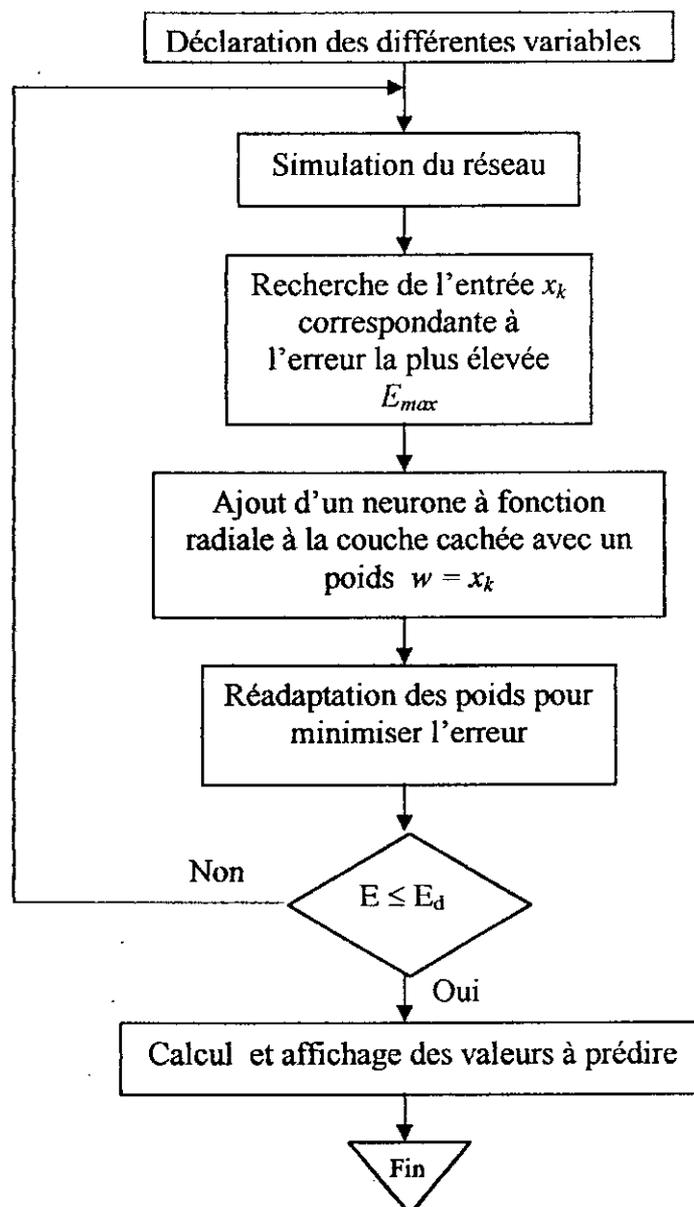


Fig.III.1. Organigramme du réseau élaboré

Nous comparerons les résultats obtenus avec un premier réseau déjà utilisé [17], qui est lui aussi un réseau RBF, ainsi qu'aux résultats obtenues à partir de résultats expérimentaux de laboratoire.

Nous ferons cette étude sur trois différentes méthodes des réseaux de neurones : RBF avec une entrée et une sortie de nature différente ; RBF avec une entrée et une sortie de même nature ; et enfin un réseau RBF avec deux entrées de nature différente.

III.2. Architecture des réseaux :

III.2.1. Méthode 1 : réseau RBF avec une entrée et une sortie de nature différente :

Dans ce réseau, nous avons une seule entrée X_i et une seule sortie correspondante Y_i . Des exemples d'entraînement $(X_1, Y_1), \dots, (X_n, Y_n)$ sont présentés au réseau. Après adaptation des paramètres, une entrée X_{n+1} est présentée au réseau qu'on considère comme valeur réelle et on refait l'apprentissage dès le début en éliminant la première valeur de la base d'apprentissage qui devient $(X_{1+f}, \dots, X_{n+f})$, tel que f est la $f^{\text{ème}}$ valeur prédite ($0 \leq f < m$), où m représente le nombre de valeurs à prédire. Ceci nous permet de garder le même nombre d'exemples d'entraînement et donc d'éviter un sur-apprentissage du réseau. On refait les mêmes étapes, jusqu'à ce qu'on arrive à injecter la dernière entrée X_{n+m} et on obtient à la sortie la dernière valeur prédite.

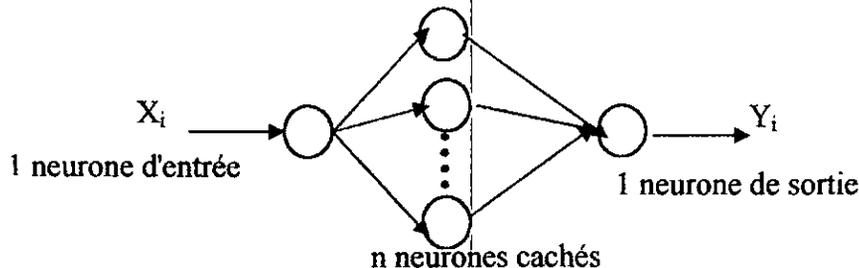


Fig.III.2. Architecture du réseau de la méthode 1

III.2.2. Méthode 2 : réseau RBF avec une entrée et une sortie de même nature :

La même procédure d'apprentissage est utilisée, sauf qu'ici au lieu d'avoir des exemples d'entraînement du type $(X_1, Y_1), \dots, (X_n, Y_n)$, nous avons des exemples d'entraînement du type $(Y_1, Y_2), \dots, (Y_{n-1}, Y_n)$. A la fin de l'apprentissage, la valeur Y_n est présentée à l'entrée du réseau, et nous obtenons à la sortie la valeur Y_{n+1} que nous considérons comme valeur exacte. Nous ajoutons, ensuite, cette valeur à la base de donnée et nous refaisons l'apprentissage dès le début. Nous reprenons ainsi les mêmes étapes jusqu'à la $(n+m-1)^{\text{ème}}$ valeur que nous injectons afin d'obtenir la dernière valeur à prédire Y_{n+m} .

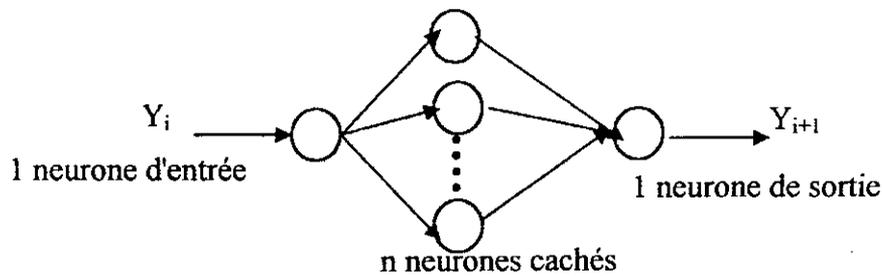


Fig.III.3. Architecture du réseau de la méthode 2

III.2.3. Méthode 3 : réseau RBF avec deux entrées et une sortie :

Dans cette méthode, deux valeurs (X_i, Y_i) sont injectées à l'entrée pour une sortie Y_{i+1} . La procédure d'apprentissage est la même que celle des méthodes précédentes. Donc, les exemples d'entraînement seront $(X_{1+f}, Y_{1+f} \text{ et } Y_{1+f}), \dots, (X_{n+f}, Y_{n+f} \text{ et } Y_{n+f})$.

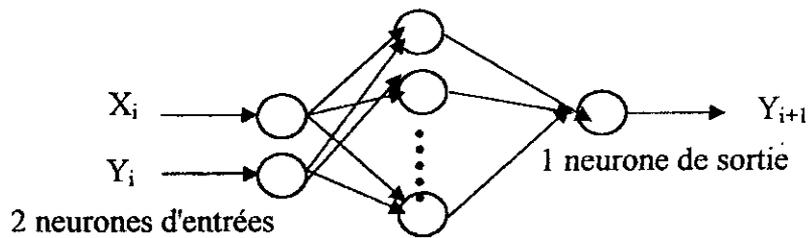


Fig.III.4. Architecture du réseau de la méthode 3

III.3. Résultats de la première méthode :

III.3.1. Variation du courant de fuite en fonction du rayon de l'électrode circulaire :

Afin de bien comparer les deux réseaux de prédiction, nous avons pris différents domaines d'apprentissage en faisant varier le nombre de points d'apprentissage et de prédiction. Les résultats obtenus par les deux réseaux sont présentés dans les figures III.5, III.6, III.7 et III.8 et sont comparées aux résultats expérimentaux.

Dans les graphes, nous appellerons les résultats obtenus par l'ancienne méthode (premier réseau) 'A-M' [17], et ceux de notre réseau, la nouvelle méthode (deuxième réseau) 'N-M'.

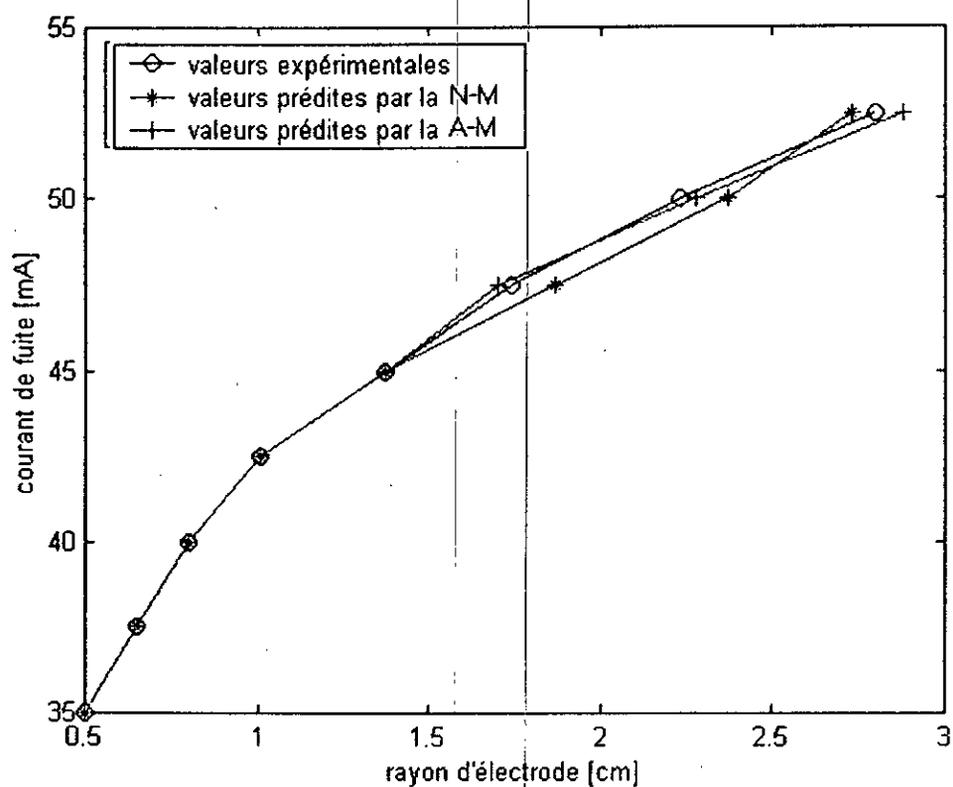


Fig.III.5. Courant de fuite en fonction du rayon de l'électrode circulaire (5 points d'apprentissage, 3 points prédits)

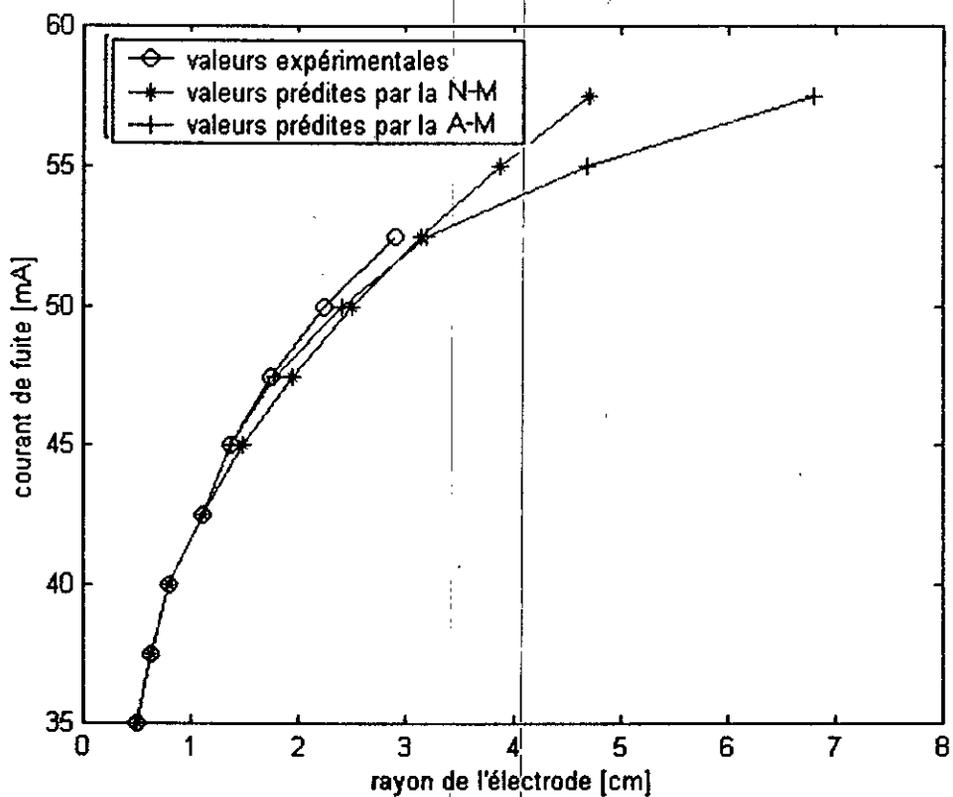


Fig.III.6. Courant de fuite en fonction du rayon de l'électrode circulaire (4 points d'apprentissage - 6 points prédits)

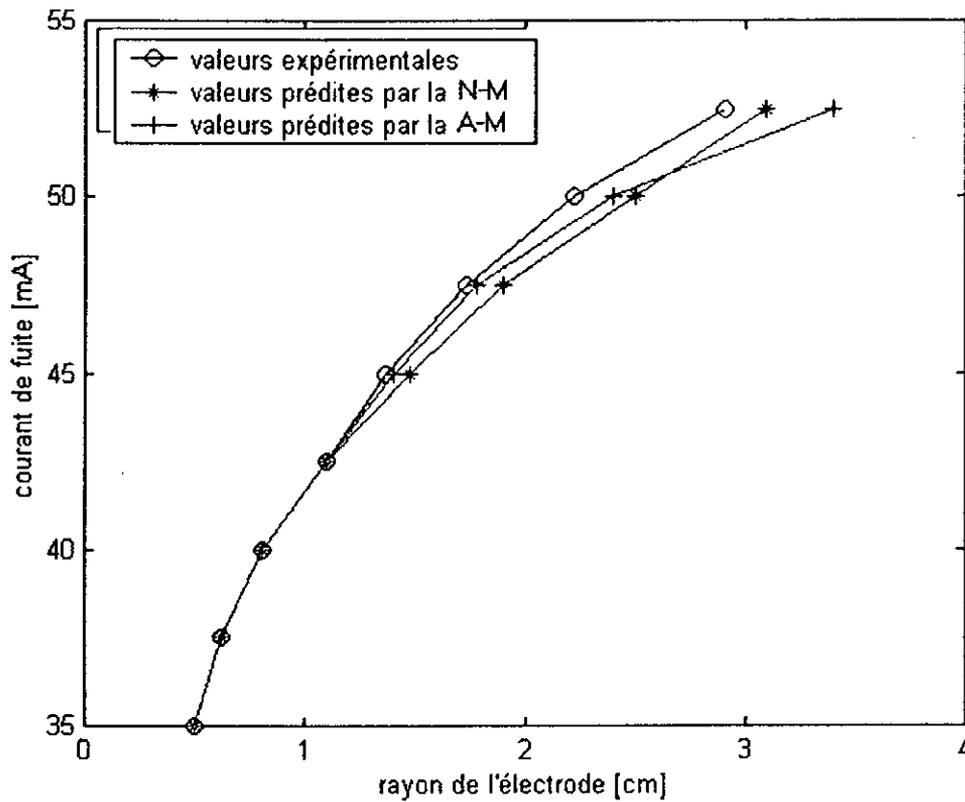


Fig.III.7. Courant de fuite en fonction du rayon de l'électrode circulaire
4 points d'apprentissage – 4 points prédits

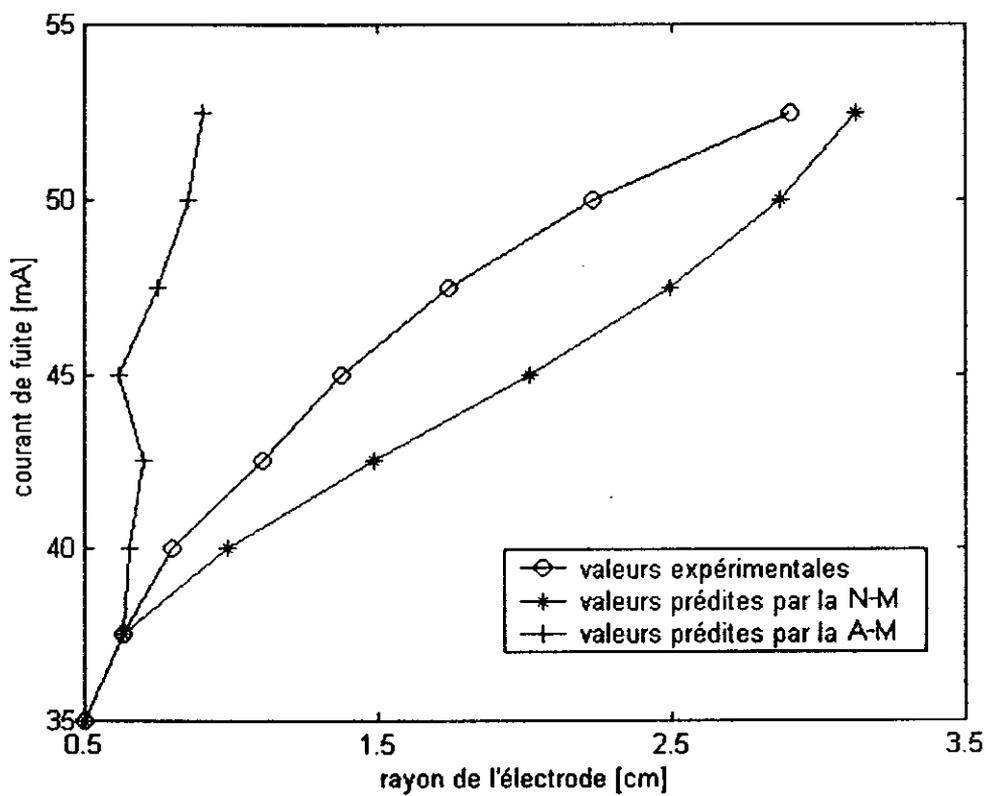


Fig.III.8. Courant de fuite en fonction du rayon de l'électrode circulaire
2 points d'apprentissage – 6 points prédits

D'après les résultats que nous avons obtenus, nous pouvons dire que la nouvelle méthode donne une bonne prédiction de cette caractéristique, même si l'ancienne méthode a montré une meilleure précision et donnée de meilleurs résultats dans les deux premières courbes où l'erreur relative est entre 2.24% et 2.85% pour la première courbe.

