

15/99

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE
LA RECHERCHE SCIENTIFIQUE



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

ECOLE NATIONALE POLYTECHNIQUE
D.E.R DE GENIE ELECTRIQUE
DEPARTEMENT D'ELECTROTECHNIQUE

المدرسة الوطنية المتعددة التقنيات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique

PROJET DE FIN D'ETUDES

UTILISATION DES RESEAUX DE
NEURONES AU VIEILLISSEMENT
DES POLYMERES UTILISES DANS
LES CABLES MT

Dirigé par :

M^r : A. BOUBAKEUR

Etudié par :

M^r : A.L. NEMMOUR

M^r : A. KHELIFI

PROMOTION : JUIN 99

E.N.P. 10, Avenue Hacén-Badi, El-Harrach, ALGER

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

المدرسة الوطنية المتعددة التقنيات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique



إهداء

أهدي هذا العمل المتواضع
إلى الوالدين الكريمين ...
إلى كل الإخوة و الأخوات ...
إلى كافة أفراد العائلة ...
إلى كل من ساهم في تكويني ...
إلى كل من عرفني و عرفته ...

أحمد لقمان

أهدي ثمرة جهدي
إلى الوالدين الكريمين ...
إلى كل الإخوة و الأخوات ...
إلى كافة أفراد العائلة ...
إلى كل الأصدقاء ...

علي



REMERCIEMENT



Nous présentons nos vifs remerciements à monsieur A.BOUBAKEUR, Maître de Conférence à l'ENP, pour l'assistance qu'il nous a témoigné tout au long de ce travail, pour l'aide morale et matérielle qu'il nous a toujours prodigué et pour la discussion entreprise avec lui sur le sujet et qui nous a été d'un apport appréciable à l'amélioration de ce travail.

Nous exprimons toute notre reconnaissance à tous les enseignants qui ont contribué à notre formation. Qu'ils trouvent ici l'expression de notre profonde gratitude.

Que nos amis A.KERBOUA, H.CHAIBI, B. SEMMAR, L.MOKHTARI, K.SADAT, R.HEMMADI, M.BOUCOUICHA, M.BESSAED trouvent ici toutes notre reconnaissance.

في هذه الدراسة، نستعمل الشبكات العصبونية الاصطناعية كوسيلة جديدة لدراسة التدهور الحراري للبوليميرات المستعملة في عزل الكوابل الكهربائية ذات التوتر المتوسط. في هذا الإطار، تستطيع الشبكات الكهربائية ان تعطينا تغيرات مختلف الخصائص العازلية والكهربائية والميكانيكية تحت عدة توترات و درجات حرارة مختلفة بعد عدة آلاف من ساعات التدهور وهذا بإجراء تجارب التدهور خلال بضع مئات من الساعات فقط.

كلمات مفتاحية : الشبكات العصبونية، التدريب، التنبؤ، التدهور الحراري، البوليميرات.

RESUME

Dans ce travail nous utilisons les réseaux de neurones artificiels (RNA) comme un nouveau outil d'étude du vieillissement des polymères utilisés dans l'isolation des câbles de moyenne tension. En effet, l'application des RNA pourrait nous fournir les variations de différentes propriétés diélectriques et mécaniques à différentes tensions et/ou différentes températures après plusieurs milliers d'heures de vieillissement en effectuant des essais de vieillissement de quelques centaines d'heures seulement.

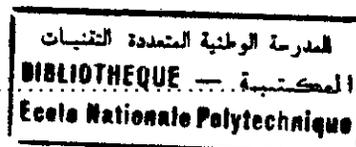
Mots clés : réseaux de neurones, apprentissage, prédiction, vieillissement thermique, polymères.

ABSTRACT

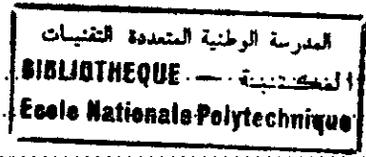
The objective of this work is the Artificial Neural Networks (ANN) using as a new tool to study the polymer ageing which are used in MT cables insulation. the use of the ANN can provide us information about the variation of dielectric and mechanical properties at different tensions and/or different temperatures after many thousands hours of ageing by few hundred hours tests proceeding only.