On constate aussi que notre réseau est moins influencé par le domaine d'apprentissage que celui de A.El Hadj. Même si la précision diminue, les résultats obtenus convergent et suivent la même allure que la courbe expérimentale, tandis que pour le premier réseau, même s'il donne dans les premières valeurs prédites une meilleure précision, il a ensuite tendance à diverger et s'éloigner de la courbe expérimentale.

III.3.2. Variation du courant de fuite en fonction de la conductivité de la couche de pollution :

Comme pour la caractéristique précédente, nous présenterons différents points d'apprentissage pour pouvoir comparer le comportement des deux réseaux, et aussi mettre en évidence l'influence de ce paramètre sur la qualité de la prédiction.

Les résultats sont présentés dans les figures III.9, III.10 et III.11 ci-dessous.

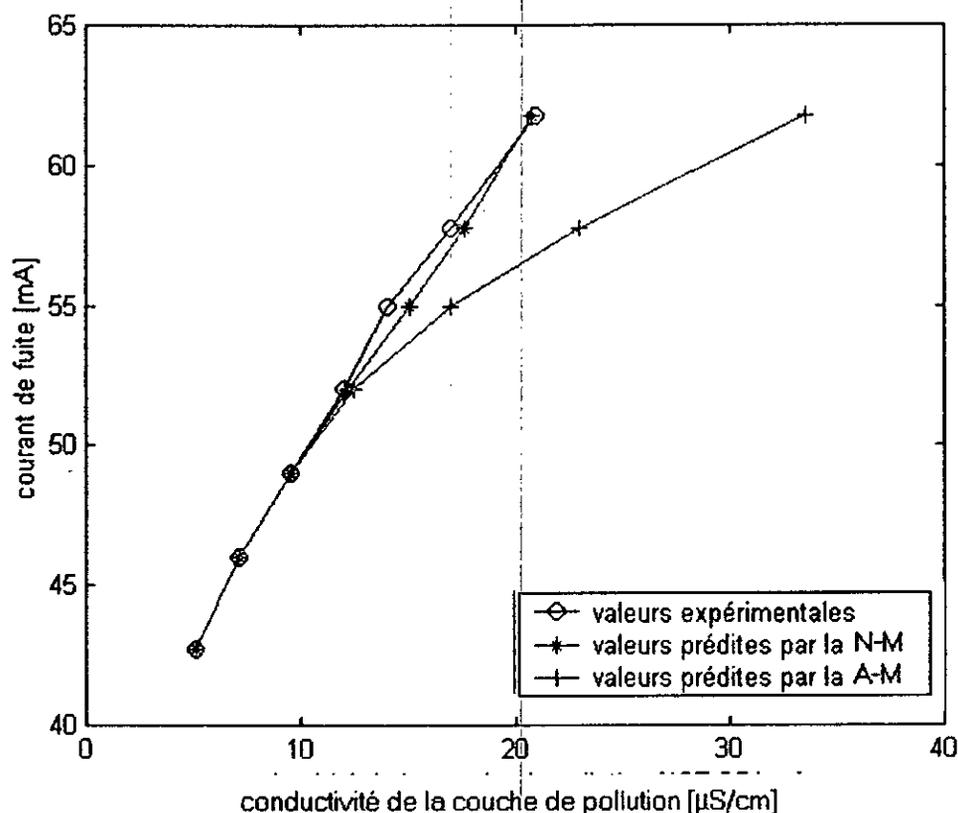


Fig.III.9. Courant de fuite en fonction de la conductivité de la couche de pollution (3 points d'apprentissage, 4 points prédits)

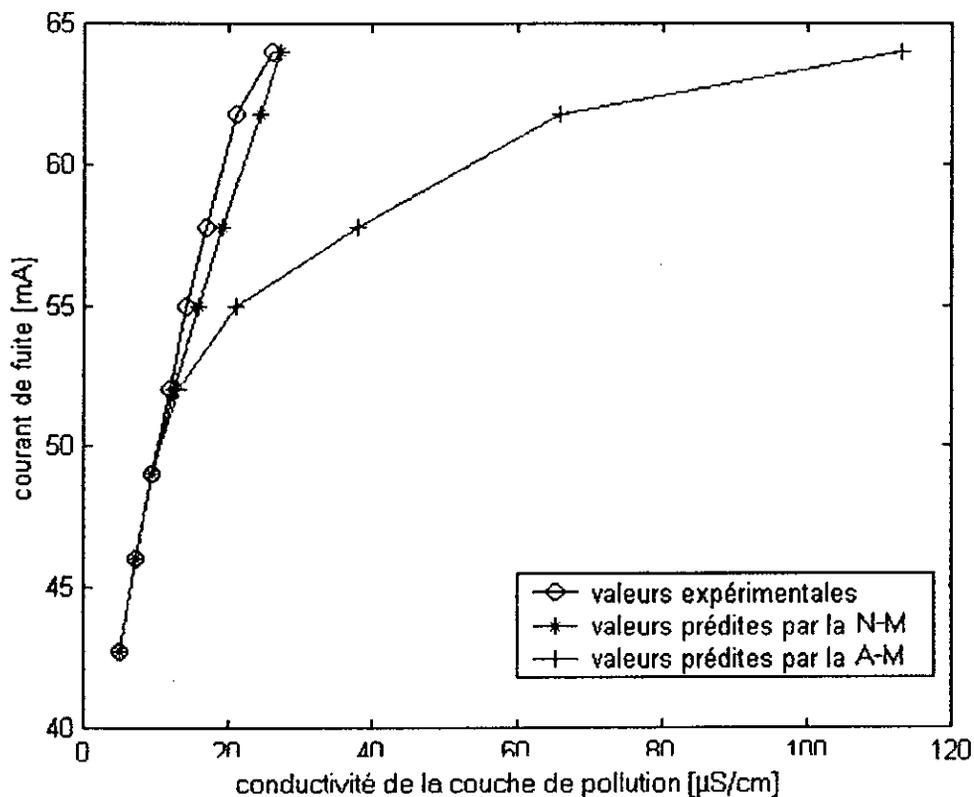


Fig.III.10. Courant de fuite en fonction de la conductivité de la couche de pollution (3 points d'apprentissage, 5 points prédits)

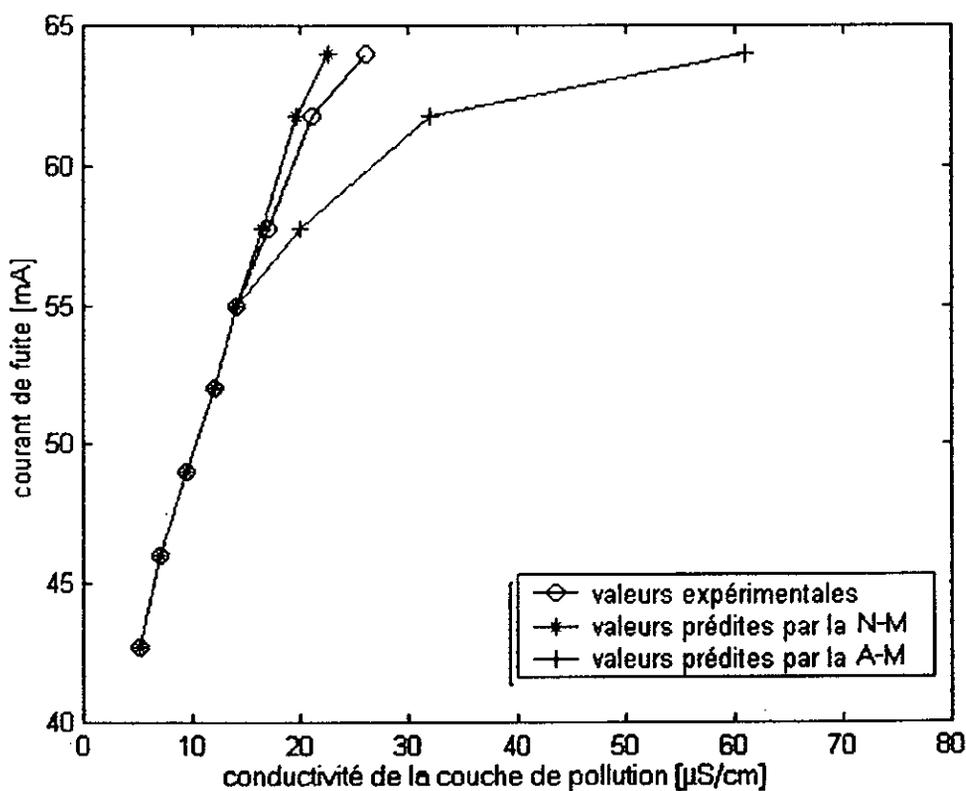


Fig.III.11. Courant de fuite en fonction de la conductivité de la couche de pollution (5 points d'apprentissage, 3 points prédits)

On remarque que dans la prédiction de cette caractéristique, le deuxième réseau est nettement meilleur que le premier, et montre une meilleure convergence alors que l'autre réseau diverge dès la première valeur prédite et donne une mauvaise prédiction.

De plus, on remarque que le domaine d'apprentissage n'a pas d'influence sur la qualité de prédiction des deux réseaux.

III.3.3. Variation de la longueur d'arc électrique en fonction de la tension appliquée :

Dans cette caractéristique nous remarquons, comme dans la précédente, que la prédiction est mauvaise dans le cas du premier réseau, tandis que le second donne de bons résultats.

Ceci met déjà en évidence les limites du premier réseau pour la prédiction de certaines caractéristiques, tandis que le nouveau réseau a jusqu'à présent fait preuve d'un bon comportement qui se traduit par la convergence de sa courbe même si la précision n'est pas toujours très bonne.

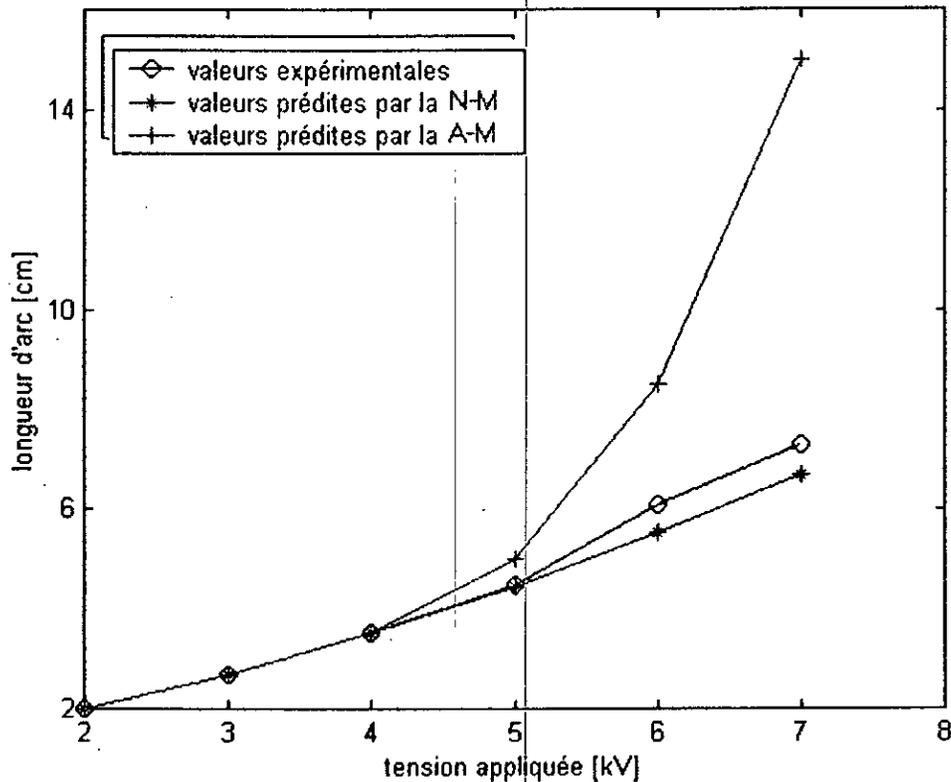


Fig.III.12. Variation de la longueur d'arc en fonction de la tension appliquée (3 points d'apprentissage – 3 points prédits)

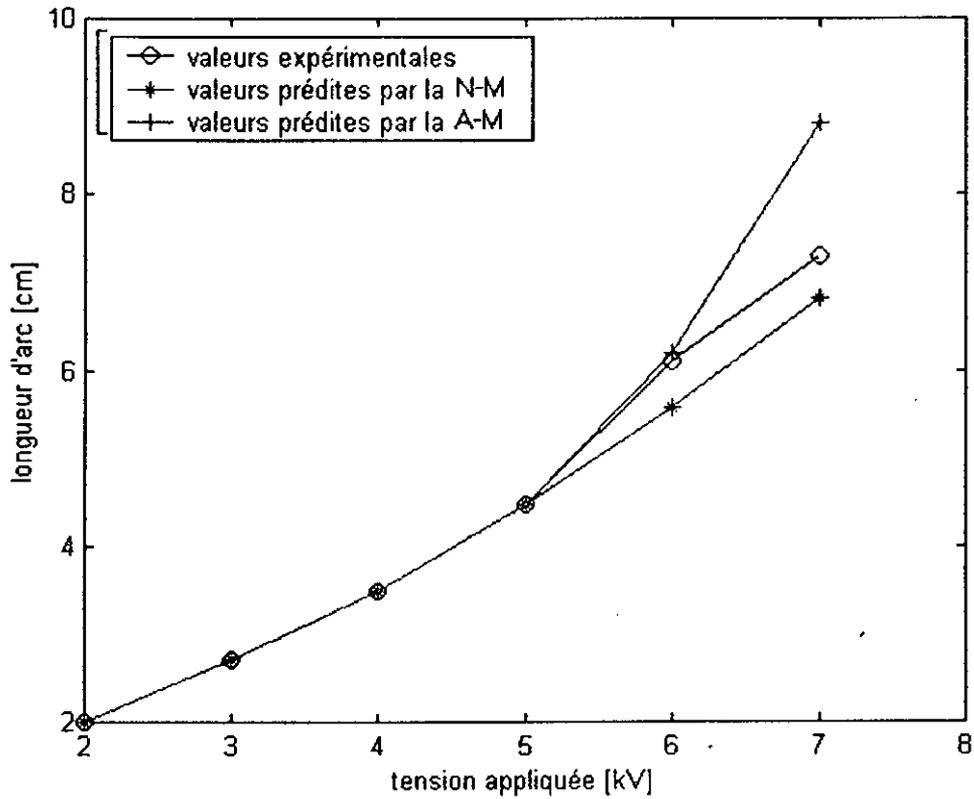


Fig.III.13. Variation de la longueur d'arc en fonction de la tension appliquée (4points d'apprentissage – 2 points prédits)