Key words: neural networks, learning, prediction, thermal ageing, polymer.

SOMMAIRE



Introduction	1
Chapitre I: Introduction aux Réseaux de Neurones	
I.1. Historique	3
I.2 Fondement biologique	4
I.2.1 Le neurone	4
I.2.2 Structure des neurones	4
I.3 Fonctionnement du neurone	5
I.3.1 Fonction discriminant	6
I.3.2 Fonction dynamique linéaire	7
I.3.3 fonction non dynamique non linéaire	7
I.4 Architecture des réseaux de neurones	7
I.4.1 Réseaux statiques	7
I.4.2 Réseaux dynamiques	7
I.5 Apprentissage des réseaux de neurones	8
I.6 Réseaux de neurones à apprentissage non supervisé	9
I.6.1 Fondement de l'apprentissage non supervisé	9
I.6.2 Mémoires associatives	9
I.6.3 Réseaux de Hopfield	10
I.6.4 Mémoires associatives bidirectionnelles (BAM)	11
I.6.5 Mémoires associatives linéaires optimisées (OLAM)	11
I.6.6 Réseau de Hamming	11
I.6.7 Réseaux compétitifs	12
I.7 Réseaux de neurones à apprentissage supervisé	14
I.7.1 Réseaux de neurones basés sur la décision	15
I.7.2 Adaline, Madaline	16
I.7.3 Réseaux multicouches statiques	18
I.7.4 Réseaux à fonction de base linéaire(LBF)	18
I.7.5 Etude du taux d'apprentissage	20
I.7.6 Variantes de la backpropagation	21
I.7.7 Réseaux LBF et approximation de fonction	23
I.7.8 Dimension du réseau, l'entraînement et le dilemme Précision- Généralisation	24
I.7.9 Réseaux à fonction de base radiale RBF	24
Chapitre II: Etude bibliographique sur le vieillissement des polymères	
II.1. Introduction	30
II.2. Polarisation des diélectriques	31
II.2.1.Principe de la polarisation	31
II.2.2. Différents types de polarisation	32
II.2.3. Conduction des isolants polymères	34
II.3 Mécanisme de rupture diélectrique des isolants solides	34
II.3.1. Rupture intrinsèque	35
II.3.2. Rupture diélectrique par collisions ionisantes	35
II.3.3. La rupture thermique	36
II.3.4. Rupture diélectrique provoquée par les décharges partielles	37
II.3.5. Rupture diélectrique provoquée par les arborescences	37



II.4. Polymères et polymérisation	38
II.4.1. Classification des polymères	38
II.4.1.1. Origine du matériau	38
II.4.1.2. Structure chimique et fonctionnalité des motifs	38
II.4.1.3. Géométrie de la macromolécule	39
II.4.1.4. Type de polymérisation utilisé	40
II.4.1.5. Techniques de polymérisation	40
II.5. Caractéristiques des polymères	41
II.5.1. Comportement chimique des polymères	41
II.5.2. Propriétés mécaniques des polymères	41
II.5.3. Propriétés électriques des polymères	41
II.6. Présentation du matériau	42
II.6.1. Présentation du PVC	42
II.6.2. formulation et mise on œuvre du PVC	42
II.6.2.1. Préparation du chlorure de vinyle	42
II.6.2.2. Molécule de PVC	42
II.6.2.3. Mélange de PVC prêt à l'emploi	43
II.6.2.4. Les adjuvants	43
II.6.2.5. Propriétés du PVC	44
II.6.3. Vieillessement thermique du PVC	45
II.6.3.1. Processus de dégradation des polymères	45
II.6.3.2. dégradation thermique du PVC	46
II.6.4. Endurance thermique du PVC	47
II.6.4.1. Position du problème	47
II.6.4.2. Principe de la détermination de l'endurance thermique des isolants solides	48
II.6.4.4. Durée de vie	48
II.6.4.5. Vitesse de dégradation	49
II.7. Présentation de quelques Résultats obtenus par le laboratoire de l'emp.....	49
II.7.1. variation de l'allongement à la rupture en fonction du temps de vieillessement	50
II.7.2. résistance à la traction	50
II.7.3. Endurance thermique	51
II.7.4. Perte de masse.....	53
Chapitre III: Détail et architecture du réseau utilisé	
III.1. Nature du problème étudié	54
III.2. Procédure d'apprentissage suivie dans un problème de prédiction	54
III.3 Architecture du réseau utilisé	55
Organigramme d'un RBF généralisé	61
Organigramme d'un RBF basé sur la ROM.....	63
Chapitre IV : Résultats et interprétations	
Résultats et interprétations	65
Conclusions et suggestions	75
Références bibliographiques	77
Annexe	79

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

INTRODUCTION GENERALE

*"L'information sur le monde extérieur est acquise
par l'observation et la mesure ;
ses résultats sont donc susceptibles d'erreurs"*
Helström, Quantum, Detection and Estimation

Introduction Générale

INTRODUCTION GENERALE

Sous le terme de réseaux de neurones artificiels, on regroupe aujourd'hui un certain nombre de modèles dont l'intention est d'imiter certaines des fonctions du cerveau humain en reproduisant certaines de ses structures de base.

Historiquement, les origines de cette discipline sont très diversifiées. En 1949, Hebb, dans une perspective psycho-physiologique, souligna l'importance du couplage synaptique dans les processus d'apprentissage. C'est en 1958 que Rosenblatt décrit le premier modèle opérationnel de réseaux de neurones, mettant en œuvre les idées de Hebb, McCulloch et Pitts : le perceptron, inspiré du système visuel, capable d'*apprendre* certaines fonctions logiques en modifiant ses connexions synaptiques.

Ce modèle suscita beaucoup de recherches, et sans doute trop d'espoirs. Quand deux mathématiciens, Minsky et Papert, démontrèrent en 1969 les limites théoriques du perceptron, l'effet fut presque immédiat : chercheurs et investisseurs se désintéressèrent des réseaux de neurones pour se tourner vers l'approche symbolique de l'intelligence artificielle, qui semblait beaucoup plus prometteuse.

Le renouveau actuel des réseaux de neurones est dû à des contributions originales, comme celles de Hopfield en 1982 qui montrent l'analogie des réseaux de neurones avec certains systèmes physiques, a permis de leur appliquer un formalisme riche et bien maîtrisé. Plus récemment, en 1985, des nouveaux modèles mathématiques ont permis de dépasser les limites du perceptron.

Aujourd'hui, les premières applications pratiques des réseaux de neurones commencent à avoir le jour. Cette discipline va donc concerner un public de plus en plus large d'étudiants, de chercheurs, d'ingénieurs et d'industriels.

Mais en raison même de la pluridisciplinarité de ce domaine, il est très difficile de l'appréhender de façon cohérente. Plusieurs milliers d'articles ont été publiés sur le sujet, dans des revues de biologie, de psychologie, de mathématique, de physique, d'électronique, chacun abordant le problème dans le cadre de sa spécialité[1].

Depuis les années 80, les chercheurs dans le domaine de génie électrique ont été obligés de s'orienter vers les réseaux de neurones et de s'adapter avec eux, surtout quand il s'agit d'identifier un système dont la dynamique est inconnue, de filtrer des signaux bruités, ou de reconnaître les formes et les paroles, etc....

Dans le domaine de la technique de la haute tension c'est durant les années quatre vingt dix que des travaux ont été communiqués sur l'utilisation des réseaux de neurones artificiels[2].

L'objectif principal est d'aboutir à un diagnostic rapide de l'état de l'isolation dans les appareils et équipement de haute tension.

Au laboratoire de haute tension de l'E.N.P, ce nouveau thème a été introduit afin de définir de nouvelle méthode d'étude du vieillissement des polymères utilisés dans l'isolation des câbles fabriqués en Algérie. En effet, l'application des réseaux de neurones pourrait nous fournir les variations des différentes propriétés diélectriques et mécaniques à différentes tensions et/ou différentes températures, après plusieurs milliers d'heures de vieillissement, en effectuant des essais de vieillissement de quelques centaines d'heures seulement.