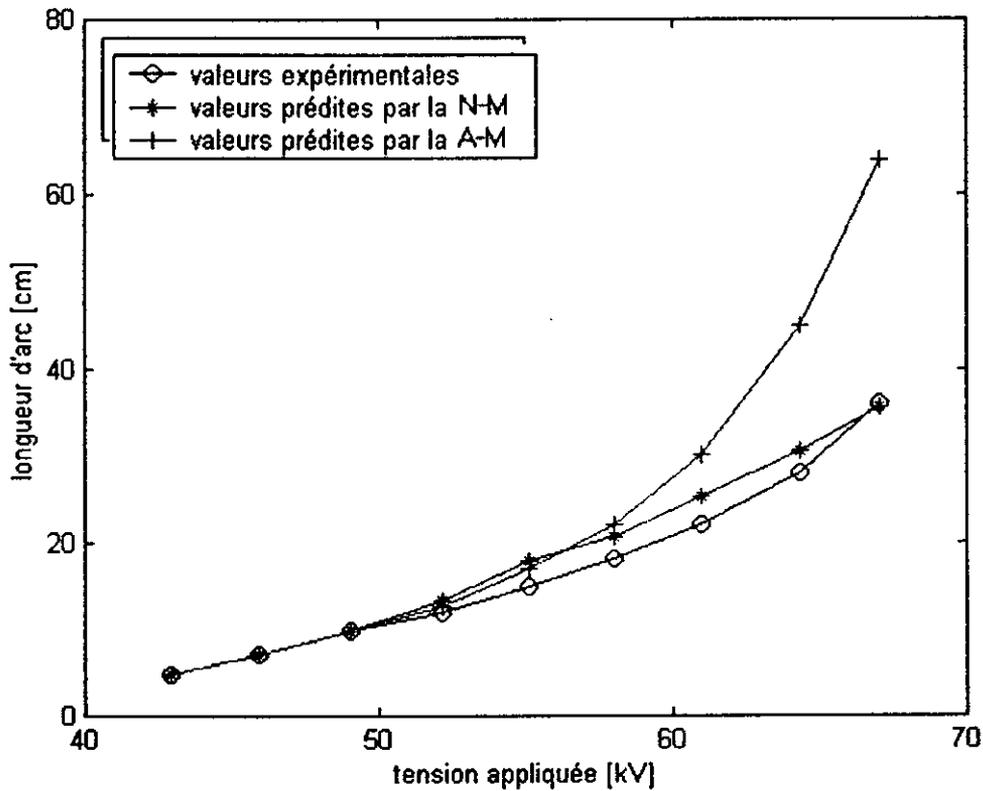


Fig.III.14. Variation de la longueur d'arc en fonction de la tension appliquée (3points d'apprentissage – 6 points prédits)

III.3.4. Variation de la tension de contournement en fonction de la longueur de la couche de pollution :

Nous avons utilisé deux domaines d'apprentissage pour cette caractéristique. D'abord en utilisant quatre points d'apprentissage pour prédire quatre points, ensuite nous avons utilisé seulement trois points d'apprentissage afin de prédire cinq points.

Les résultats obtenus pour ces deux domaines d'apprentissage sont montrés dans les courbes ci-dessous.

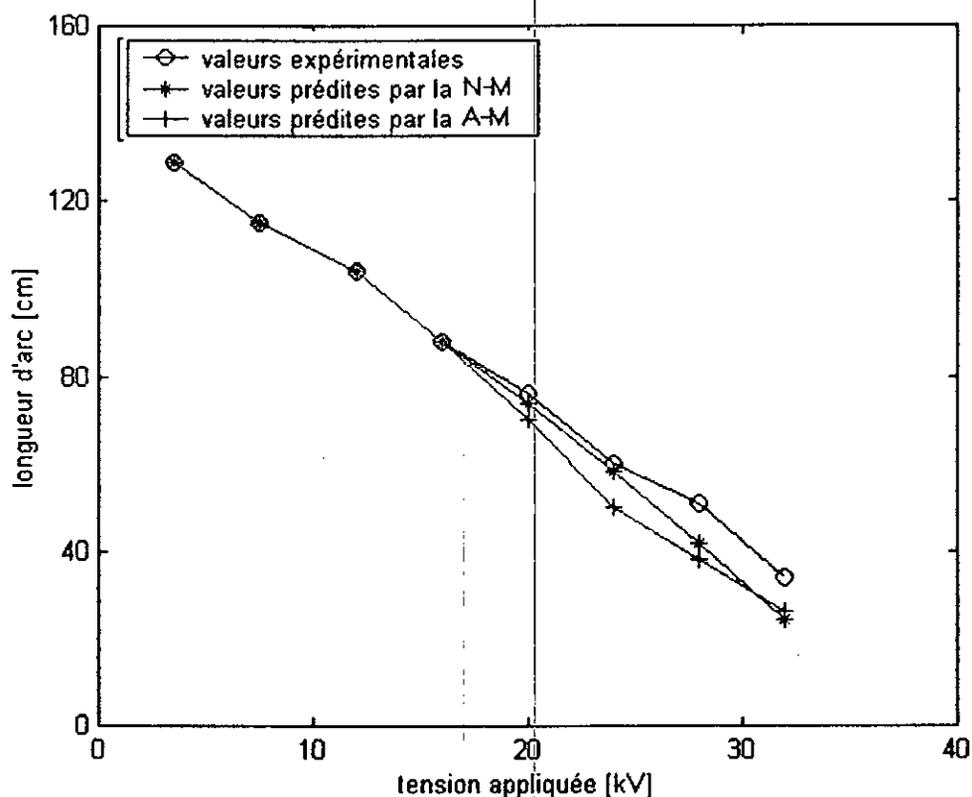


Fig.III.15. Variation de la tension de contournement en fonction de la longueur de la couche de pollution.
(4 points d'apprentissage – 4 points prédits)

Pour cette première courbe, on constate que les deux méthodes font une bonne prédiction avec une bonne précision. On voit que la courbe du premier réseau diverge, ou encore, s'éloigne de la courbe expérimentale mais ensuite elle converge de nouveau.

Le nouveau réseau quant à lui, sa courbe ne s'éloigne de la courbe expérimentale que pour les dernières valeurs prédites.

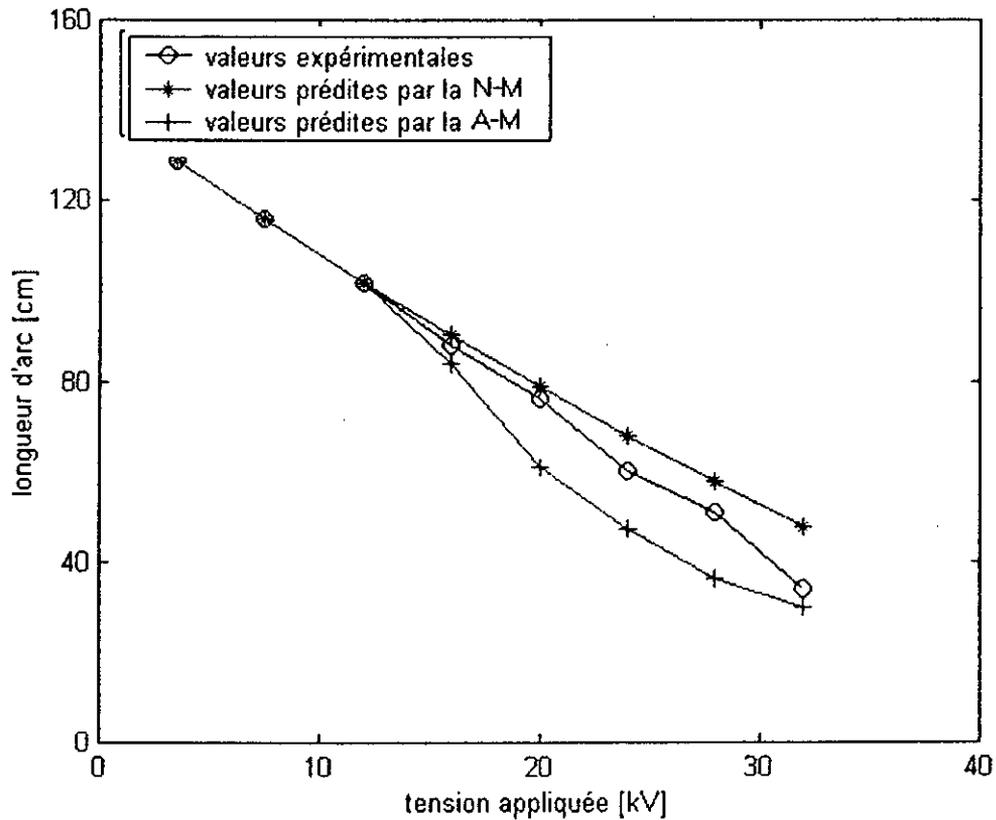


Fig.III.16. Variation de la tension de contournement en fonction de la longueur de la couche de pollution.
(3 points d'apprentissage – 5 points prédits)

On constate dans le cas de cette courbe, que les deux courbes prédites ont le même comportement que dans le cas précédent et effectuent une bonne approximation sachant que le domaine d'apprentissage se limite seulement à trois exemples.

III.4. Résultats de la deuxième méthode :

III.4.1. Variation du courant de fuite en fonction du rayon de l'électrode circulaire :

Afin d'illustrer le comportement des deux réseaux élaborés en utilisant la deuxième méthode, nous avons utilisé trois domaines d'apprentissage différents.

On constate que quelque soit le domaine d'apprentissage utilisé, l'ancien réseau donne de mauvais résultats et diverge très vite, même si le nombre de point d'apprentissage est important comme dans la seconde courbe où nous avons utilisé six points d'apprentissage.

Par contre, le nouveau réseau fait une bonne prédiction de la caractéristique dans les deux premiers cas, tandis que pour le troisième cas, il diverge rapidement.

On peut dire que le second réseau est meilleur pour la prédiction de cette caractéristique.

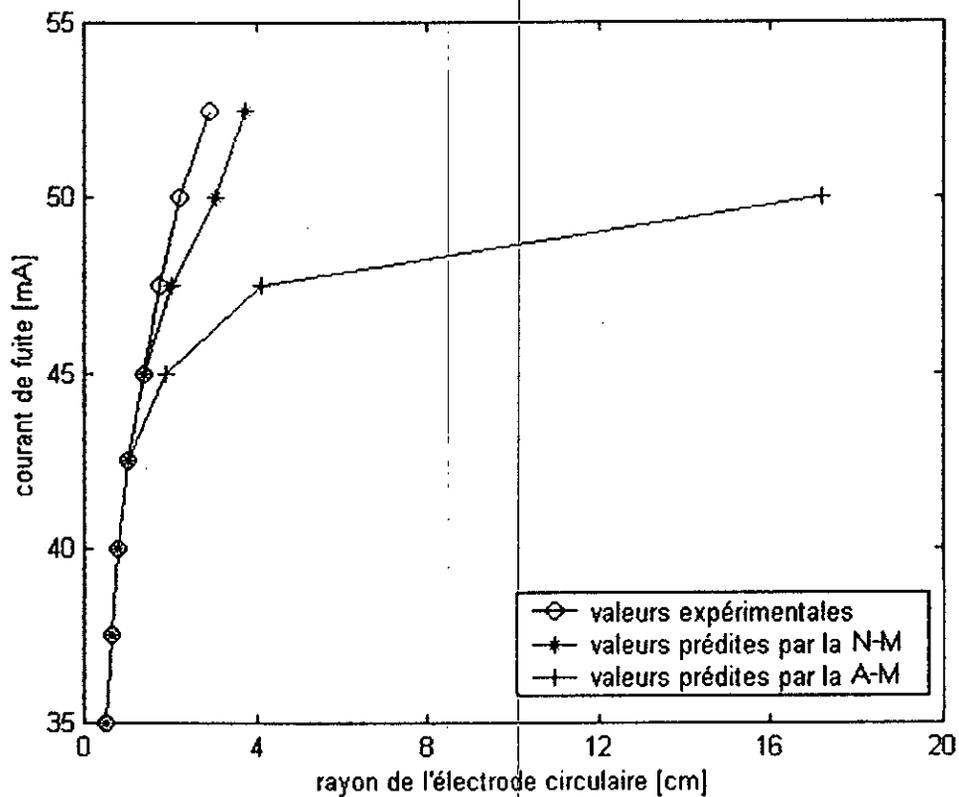


Fig.III.17. Variation du courant de fuite en fonction du rayon de l'électrode circulaire (4 points d'apprentissage - 4 points prédits)

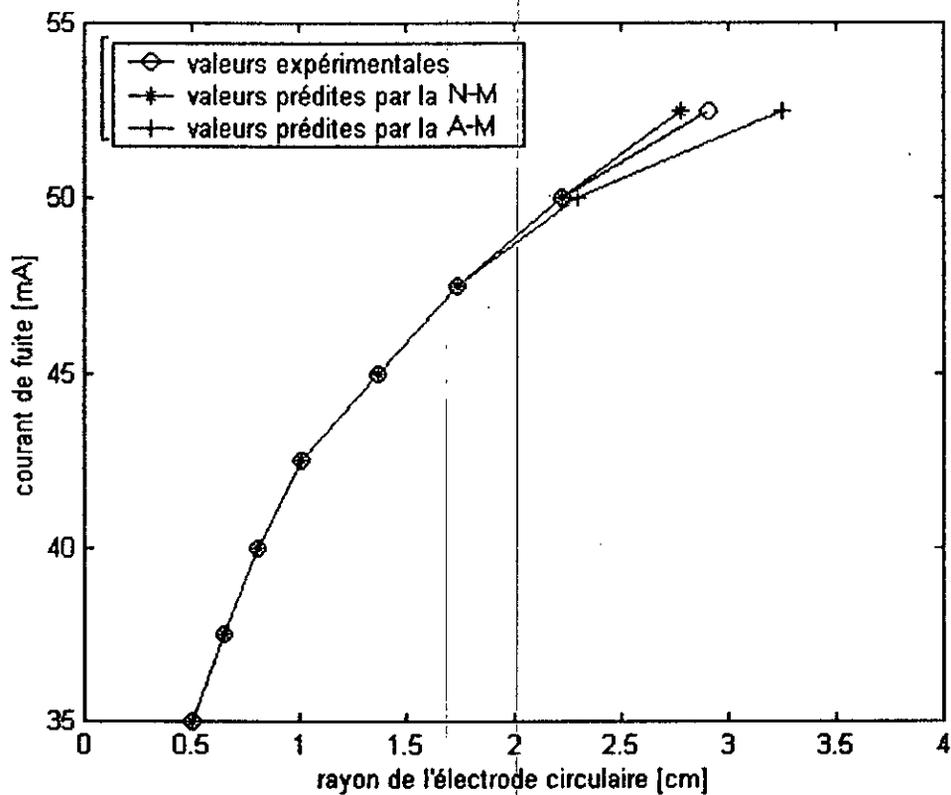


Fig.III.18. Variation du courant de fuite en fonction du rayon de l'électrode circulaire (6 points d'apprentissage - 2 points prédits)

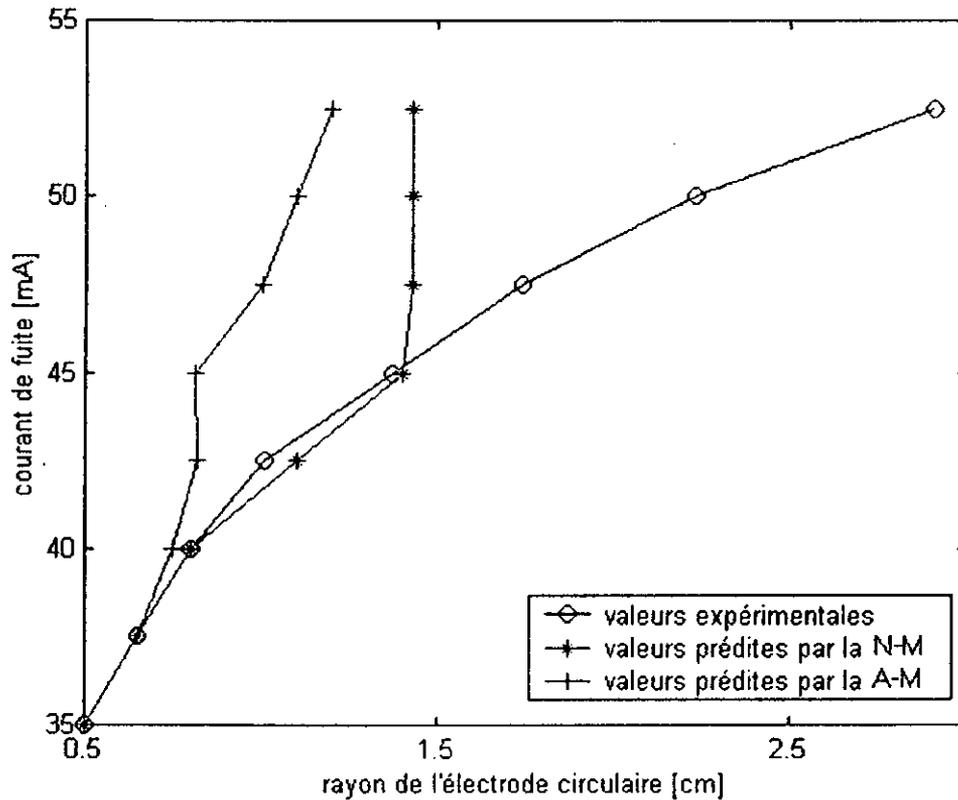


Fig.III.19. Variation du courant de fuite en fonction du rayon de l'électrode circulaire (2 points d'apprentissage – 6 points prédits)

III.4.2. Variation du courant de fuite en fonction de la conductivité de la couche de pollution :

Pour cette caractéristique, nous utiliserons, comme dans le cas de la variation du courant de fuite en fonction du rayon de l'électrode circulaire, trois domaines de prédiction.