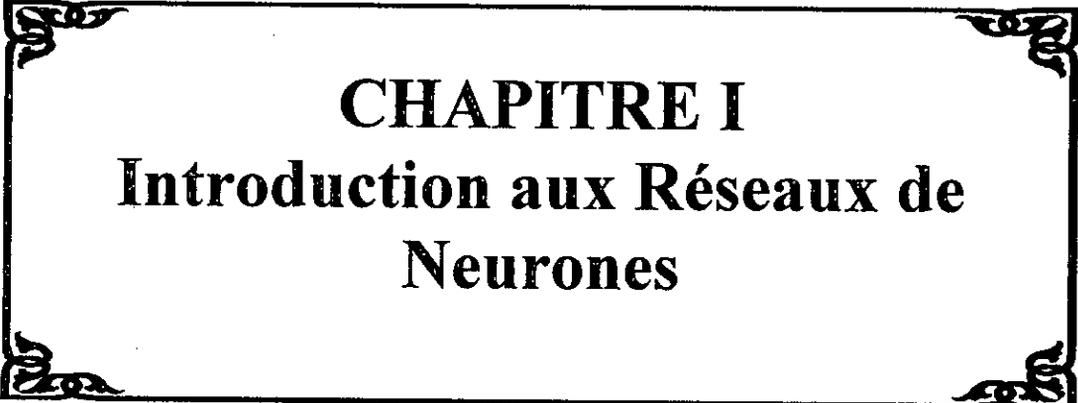
Introduction Générale

Dans notre travail nous nous sommes intéressés au vieillissement thermique du Polychlorure de Vinyle PVC, utilisé dans l'isolation des câbles de moyenne tension fabriqués par l'ex-ENICAB de Gué de Constantine.

Dans un premier chapitre nous présentons une synthèse bibliographique sur les réseaux de neurones artificiels.

Dans le deuxième chapitre nous décrivons le vieillissement des polymères, plus particulièrement le vieillissement thermique du PVC, étudié par M. Nedjar [15].

Dans un troisième chapitre nous détaillons le modèle mathématique que nous avons utilisé et dans un quatrième chapitre nous présentons les résultats obtenus et leur discussion. Nous terminons ce travail par une conclusion générale.



CHAPITRE I
Introduction aux Réseaux de
Neurones

I.1 HISTORIQUE [1]

Le cerveau s'enracine très loin dans le passé des espèces. Il a évolué en taille et en complexité neurologique sur plusieurs millions d'années ; des poissons aux amphibiens puis aux reptiles et aux mammifères, la saga des vertébrés est une lutte incessante, l'expression d'une volonté de puissance, pour s'évader du milieu aquatique et faire la conquête du milieu terrestre. Cette longue marche, exigeant une perpétuelle adaptation à un environnement capricieux, quelquefois sévère, voire dangereux et mortel, a été rendue possible par le développement du cerveau, accroissant précision de l'information sensorielle, permettant la planification et la coordination des actions. Cependant, de façon paradoxale, cet organe fabuleux est resté un mystère pour l'homme. Pourtant, tous les grands esprits depuis les Grecs ont élaboré des théories à son sujet.

Depuis les années 40, il semble que nous assistons à une mutation conceptuelle dans ce domaine, mutation rendue possible par les efforts conjugués de la biologie, des sciences cognitives et des sciences de l'ingénieur. Cette mutation s'est opérée selon le schéma suivant : dans les années 40, une première école de pensée, à la suite de Von Newmann, Wiener, Turing et Mc Culloch, chercha à établir les fondements d'une science des systèmes capable d'auto-organisation. Wiener proposa le terme de cybernétique. En 1943, Mc Culloch et Pitts proposèrent de modéliser la cellule nerveuse ou neurone par un automate à seuil dénommé neurone formel. Quelques années plus tard, Rosenblatt eut l'idée d'organiser ces automates en réseau et conçut le perceptron. Ce modèle pouvait déjà reconnaître des formes simples. Widrow conçut la machine Adaline qui fut utilisé pour la reconnaissance de la parole. Cependant, cette première direction de recherche, après avoir suscité de grands espoirs, fut abandonnée à la suite de l'étude de Minsky et Papert qui mit en évidence les sévères limitations du perceptron. Les chercheurs en intelligence artificielle rejoignirent alors une nouvelle école de pensée, animée par Simon Chomsky, Minsky et McCarthy. Cette école de pensée faisait porter ses efforts sur le calcul symbolique, basé sur l'hypothèse que le raisonnement modélisant la pensée est une combinaison de symboles soumise à des règles logiques. Cette approche symbolique, utilisant l'ordinateur comme instrument eut un énorme succès et entraîna le développement de l'intelligence artificielle. Ce développement fut très fructueux : il dégagait les concepts importants, et permit de mettre au point, dans des domaines sémantiques bien délimités, les fameux systèmes experts. Enfin, il révolutionna en partie le génie logiciel.

Cependant, cette approche symbolique a des faiblesses. Une première faiblesse tient à la lenteur des méthodes séquentielles utilisées, difficilement parallélisables, ce qui donne lieu à une explosion combinatoire dès que le nombre des données devient important. Une seconde faiblesse concerne la représentation des connaissances. Elle est localisée dans ce sens qu'un atome de connaissance est représenté par un objet précis qui peut être par exemple une règle de production, un octet en mémoire, etc.... Ces faiblesses ne se trouvent pas dans l'approche privilégiant les réseaux de neurones, approche dite connexionniste. D'une part, ces réseaux travaillent en parallèle, et d'autre part, la représentation des connaissances est distribuée : la mémoire d'un fait correspond à la mise en activité d'un groupe spécifique de neurones. Ce mode de représentation délocalisé permet une certaine résistance à la détérioration.

Une troisième faiblesse de l'approche symbolique concerne l'apprentissage. En effet, malgré les efforts importants de plusieurs équipes de chercheurs, il semble difficile au niveau symbolique, de simuler l'apprentissage. Or, les réseaux, dans l'approche connexionniste, possèdent une aptitude à l'apprentissage. L'apprentissage d'un fait se

réalise par renforcement des connexions entre les neurones mémorisant ce fait, et le réseau s'auto-organise à partir d'exemples qui lui sont présentés.

En résumé l'approche connexionniste s'oppose à l'approche symbolique, et privilégie les avantages suivants :

- L'activité parallèle et en temps réel de nombreux composants,
- La représentation distribuée des connaissances,
- L'apprentissage par modification des connexions.

Ce furent ces avantages qui poussèrent les chercheurs, dans les années 80, à effectuer une ré-évaluation de l'approche connexionniste. La remise en piste des réseaux de neurones fut alors décidée.

Actuellement, les deux approches coexistent, et l'avenir verra sans doute des tentatives de synthèse, l'approche connexionniste prenant en charge les tâches de bas niveau (perception de formes par exemple) et l'approche symbolique modélisant, intégrant et supervisant différents domaines d'auto-organisation et élaborant des synthèses selon différents niveaux d'abstraction.

I.2 FONDEMENT BIOLOGIQUE

I.2.1 Le neurone

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

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

I.2.2 Structure des neurones

Un neurone est constitué de trois parties :

- Le corps cellulaire,
- Les dendrites,
- L'axone.

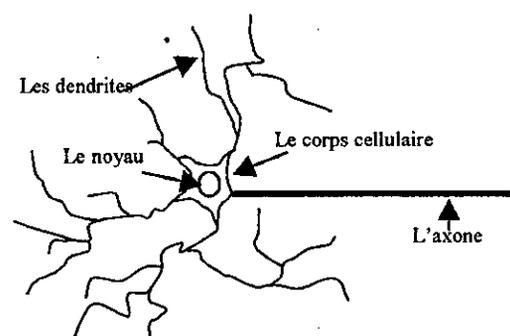
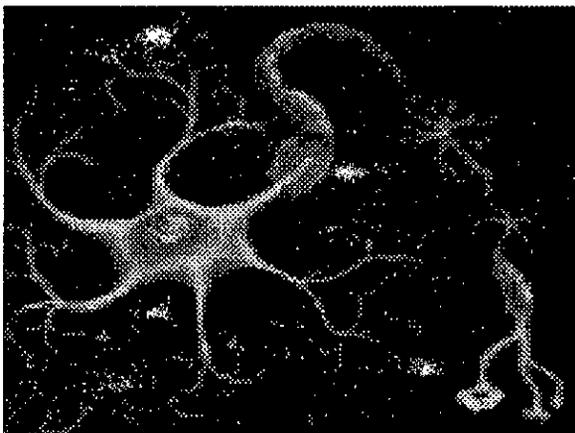