Pour cette caractéristique aussi, le premier réseau a montré ses limites pour la prédiction. Les courbes prédites ont tendance à diverger dès les premiers points prédits sauf pour le dernier cas (fig.III.22) où il a donné des résultats acceptables.

Le deuxième réseau a donné de bons résultats par rapport au premier, surtout dans les deux premières courbes.

Pour ce qui est de la troisième courbe (fig.III.22), les courbes prédites par les deux réseaux suivent l'allure de la courbe expérimentale.

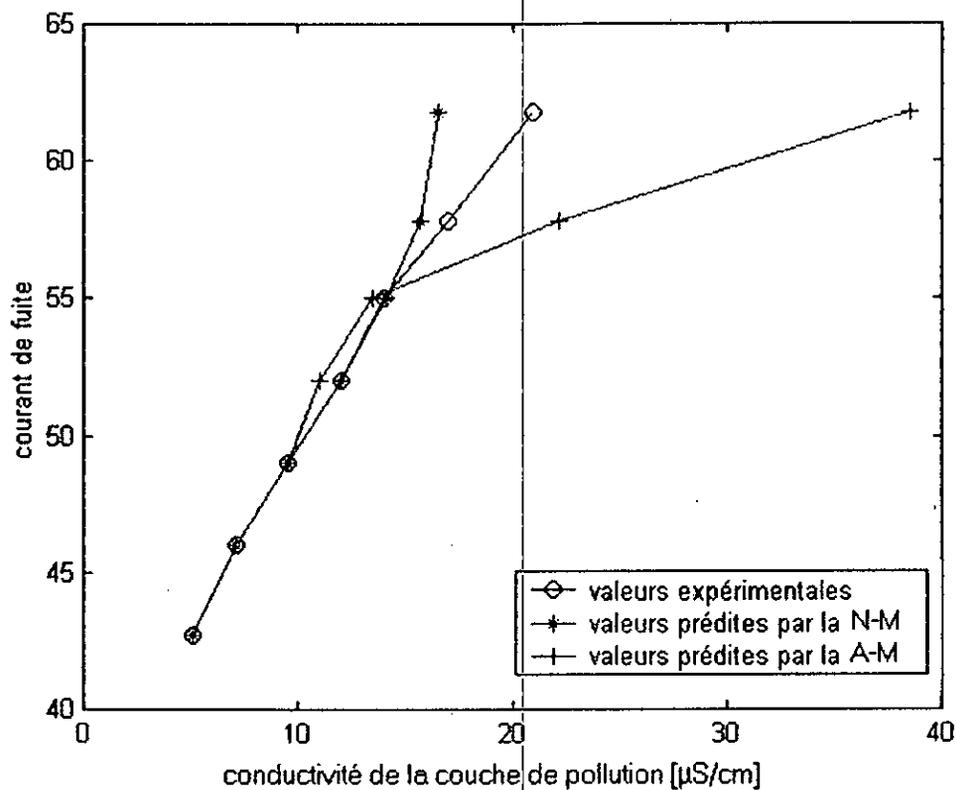


Fig.III.20. Variation du courant de fuite en fonction de la conductivité de la couche de pollution (3 points d'apprentissage – 4 points prédits)

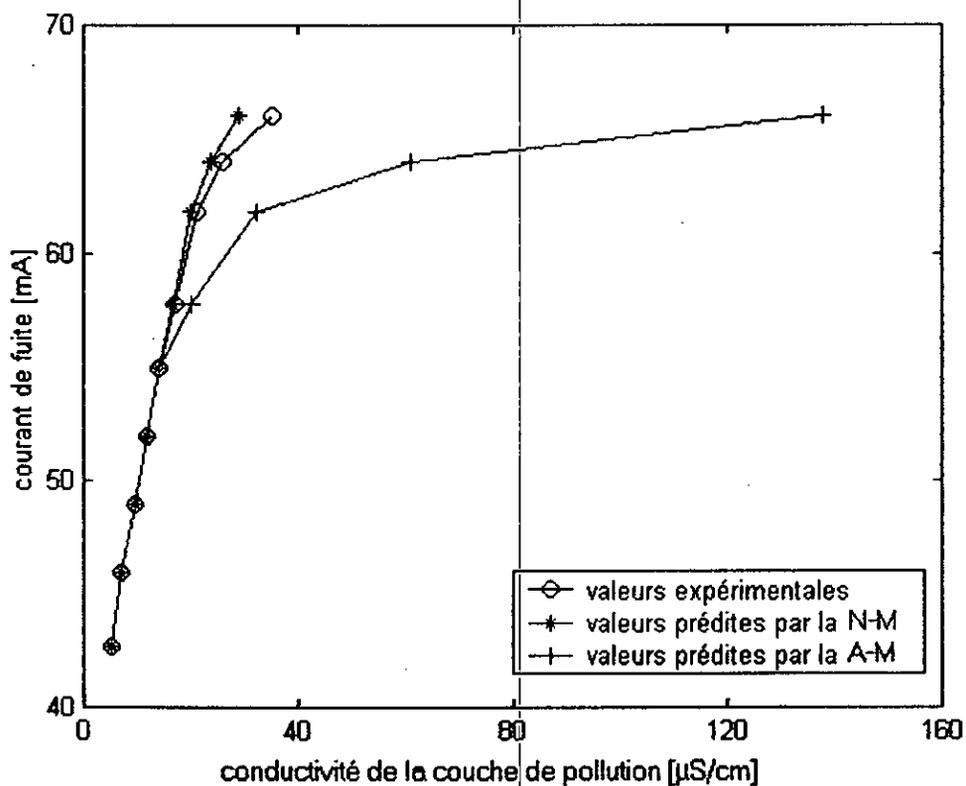


Fig.III.21. Variation du courant de fuite en fonction de la conductivité de la couche de pollution (5 points d'apprentissage – 4 points prédits)

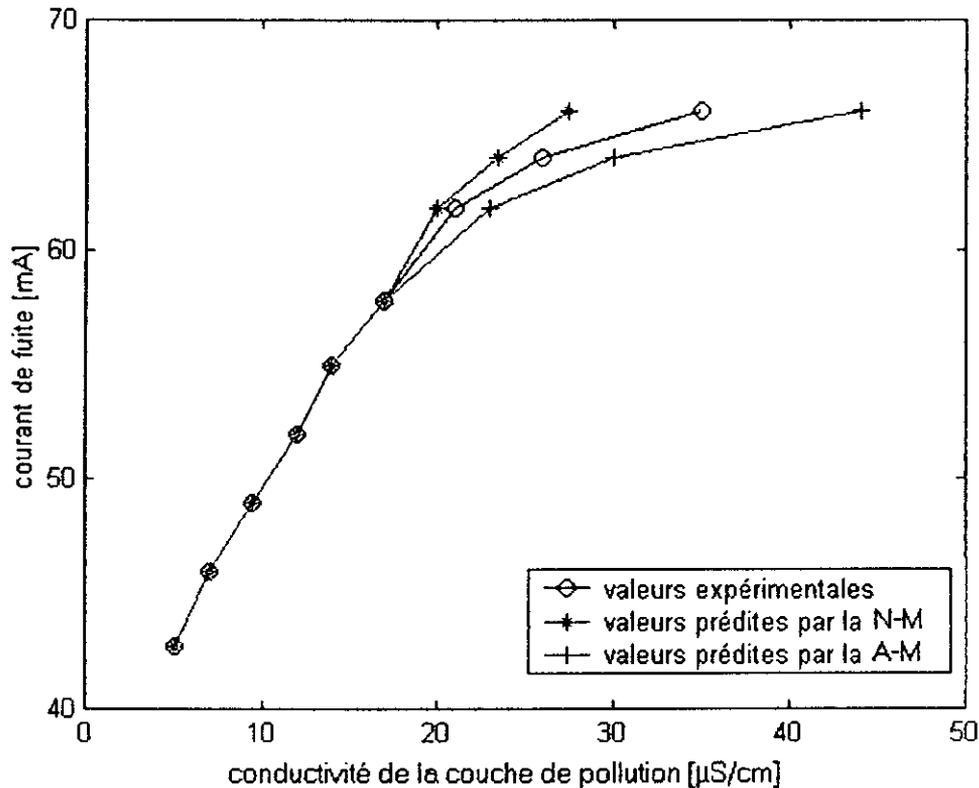


Fig.III.22. Variation du courant de fuite en fonction de la conductivité de la couche de pollution
(6 points d'apprentissage – 3 points d'apprentissage)

III.4.3. Variation de la longueur de l'arc électrique en fonction de la tension appliquée :

Nous avons utilisé trois domaines d'apprentissage. Le premier avec quatre points d'apprentissage et deux points à prédire, le second avec trois points d'apprentissage pour la prédiction de cinq points, et enfin le dernier avec trois points d'apprentissage et trois points prédits.

Encore une fois, le deuxième réseau a donné de très bons résultats, meilleurs que dans le cas des autres caractéristiques (courant de fuite en fonction du rayon de l'électrode et en fonction de la conductivité de la couche de pollution).

Le réseau de neurone de l'ancienne méthode a donné de mauvais résultats. Les courbes obtenues divergent rapidement dès le premier point à prédire.

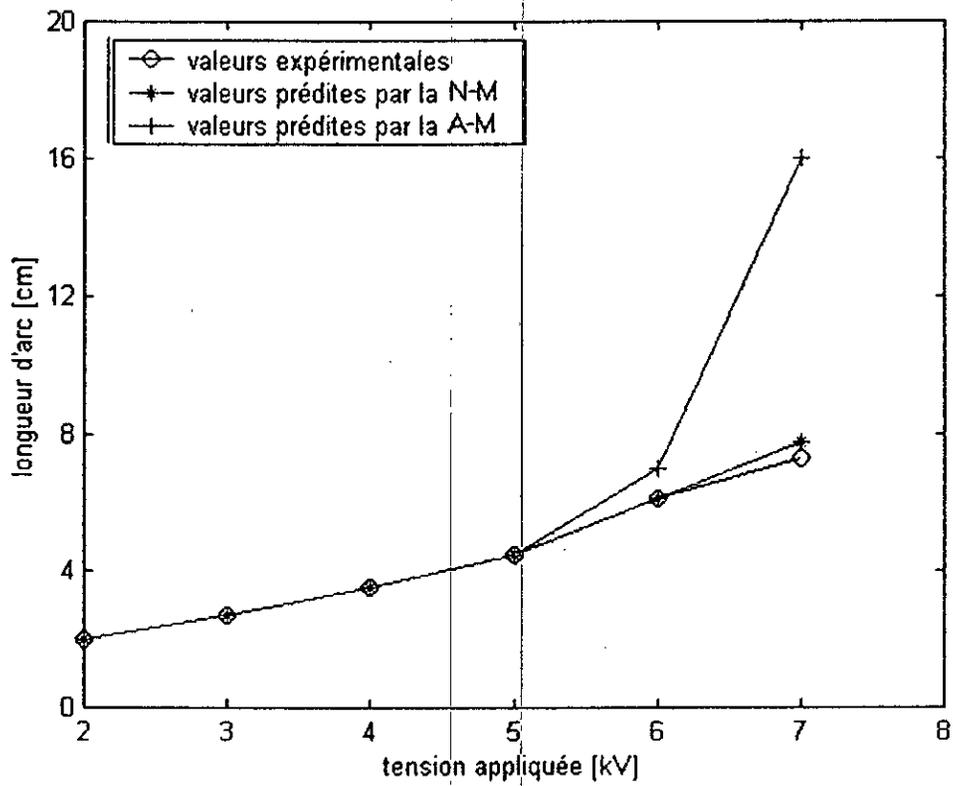


Fig.III.23. Variation de la longueur d'arc en fonction de la tension appliquée (4 points d'apprentissage – 2 points prédits)

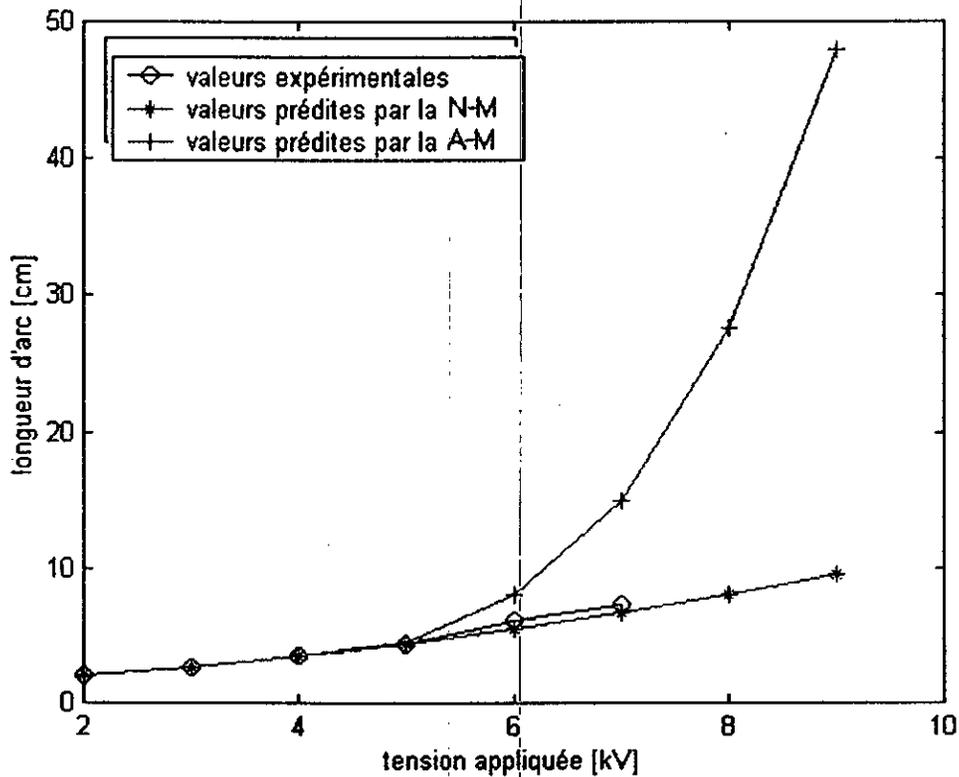


Fig.III.24. Variation de la longueur d'arc en fonction de la tension appliquée (3 points d'apprentissage – 5 points prédits)

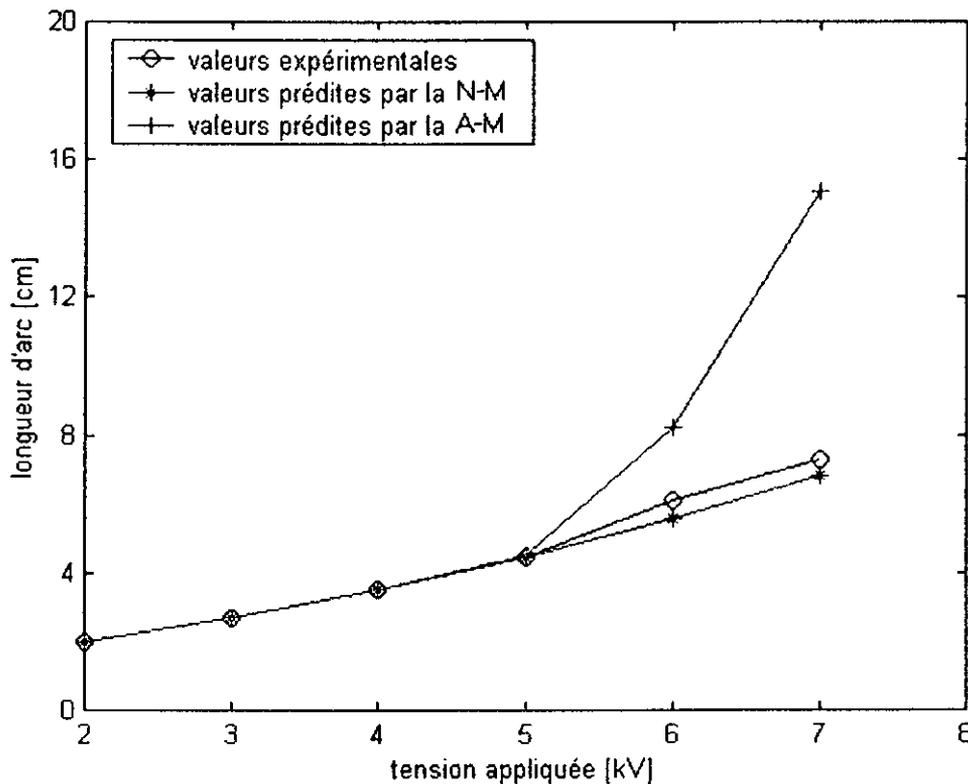


Fig. III.25. Variation de la longueur d'arc en fonction de la tension appliquée (3 points d'apprentissage – 3 points prédits)

III.4.4. Variation de la tension de contournement en fonction de la longueur de la couche polluante :

Les résultats de simulation de cette caractéristique démontrent encore une fois l'incapacité du réseau de neurones de l'ancienne méthode à faire une bonne prédiction en comparaison avec la nouvelle.