Fig.I.1 le neurone biologique

a. Le corps cellulaire

Il contient le noyau du neurone et effectue les transformations biochimiques nécessaires à la synthèse des enzymes et des autres molécules qui assurent la vie du neurone. Sa forme est pyramidale ou sphérique dans la plupart des cas. Elle dépend souvent de sa position dans le cerveau.

b. Les dendrites

Chaque neurone possède une « chevelure » de dendrites. Celles-ci sont de fines extensions tubulaires. Elles se ramifient, ce qui les amène à former une espèce d'arborescence autour du corps cellulaire. Elles sont les récepteurs principaux du neurone pour capter les signaux qui lui parviennent.

c. L'axone

L'axone, qui est à proprement parlé la fibre nerveuse, sert de moyen de transport pour les signaux émis par le neurone. Il se distingue des dendrites par sa forme et par les propriétés de sa membrane externe. En effet, il est généralement plus long que les dendrites, et se ramifie à l'extrémité, là où il communique avec d'autres neurones, alors que les ramifications des dendrites se produisent plutôt près du corps cellulaire. Pour former le système nerveux, les neurones sont connectés les uns aux autres suivant des répartitions spatiales complexes. Les connexions entre deux neurones se font en des endroits appelés synapses où ils sont séparés par des espaces synaptiques.

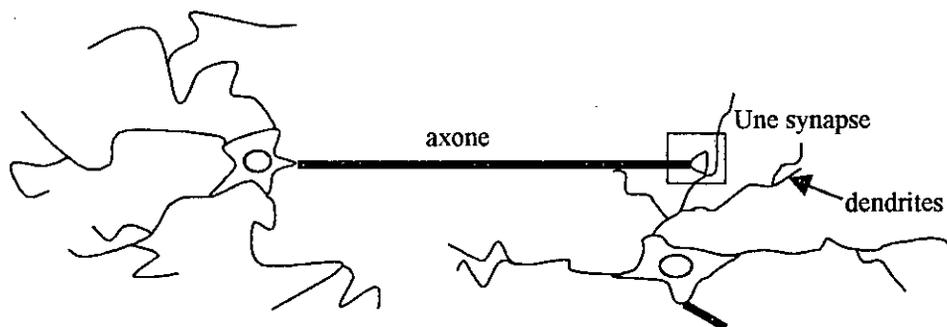


Fig. I.2 le neurone biologique

I.3 Fonctionnement du neurone

Les fonctions spécifiques réalisées par un neurone dépendent essentiellement des propriétés de sa membrane externe.

a. La membrane externe

La membrane externe d'un neurone remplace cinq fonctions principales : elle sert à propager des impulsions électriques tout au long de l'axone et des dendrites,

- Elle libère des médiateurs à l'extrémité de l'axone,
- Elle réagit à ces médiateurs au niveau des dendrites,
- Elle réagit au niveau du corps cellulaire aux impulsions électriques que lui transmettent les dendrites pour générer ou non une nouvelle impulsion,

- Enfin, elle permet au neurone de reconnaître les autres neurones afin qu'il puisse se situer au cours de la formation du cerveau et trouver les cellules auxquelles il doit être connecté.

b. Le corps cellulaire comme sommateur à seuil

D'une façon simple, on peut dire que le sommateur du neurone traite les courants électriques qui lui parviennent de ses dendrites, et qu'il transmet le courant électrique résultant de ce traitement aux neurones auxquels il est connecté par l'intermédiaire de son axone.

Le schéma classique présenté par les biologistes est celui d'un sommateur effectuant une sommation des influx nerveux transmis par ses dendrites. Si la sommation dépasse un seuil, le neurone répond par un influx nerveux ou potentiel d'action qui se propage le long de son axone. Si la sommation est inférieure à ce seuil, le neurone reste inactif.

L'influx nerveux qui se propage entre différents neurones est, au niveau de ces neurones, un phénomène électrique .

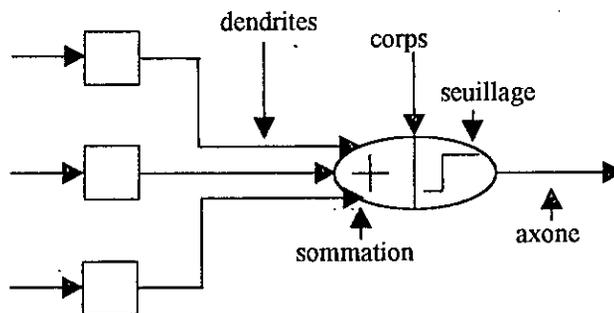


Fig. I.3 le modèle de Mc Culloch et Pitts

I.3.1 Fonction discriminant

Elle définit l'activité du neurone, dans le cas du perceptron c'est la fonction linéaire qui est utilisée, et elle constitue la base des réseaux LBF (linear basis fonction). Afin d'améliorer les performances de ces réseaux, d'autres fonctions non linéaires ont été élaborées. Nous citons ci-dessous les plus utilisées.

a. Fonction de base radiale RBF (radial basis fonction)

Sa forme générale est :

$$\Phi = \frac{\|x - w\|^2}{2} \tag{I.1}$$

où w représente les centroïdes de chaque classe de l'espace d'entrée. Ce dernier sera échantillonné par cette fonction.

b. Fonction de base elliptique EBF (elliptique basis fonction)

La forme de cette fonction est la généralisation de la fonction de base radiale :

$$\Phi = \sum_{k=1}^n \alpha_k (x - w)^2 + \theta \tag{I.2}$$

elle est modulée par des paramètres α_k et dotée d'un biais θ .

I.3.2 Fonction dynamique linéaire

Cette fonction détermine la dynamique du réseau, elle est régit par l'équation différentielle suivante :

$$\alpha_0 \dot{u}_i(t) + \alpha_1 u_i(t) = v_i(t) \quad (\text{I.3})$$

où u_i est l'activation du $i^{\text{ème}}$ neurone et v_i est l'entrée du système dynamique décrit par l'équation (I.3)

I.3.3 fonction non dynamique non linéaire

Elle a pour rôle de borner l'activité du neurone. Son premier modèle proposé est la fonction seuil qui est une fonction binaire, mais dont le problème majeur est qu'elle n'est pas dérivable. D'autres fonctions qui sont dérivables, monotones, continues et qui peuvent effectuer le seuillage ont été mises au point. Parmi ces fonctions on retrouve la sigmoïde, la tangente hyperbolique ou la log-sigmoïde.

Une autre fonction qui est très utilisée malgré qu'elle ne soit pas monotone, est la fonction gaussienne dont le grand avantage est son comportement local.

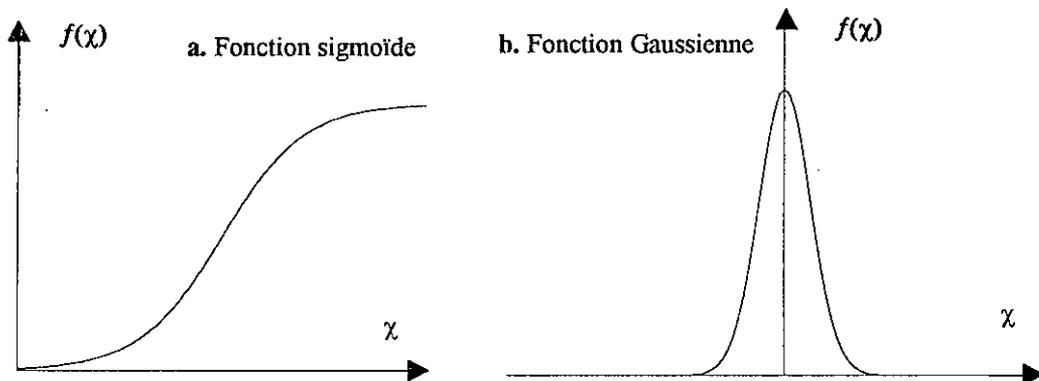


Fig.I.4 Allure de fonctions d'activation

I.4 Architecture des réseaux de neurones

Selon leurs architectures, on distingue trois types de réseaux de neurones.

I.4.1 Réseaux statiques

Ces réseaux sont constitués de plusieurs couches, chaque neurone d'une couche reçoit ses entrées à partir des neurones de la couche précédente. Dans de tels réseaux, il n'existe pas de retour d'informations. Ils peuvent être utilisés dans les problèmes de classification et d'approximation de fonctions.