Cette dernière a fait une bonne prédiction de la variation de la tension de contournement en fonction de la longueur de la couche polluante. On voit que la courbe obtenue suit exactement l'allure de la courbe expérimentale.

En résumé, nous pouvons dire que pour la méthode 2, dans le cas de la prédiction, la première approche est mauvaise pour la prédiction des quatre caractéristiques étudiées, et que pour pouvoir utiliser la méthode 2, il est préférable d'utiliser la seconde approche, la notre. Cette dernière a donné de bons résultats et est peu influençable par le domaine d'apprentissage.

Le deuxième réseau dépend plus du nombre de point d'apprentissage, même s'il reste peu influençable, que le nombre de points à prédire.

Les courbes obtenues pour cette caractéristique sont présentées dans les figures fig. III.26 et fig. III.27.

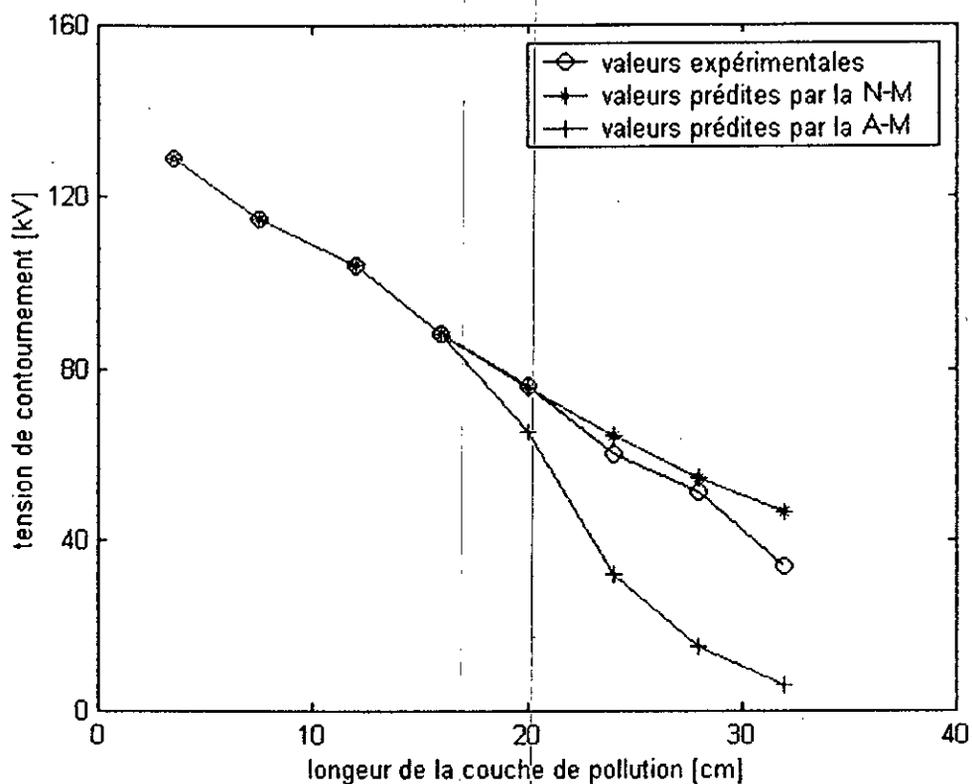


Fig.III.26. Variation de la tension de contournement en fonction de la longueur de la couche de pollution
(4 points d'apprentissage – 4 points prédits)

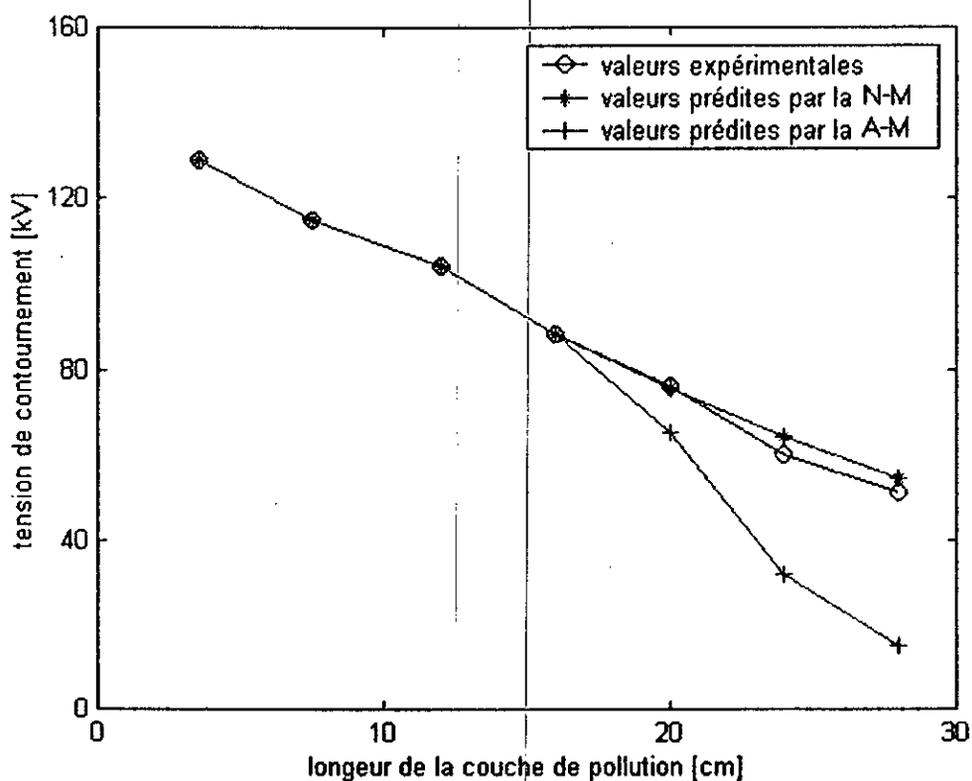


Fig.III.27. Variation de la tension de contournement en fonction de la longueur de la couche de pollution
(4 points d'apprentissage – 3 points prédits)

III.5. Résultats de la troisième méthode :

III.5.1. Variation du courant de fuite en fonction du rayon de l'électrode circulaire :

Pour cette caractéristique, le réseau de neurone de la méthode 3 a montré son efficacité à prédire avec une excellente précision la courbe expérimentale dans le cas des deux approches.

Si on compare entre ces deux la, on voit que dans la première courbe, le premier réseau a une meilleure précision que le second, mais pour les autres courbes, la nouvelle méthode a prédit exactement la courbe expérimentale, avec une erreur pour certains points prédits de moins de 10^{-4} .

Même si la prédiction dans le cas des deux méthodes est très bonne et précise, le deuxième réseau donne une meilleure précision.

Les résultats obtenus sont illustrés dans les courbes suivantes :

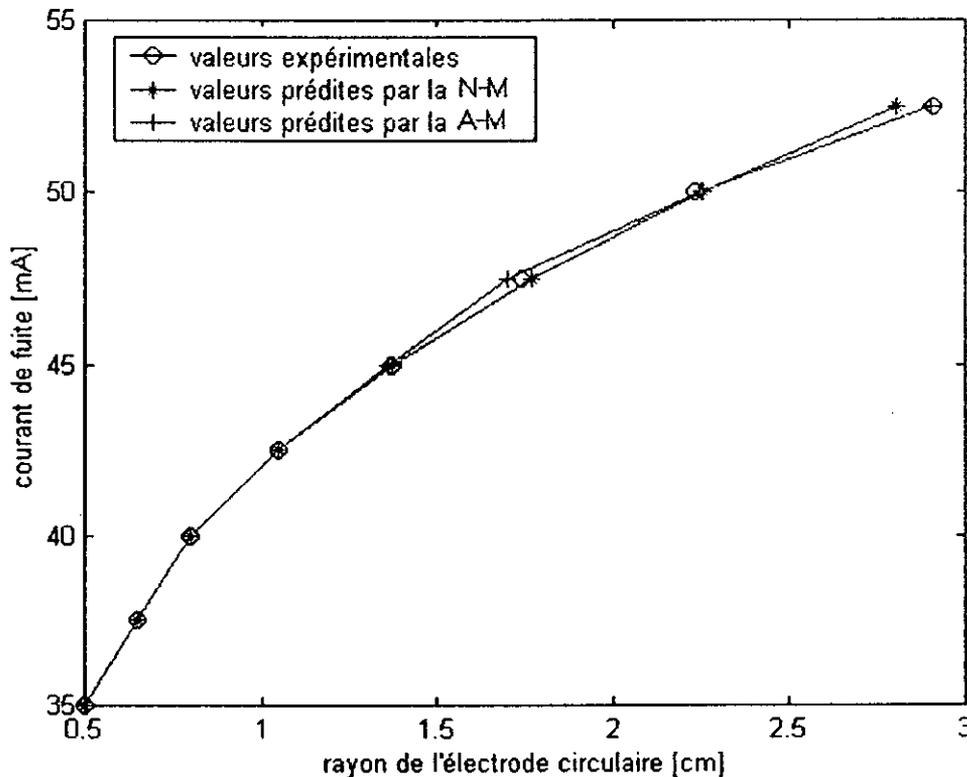


Fig III.28. Variation du courant de fuite en fonction du rayon de l'électrode circulaire (4 points d'apprentissage – 4 points prédits)

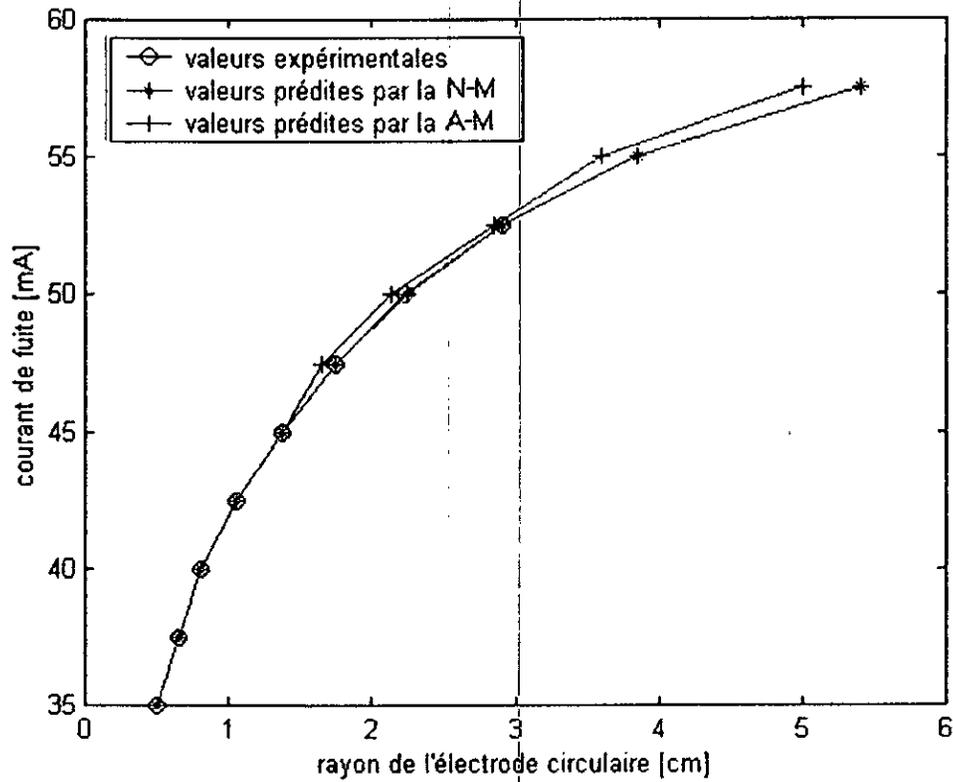


Fig.III.29. Variation du courant de fuite en fonction du rayon de l'électrode circulaire (5 points d'apprentissage – 5 points prédits)

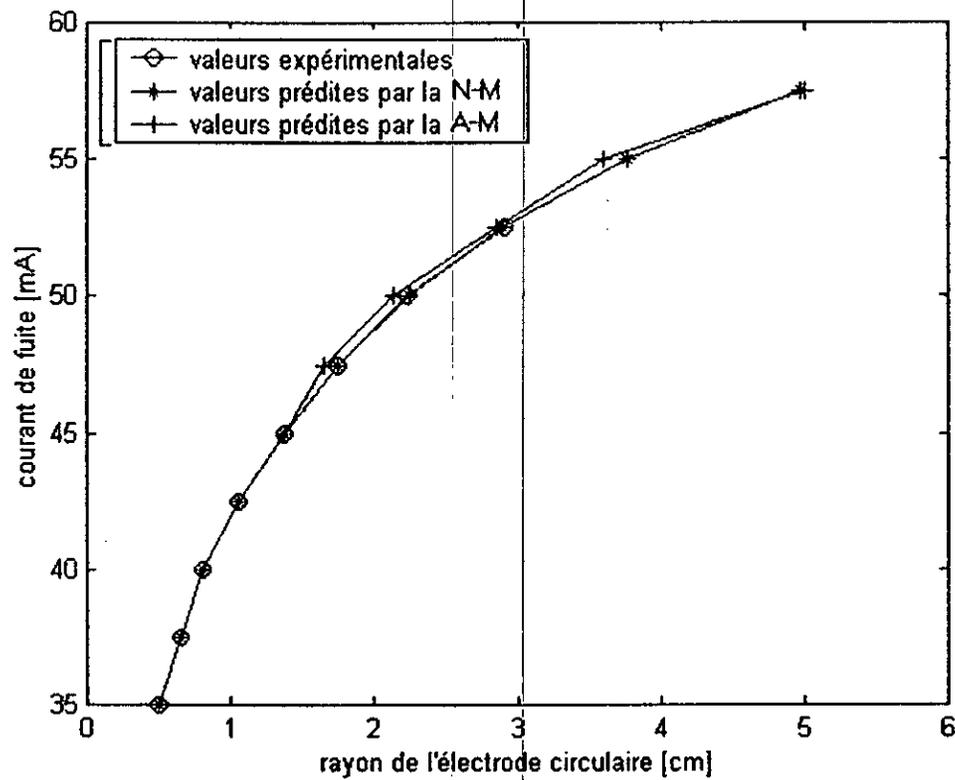


Fig.III.30. Variation du courant de fuite en fonction du rayon de l'électrode circulaire (4 points d'apprentissage – 6 points prédits)

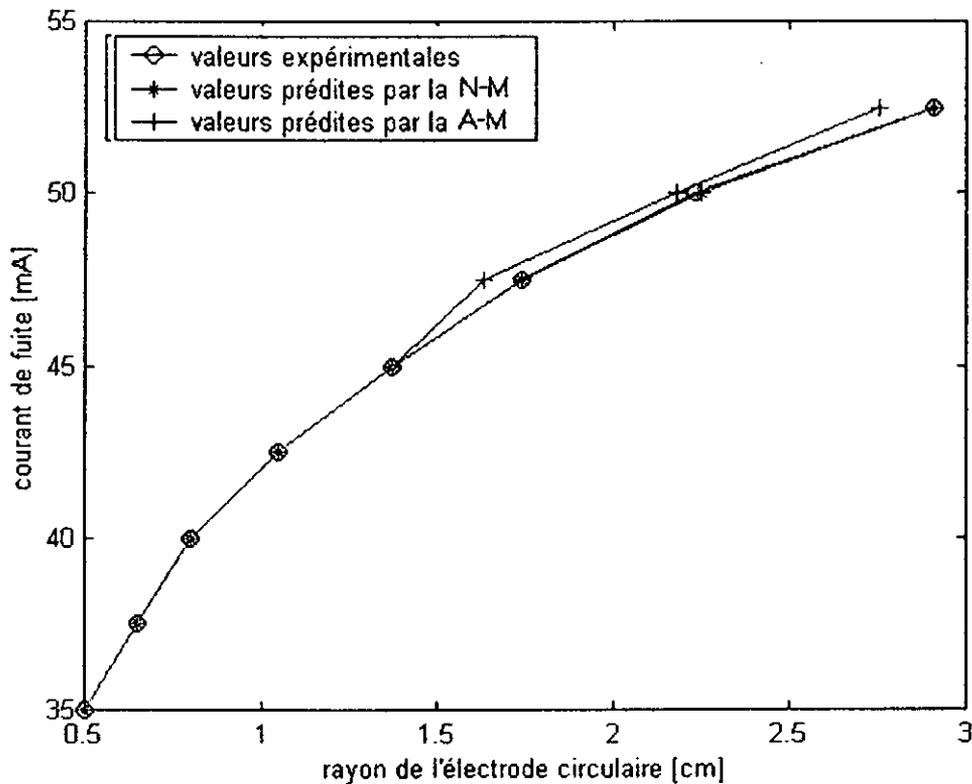


Fig.III.31. Variation du courant de fuite en fonction du rayon de l'électrode circulaire (5 points d'apprentissage – 3 points prédits)

III.5.2. Variation du courant de fuite en fonction de la conductivité de la couche de pollution :

Dans cette caractéristique, nous avons utilisé trois domaines d'apprentissage en changeant le nombre de points à prédire et en gardant le même nombre de point d'apprentissage.

Les deux approches ont fait une bonne prédiction de la courbe expérimentale, pour un différent nombre de points à prédire. L'erreur de prédiction est pratiquement la même dans les deux cas, entre 1.22% à 4.65% pour la première courbe et de l'ordre de 4% pour les autres.

Même si la précision est moins bonne que dans le cas de la variation du courant de fuite en fonction du rayon de l'électrode circulaire, la prédiction de cette caractéristique reste de très bonne qualité.

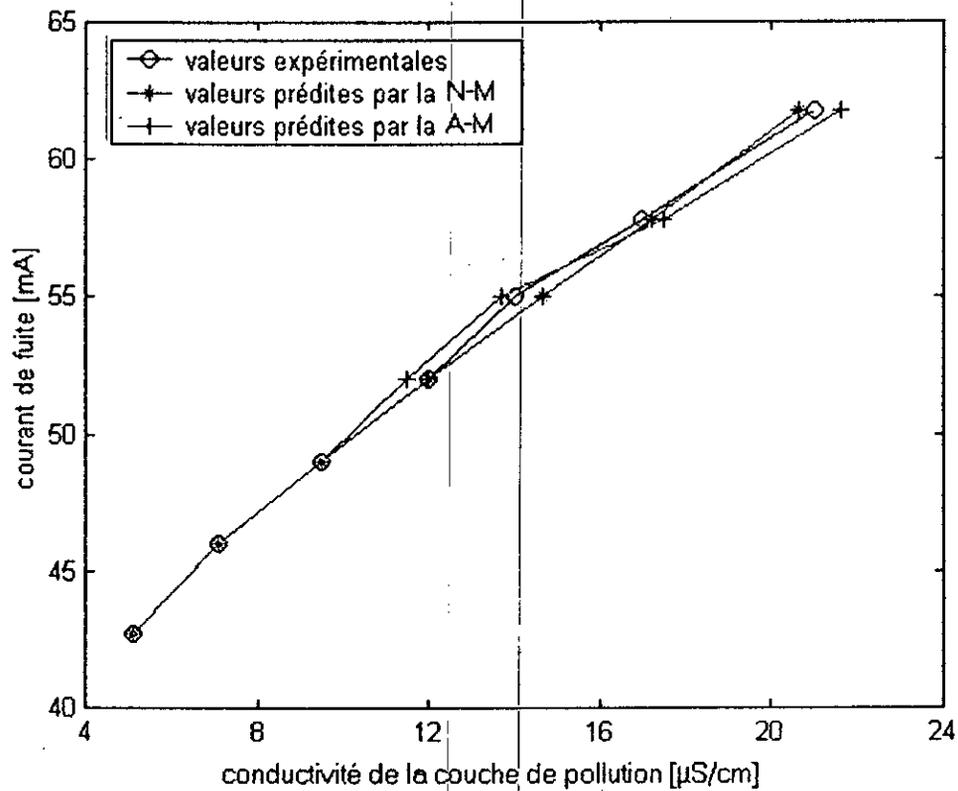


Fig.III.32. Variation du courant de fuite en fonction de la conductivité de la couche de pollution.
(3points d'apprentissage – 4 points prédits)

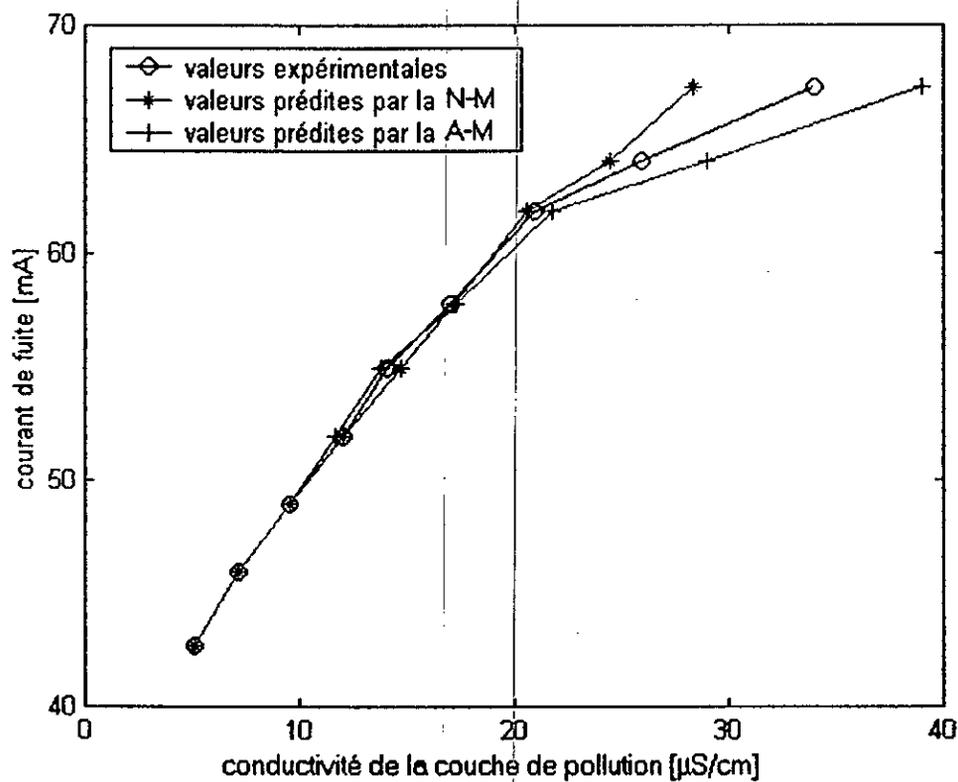


Fig.III.33. Variation du courant de fuite en fonction de la conductivité de la couche de pollution.
(3 points d'apprentissage – 6 points prédits)

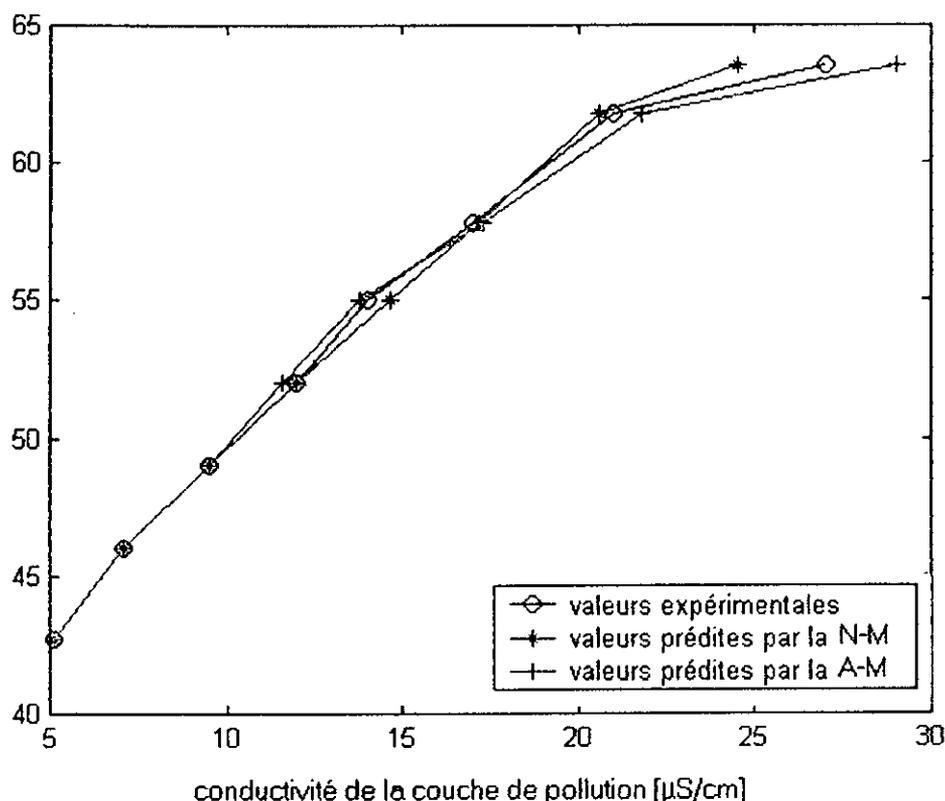


Fig.III.34. Variation du courant de fuite en fonction de la conductivité de la couche de pollution.
(3 points d'apprentissage – 5 points prédits)

III.5.3. Variation de la longueur d'arc en fonction de la tension appliquée :

Pour cette caractéristique, nous avons utilisé quatre domaines d'apprentissage.

On remarque que pour la première courbe, les deux approches donnent de bons résultats, mais la nouvelle montre une meilleure précision.

Les deux courbes prédites ont le même comportement, elles divergent ensuite convergent de nouveau vers la courbe expérimentale.

Dans la fig.III.36, la courbe prédite par le premier réseau diverge et s'éloigne de la courbe expérimentale, tandis que celle prédite par le deuxième réseau, même si la précision est mauvaise, elle suit l'allure de la courbe expérimentale.

Pour les fig.III.37 et fig.III.38, les deux courbes prédites ont la même allure. Les erreurs des courbes obtenues par la nouvelle méthode sont légèrement inférieures, de 1.13% à 2.27%, à celles de l'ancienne méthode

Dans cette caractéristique, on constate l'effet du choix du domaine d'apprentissage sur les deux réseaux. Plus le nombre de points à prédire est supérieur à celui des points d'apprentissage, moins la précision est bonne.

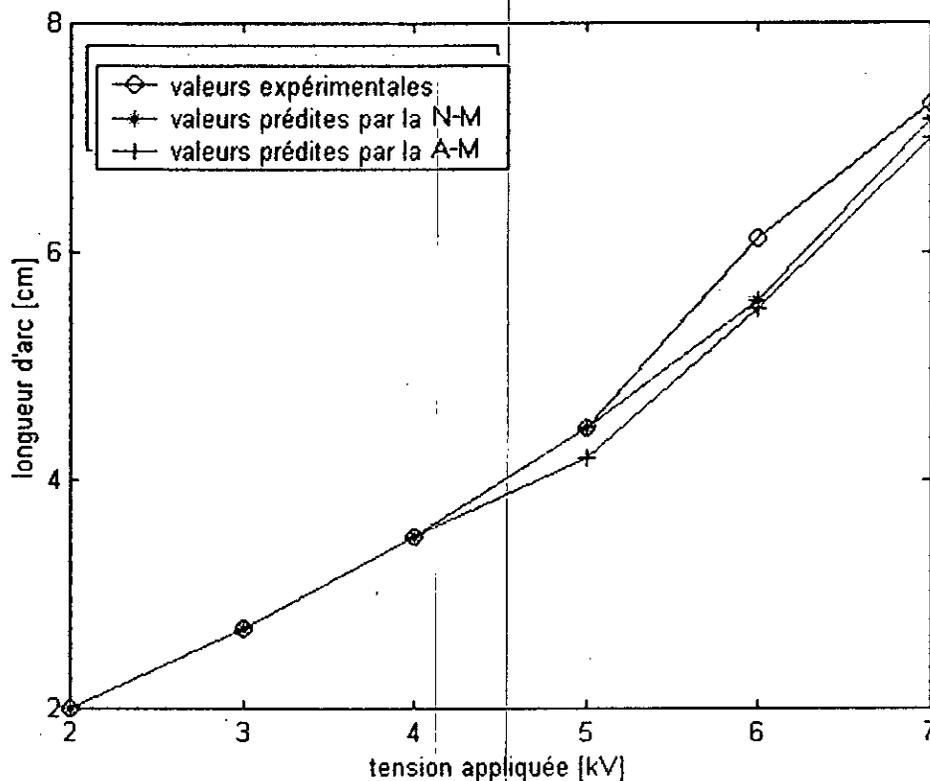


Fig.III.35. Variation de la longueur d'arc en fonction de la tension appliquée (3 points d'apprentissage – 3 points prédits)

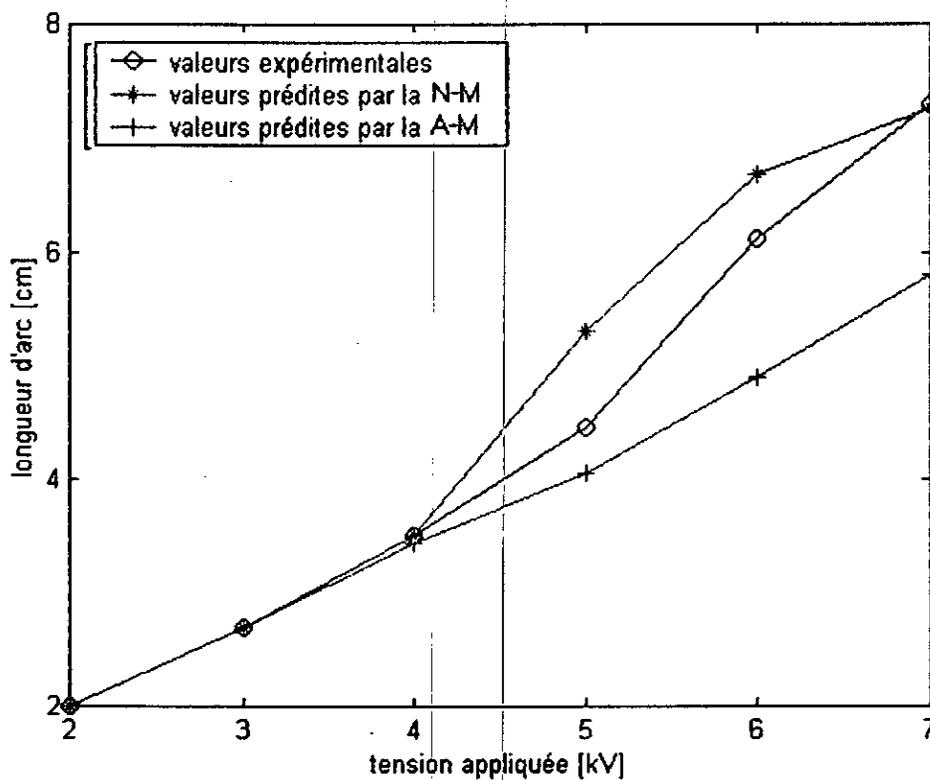


Fig.III.36. Variation de la longueur d'arc en fonction de la tension appliquée (2 points d'apprentissage – 4 points prédits)

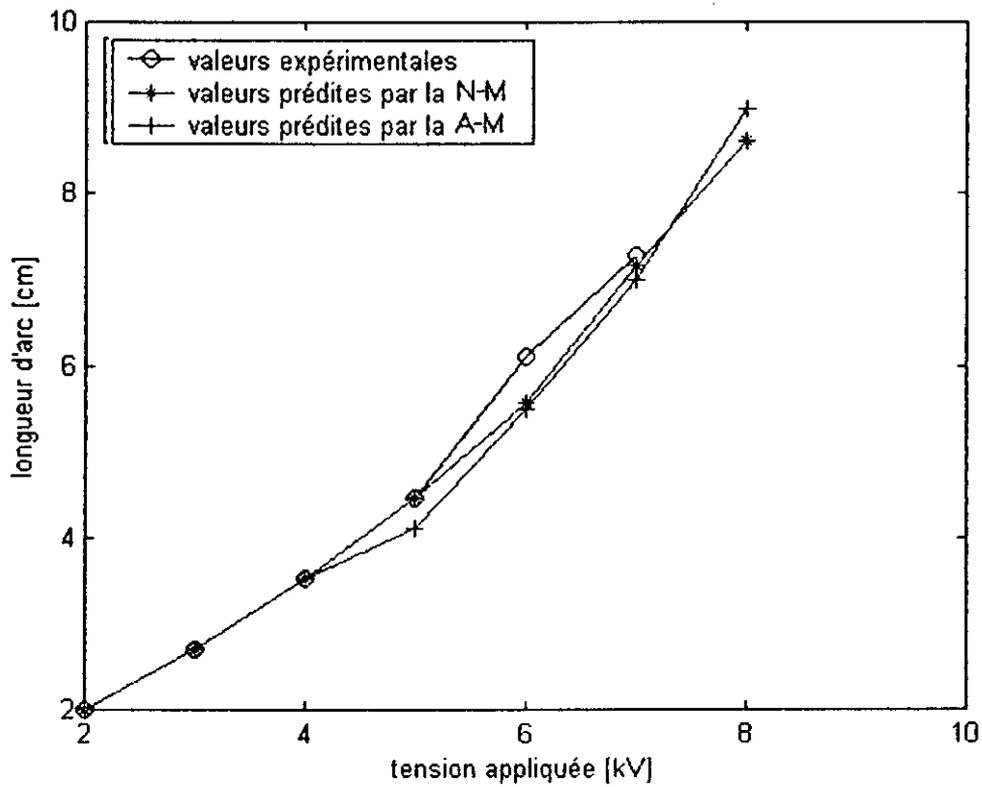


Fig.III.37. Variation de la longueur d'arc en fonction de la tension appliquée (3 points d'apprentissage – 4 points prédits)

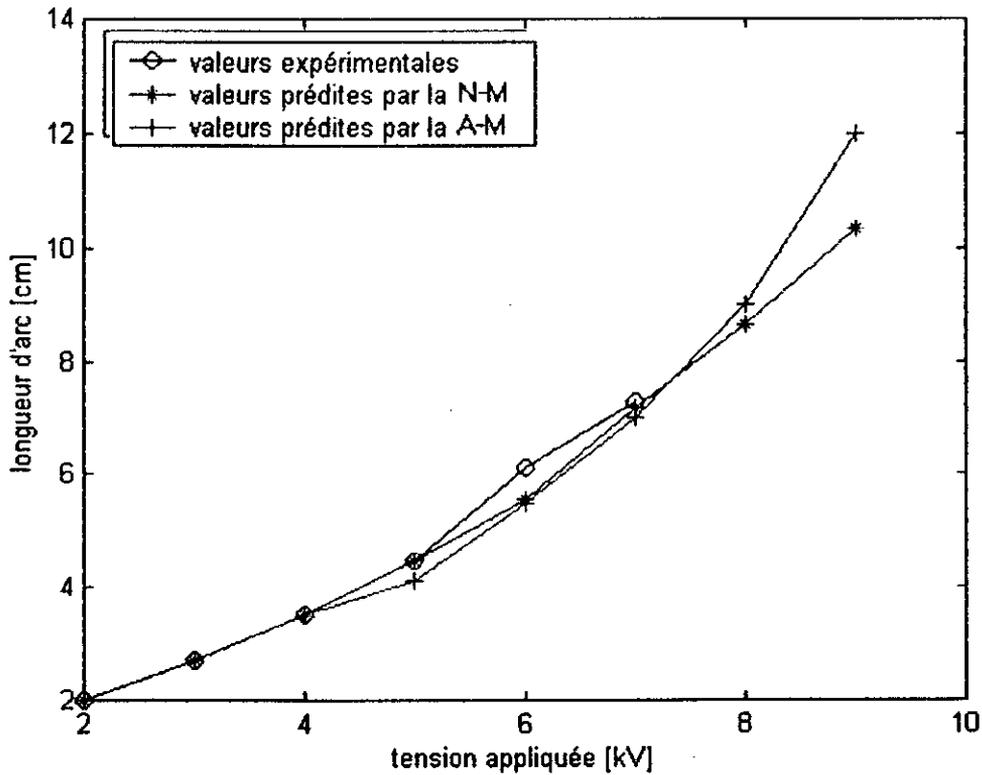


Fig.III.38. Variation de la longueur d'arc en fonction de la tension appliquée (3 points d'apprentissage – 5 points prédits)

III.5.4. Variation de la tension de contournement en fonction de la longueur de la couche de pollution :

Les deux réseaux ont fait une bonne prédiction de la variation de la tension de contournement en fonction de la longueur de la couche de pollution.

Dans la fig.III.39, qui correspond à quatre points d'apprentissage pour 4 points prédits, le deuxième réseau montre une meilleure précision et suit l'allure de la courbe expérimentale, tandis que la courbe de prédiction du premier réseau diverge en premier, converge puis diverge de nouveau de la courbe expérimentale.

Le second réseau montre un meilleur comportement et son erreur est nettement plus faible que le premier réseau, par exemple le dernier point prédit (fig.III.39), le premier réseau a une erreur de 16.17% tandis que notre réseau en a une de 4.48%.

Dans la fig.III.40, les deux courbes prédites divergent puis convergent vers la courbe expérimentale. La nouvelle méthode montre, là aussi, une meilleure prédiction que l'ancienne et converge plus rapidement qu'elle.

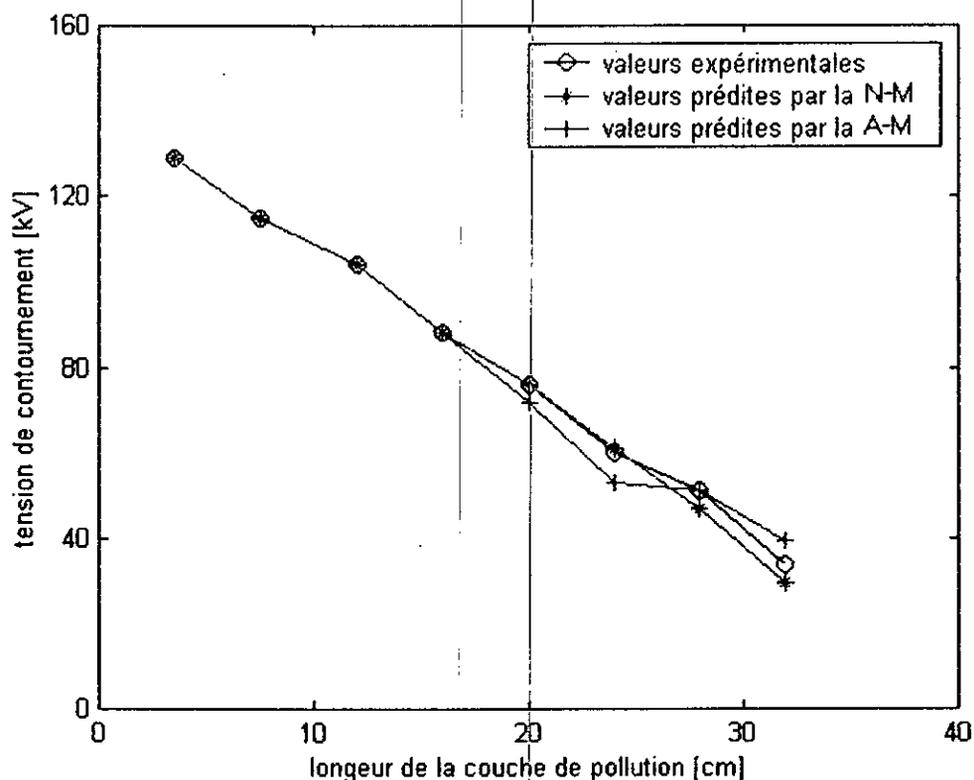


Fig.III.39. Variation de la tension de contournement en fonction de la longueur de la couche de pollution
(4 points d'apprentissage – 4 points prédits)

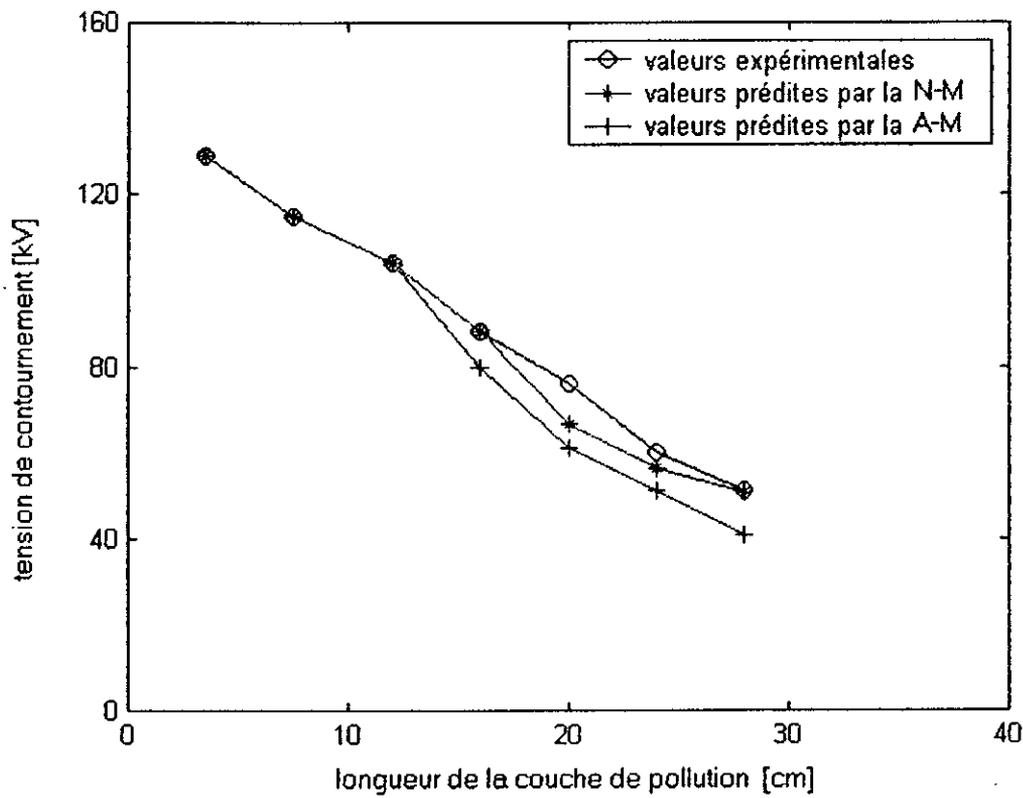


Fig.III.40. Variation de la tension de contournement en fonction de la longueur de la couche de pollution
(3 points d'apprentissage – 4 points prédits)

III.6. Comparaison entre les trois méthodes pour la nouvelle approche :

Nous avons déjà constaté l'efficacité de cette approche pour prédire les différentes caractéristiques électriques, variation du courant de fuite en fonction du rayon de l'électrode circulaire et de la conductivité de la couche de pollution, variation de la longueur d'arc en fonction de la tension appliquée et enfin la variation de la tension de contournement en fonction de la longueur de la couche de pollution.

Dans la figure.III.41 nous avons comparé les résultats obtenus par les trois méthodes pour la prédiction de la variation du courant de fuite en fonction du rayon de l'électrode circulaire. On remarque que la première et la dernière méthode ont donné de bons résultats, et donc une bonne prédiction de cette caractéristique alors que la deuxième méthode donne une mauvaise prédiction et diverge très vite de la courbe expérimentale. Même si la première méthode donne une bonne prédiction, la troisième méthode reste la plus précise.

La comparaison des trois méthodes dans le cas de la variation du courant de fuite en fonction de la conductivité de la couche de pollution est représentée dans la figure.III.41. La deuxième méthode donne une mauvaise prédiction. Elle diverge dès le premier point prédit. La première et la troisième méthode font une très bonne approximation de la courbe expérimentale et toute les deux suivent son allure.

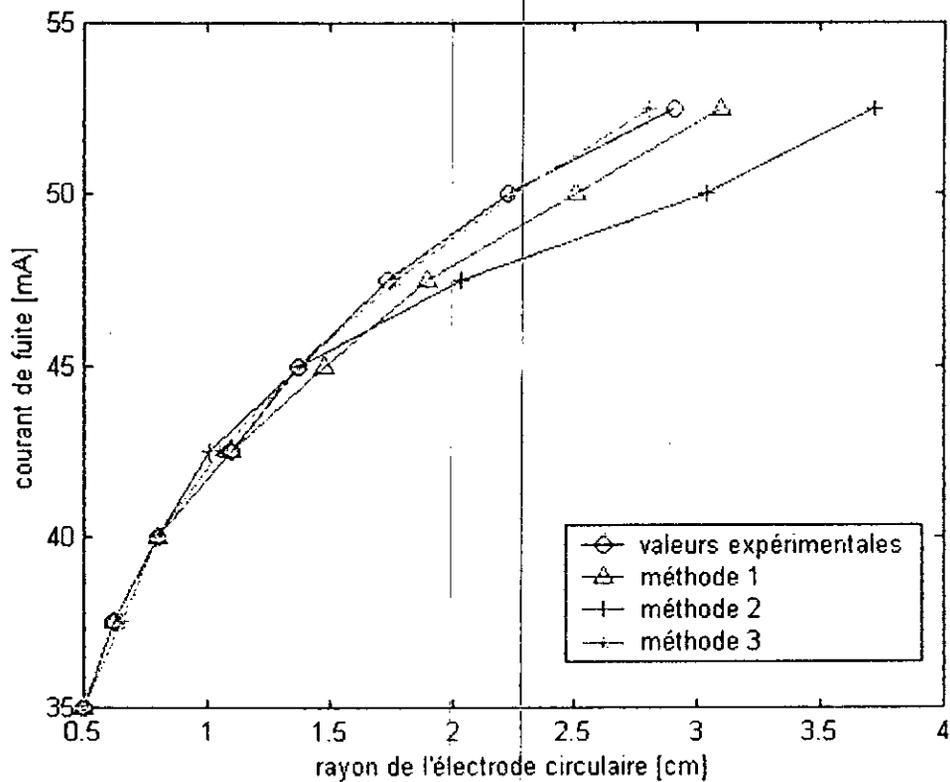


Fig.III.41. Variation du courant de fuite en fonction du rayon de l'électrode circulaire (4 points d'apprentissage – 4 points prédits)

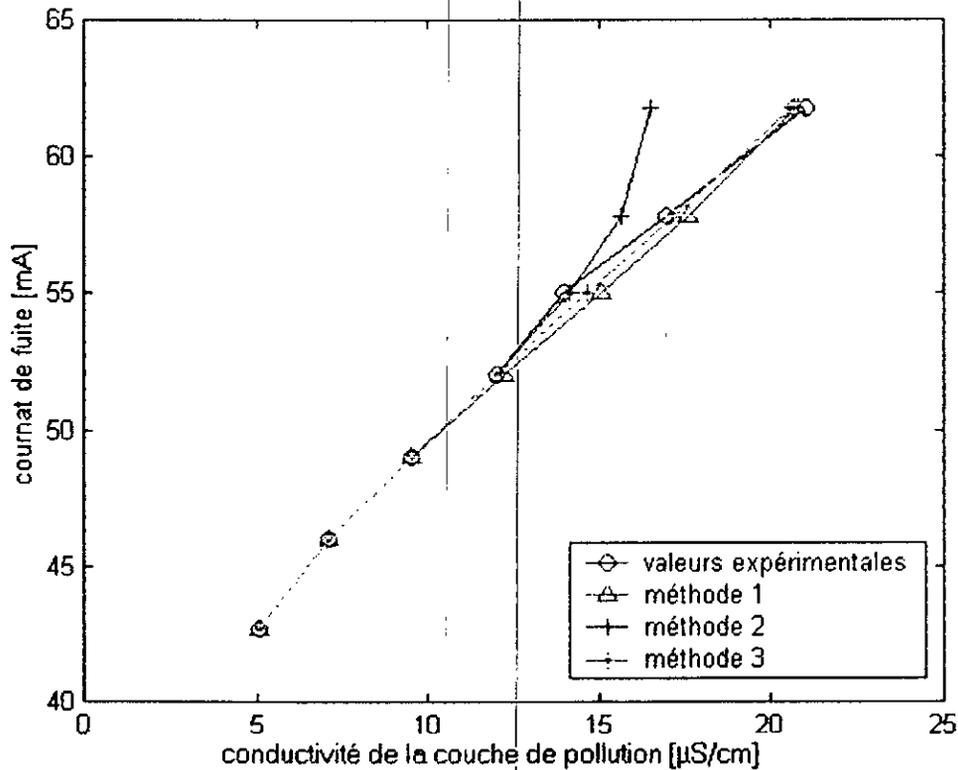


Fig.III.42. Variation du courant de fuite en fonction de la conductivité de la couche de pollution (3 points d'apprentissage – 4 points prédits)

On peut dire que la première et la dernière méthode sont de bon approximateurs de cette caractéristique

Pour la variation de la longueur d'arc en fonction de la tension appliquée, les résultats des différentes méthodes sont représentés dans la figure fig.III.42. Pour cette caractéristique, les trois méthodes ont donné une bonne approximation de la courbe expérimentale. La méthode 3 est encore une fois le meilleur approximateur. Elle converge vers la courbe expérimentale.

La dernière figure, fig.III.43, représente la variation de la tension de contournement en fonction de la longueur de la couche de pollution. La deuxième méthode a donné des résultats acceptables, mais reste un mauvais approximateur, surtout qu'après les deux premières valeurs prédites elle commence à diverger.

La troisième méthode prouve, encore une fois, son efficacité à prédire avec une bonne précision des caractéristiques non linéaires. Elle a donné un meilleur résultat que les deux autres méthodes.

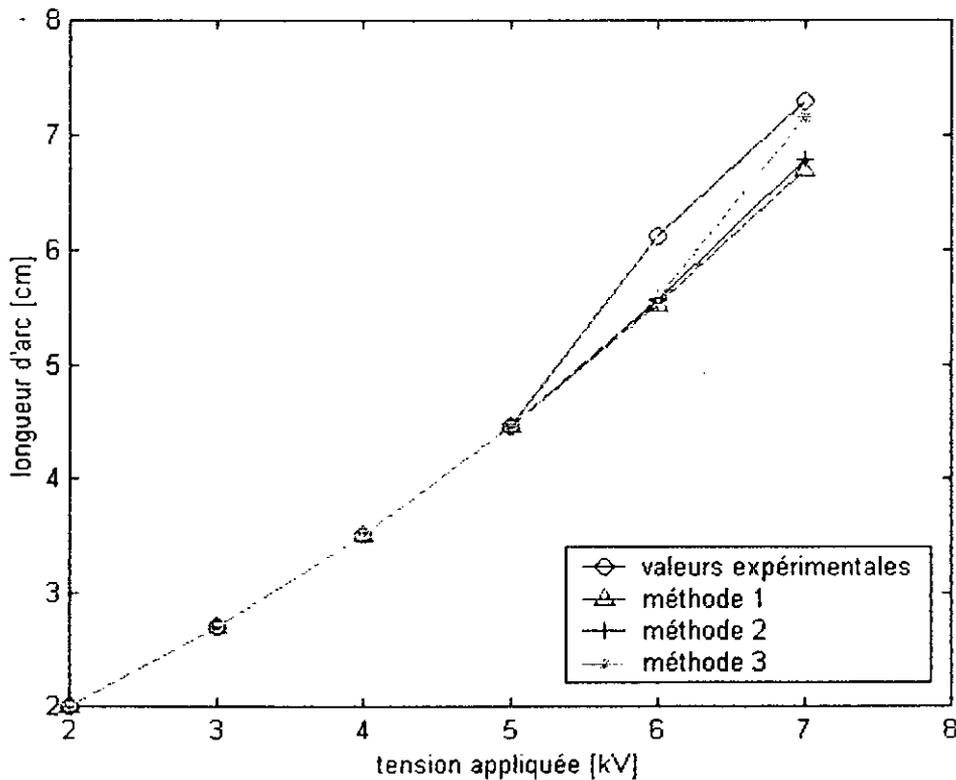


Fig.III.43. Variation de la tension de contournement en fonction de la longueur de la couche de pollution
(4 points d'apprentissage – 4 points prédits)

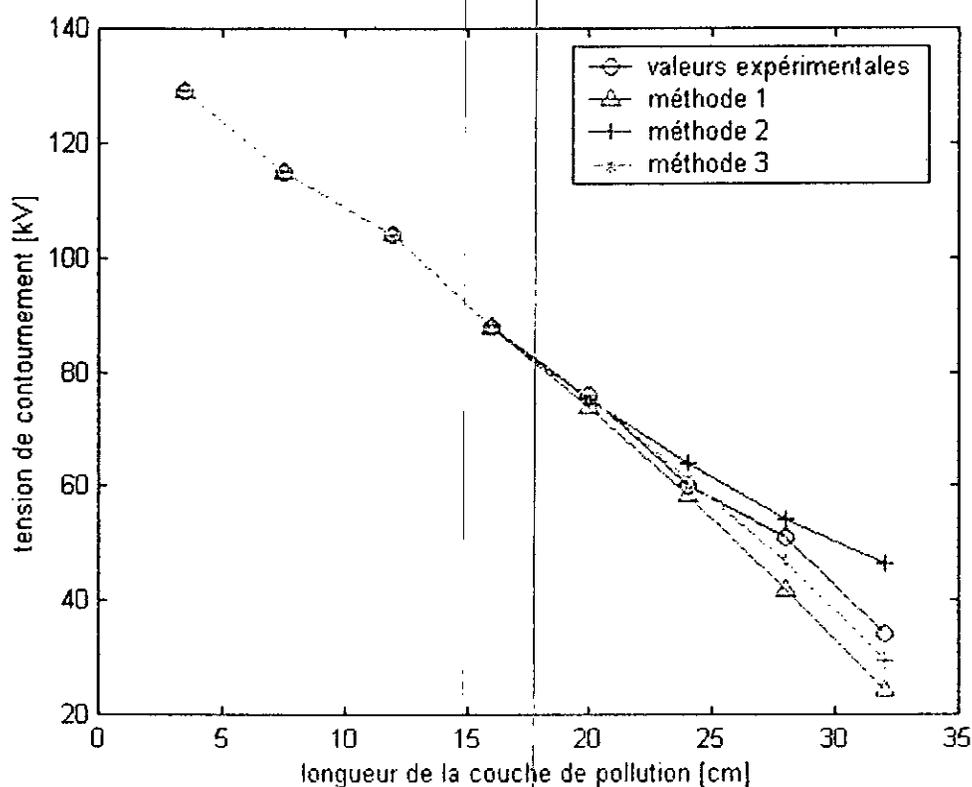


Fig.III.44. Variation de la tension de contournement en fonction de la longueur de la couche de pollution
(4 points d'apprentissage – 4 points prédits)

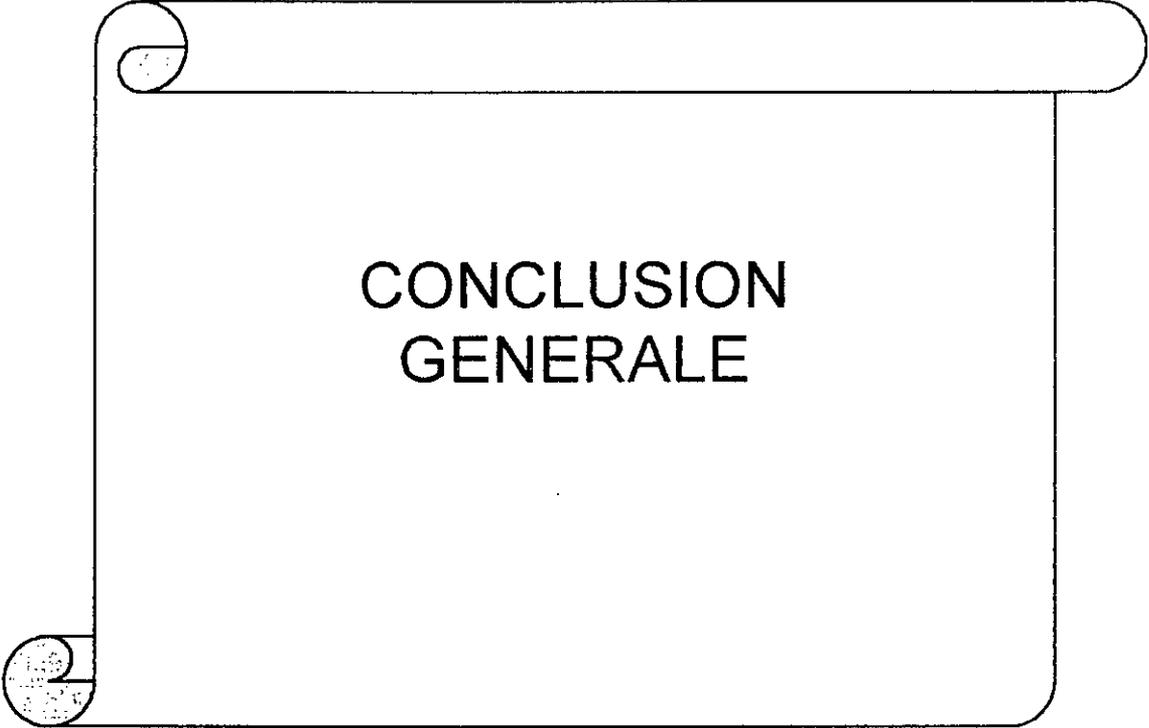
III.7. Conclusion :

Nous avons présenté dans ce chapitre une nouvelle approche pour l'approximation de quatre caractéristiques électriques qui sont la variation du courant de fuite en fonction du rayon de l'électrode circulaire et de la conductivité de la couche de pollution, la variation de la longueur d'arc en fonction de la tension appliquée et enfin la variation de la tension de contournement en fonction de la longueur de la couche de pollution.

Pour cela nous avons utilisé trois différentes architectures qui diffèrent de par la nature de leurs entrées. La première comporte une seule entrée et une seule sortie de nature différente. La deuxième architecture quant à elle, comporte une seule entrée et une seule sortie de même nature. La dernière a deux entrées de nature différente et une seule sortie.

Nous avons constaté clairement que dans le cas des trois méthodes, le nouveau réseau donne une meilleure prédiction que l'ancienne méthode, même pour la méthode 2 où le premier réseau n'arrive à prédire aucune des quatre caractéristiques.

La méthode 3 reste le meilleur approximateur dans le cas des deux réseaux. Elle a montré la meilleure capacité à prédire les différentes caractéristiques étudiées. Elle converge rapidement et suit l'allure de la courbe expérimentale de très près, réduisant ainsi l'erreur. Cette même erreur diminue dans certains cas à moins de 10^E-4 pour le deuxième réseau.



CONCLUSION
GENERALE

Conclusion générale

Dans ce travail nous avons voulu évaluer l'aptitude des réseaux de neurones artificiels (RNA) à généraliser certaines caractéristiques électriques décrivant le comportement d'un isolateur haute tension sous l'effet de la pollution.

Dans un premier temps, nous avons fait une large recherche bibliographique pour mieux définir, connaître et exploiter les réseaux de neurones artificiels. Et aussi une recherche dans les travaux qui ont déjà été effectués dans le domaine de la prédiction dans la technique de la haute tension. Ceci afin d'aboutir au meilleur réseau possible, un réseau robuste et qui nous permet la généralisation et la prédiction de certaines caractéristiques électriques qui nous permettent de connaître le comportement d'un isolateur pollué.

A cet effet, nous avons élaboré trois programmes que nous avons testé et comparé à d'autres du même type. Le premier consiste en un réseau à une seule entrée et une seule sortie de nature différente. Le deuxième programme est basé sur un réseau à une seule entrée et une seule sortie de même nature. Le troisième programme utilise deux entrées et une sortie où l'une des entrées est de même nature que la sortie, quant à la seconde entrée, elle est de nature différente.

Cette étude nous a permis de mettre en évidence la capacité des réseaux de neurones artificiels à fonction de base radiale (RBF) à être de très bons et puissants approximateurs de fonctions. Ces réseaux donne de très bons résultats de prédiction. Leur simplicité et leur capacité à apprendre ont en fait des outils très utilisés dans différentes applications.

De plus, le bon choix des paramètres et des algorithmes leur procure encore plus de puissance et de rapidité à converger. Le choix de la meilleure technique doit en fait être guidé par la nature du problème à résoudre. Ceci est mis en évidence dans notre étude, où grâce à un meilleur choix de l'algorithme et des paramètres, le réseau de neurones a donné un meilleur résultat de prédiction que le réseau utilisé par A.El Hadj [17].

D'après les courbes de simulations obtenues, on constate que les trois programmes présentent une bonne généralisation pour la plus part des caractéristiques dans le cas de la nouvelle méthode, et sont meilleurs approximateurs. La troisième méthode reste incontestablement la meilleure méthode de prédiction et présente la meilleure précision que les deux autres méthodes.

L'utilisation des réseaux de neurones artificiels dans différents domaines pour la prédiction, le diagnostic...etc. reste le moyen le plus économique et le plus rapide. Les RNA ne sont pas seulement économique, ils nous permettent surtout de gagner beaucoup de temps. Ceci permet aux chercheurs de passer moins de temps aux laboratoires, ce qui leur donne plus de temps pour se consacrer à d'autres travaux et activités.

Les bons résultats de ce travail nous encouragent à poursuivre les recherches pour atteindre une bonne optimisation d'un réseau de neurones afin d'améliorer la qualité d'apprentissage et donc de prédiction. Ces recherches nous permettront de mieux maîtriser les réseaux de neurones et de parvenir à prédire n'importe quelle caractéristique. La coopération et la communication entre chercheurs du domaine des RNA et d'autres disciplines sont nécessaires pour atteindre ce but.

Références Bibliographiques

- [1] D.O.Hebb, 'The organization of behavior, a neuropsychological theory', New York : John Wiley, 1949.
- [2] McCulloch et W.Pitts, 'A logical calculus of the ideas imminent in nervous activity', Bulletin of Mathematical Biophysics, Volume 9, pp. 127-147, 1943.
- [3] F.Rosenblatt, 'The perceptron : a probabilistic model for information storage and organization in the brain', Psychological Review, Volume 65, pp. 386-408, 1958.
- [4] M.Minsky, S.Papert, 'Perceptrons', Cambridge, Mass : MIT Press, 1969.
- [5] T.Kohonen, 'A simple paradigm for the self-organized formation of structured featuremaps', in competition and cooperation in neural nets. Ed. S.Amari, M.Arbib, Volume 45 Berlin : Springer Verlag, 1982.
- [6] J.J.Hopfield, 'Neural networks and physical systems with emergent collective computational abilities', Proc.National Academy of Science. 79 : 2554-58, 1982.
- [7] P.J.Werbos, 'Backpropagation Through Time, What it does and How to do it?', proceedings of the IEEE Vol. 78, N°10, Octobre 1990.
- [8] R.L.Harvey, 'Neural network principles', Printice Hall International INC, 1994.
- [9] Y.M.Yeddou, 'Etude de synthèse sur les réseaux de neurones et leurs applications', thèse de magister en Automatique à l'ENP, Alger 1998.
- [10] B.Widrow, M.E.Hoff, 'Adaptive switching circuits', Western Electric Showand Convention Record, part 4, pp. 96-104, août 1960.
- [11] M.J.D.Powell, 'Radial basis functions for multivariable interpolation : A review', J.C.Mason and M.G.Cox, Editors, Algorithms for Approximation, pp. 143-167, Oxford University Press, 1987.
- [12] S.Khemaissia, A.S.Morris, « Review of Networks and Choice of Radial Basis Function Networks for System Identification », Technologies Avancées, N° 6, pp 55-85, 1994.
- [13] J.M.Zurada, 'Introduction to artificial neural systems', West publishing Company, USA,1992.
- [14] R.J.Williams, D.Zipser, 'A learning algorithm for continually running fully recurrent neural networks', Neural Networks, Neural computation, Volume 1, pp. 268-280, 1989.
- [15] A.Karakasoglu, S.I.Sudharsanan, « Identification and Decentralized Adaptive Control Using Dynamical Neural Networks with Application to Robotic Manipulators », IEEE Transactions on Neural Networks, Vol.4, N° 6, Novembre 1993.
- [16] F.J.Sollis, J.B.Wells, 'Minimization by random search techniques', Mathematical Operational Research, Volume 6, pp. 19-29, 1981.

- [17] A.El Hadj, 'Utilisation des Réseaux de Neurones Artificiels dans le Domaine de la Pollution des Isolateurs', thèse de magister à l'ENP, Alger, 2001.
- [18] A.A.Mazroua, R.Bartnikas, M.M.Salama, 'Discrimination between PD pulse shapes using different neural networks paradigms', IEEE Trans. on Dielectrics and Electrical Insulation, Volume 1, N°6, pp. 1119-1131, Décembre 1994.
- [19] A.A.Mazroua, R.Bartnikas, M.M.Salama, 'Neural networks systems using the multilayer perceptron technique for the recognition of PD pulse shapes due to cavities and electrical trees', IEEE Trans. On power Delivery, Vol. 10, N°1, pp. 92-96, Jan. 1995.
- [20] K.Engel, D.Peier, « Influence of PD Fault Development on Fault Type Recognition Using an Artificial Neural Network », Ninth International Symposium on High Voltage Engineering, 28 Août- 1 Septembre 1995, Graz Convention center, Austria, Europe.
- [21] P.S.Ghosh, S.Chakravorti, N.Chatterjee, « ANN with RPROP Algorithm Applied to Modelling of Flashover Voltage Characteristics of Polluted Insulators », Ninth International Symposium on High Voltage Engineering, 28 Août- 1 Septembre 1995, Graz Convention center, Austria, Europe.
- [22] M.E.Ghourab, « Application of Artificial Neural Networks for Insulation Diagnosis », Ninth International Symposium on High Voltage Engineering, 28 Août - 1 Septembre 1995, Graz Convention Center, Austria, Europe.
- [23] T.Kexiong, Z.Deheng, W.Zhenyuan, « Recognition of Partial Discharge Based on Artificial Neural Network », Tenth International Symposium on High Voltage Engineering, August 25-29, 1997, Montréal, Québec, Canada.
- [24] L.Tang, M.R.Raghuveer, 'Classification of Transformer Faults by Using DGA Based Artificial Neural Network', 10th ISH, Montréal, Canada, Aug. 1997.
- [25] H.Okubo, T.Otsuka, K.kato, N.Hayawaka, M.Hikita, 'Electric Field Optimization of High Voltage Electrode Based on Neural Network', IEEE Transaction on Power Systems, Vol.12, N°4, Novembre 1997.
- [26] S.Happe, H.G.Kranz, « Real-Time Partial Discharge Disturbing Pulse Suppression with a Neural Network Hardware », Tenth International Symposium on High Voltage Engineering, August 25-29, 1997, Montréal, Québec, Canada.
- [27] S.Chowdhury, M.S.Naidu, « Application of Artificial Neural Network (ANN) in SF6 Breakdown Studies in Non-Uniform Field Gaps », Eleventh International Symposium on High Voltage Engineering, August 23-27, 1999, London, United Kingdom.
- [28] Y.Tian, P.L.Lewin, A.E.Davies, « PD Pattern Identification Using Acoustic Emission Measurement And Neural Networks », Eleventh International Symposium on High Voltage Engineering, August 23-27, 1999, London, United Kingdom.

- [29] I.Nunes Da silvan, A. Nunes Da Souza, « Analysis of High-Voltage Substations Design Using Artificial Neural Networks », High Voltage Engineering Symposium, 22-27 August, 1999, London, United Kingdom.
- [30] X.H.Jin, C.C.Wang & al., 'Comparaison of PD classification capabilities for transformer failure and typical noise models with neural networks application', CEIDP 2000, IEEE Annual Report Conf. On Electrical Insulation and Dielectric Phenomena, Vol.1, pp. 288-291, Victoria, BC, Canada, octobre 2000.
- [31] L.Mokhnache, 'Application des Réseaux de Neurones dans le Diagnostic et la Prédiction des Isolations de Haute Tension', thèse de doctorat à l'ENP, Alger, 2004.