I.4.2 Réseaux dynamiques

Contrairement aux précédents, ces réseaux contiennent des retour d'informations (figI.5). Cet aspect leur confère la possibilité de commander et d'identifier les systèmes dynamiques.

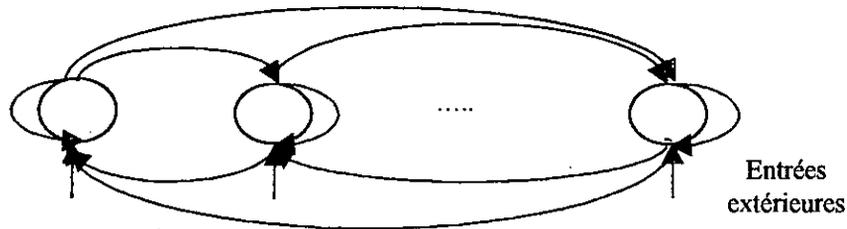


Fig.I.5 Architecture générale d'un réseau dynamique

I.5 Apprentissage des réseaux de neurones

En général, l'apprentissage désigne tout processus susceptible de modifier un comportement ultérieur. Pour un réseau, il lui permet de s'adapter avec des situations désirées tout en modifiant ses poids internes au fur et à mesure qu'on lui présente de nouveaux exemples, et plus il y a d'associations reliant les éléments nouveaux à ce qui est déjà connu, meilleur est l'apprentissage. On peut distinguer deux types d'apprentissage (figI.6)

a. Apprentissage supervisé

Cet apprentissage se fait en introduisant des couples d'entrées et leurs sorties désirées, et on détermine les performances du réseau par l'intermédiaire d'un critère à optimiser.

b. Apprentissage non supervisé

L'apprentissage non supervisé nécessite la présence des entrées seulement. L'entraînement est basé seulement sur l'information locale existant aux niveaux des neurones.

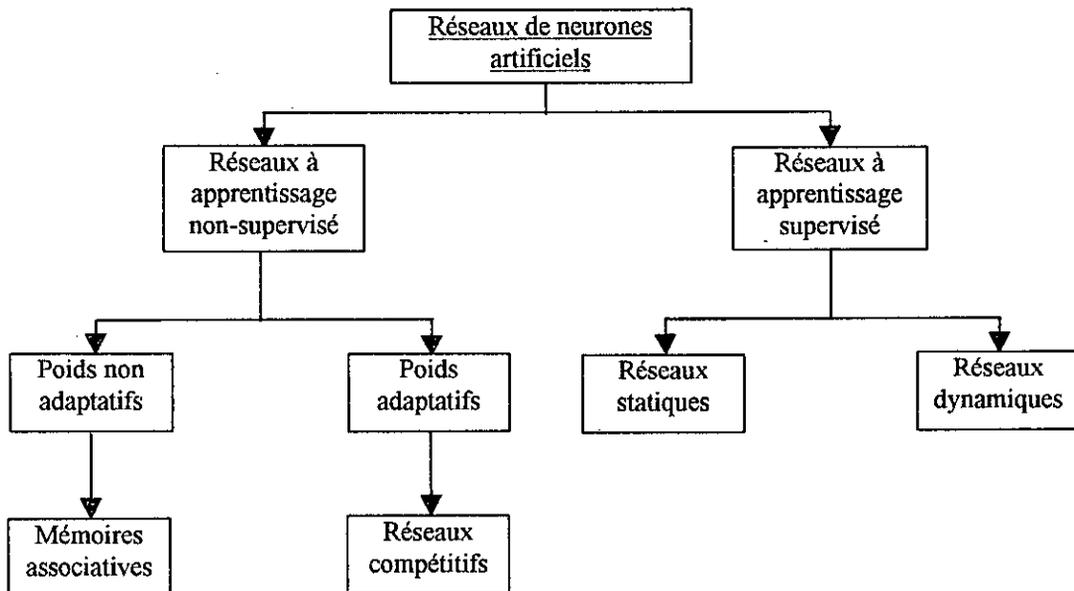


Fig.I.6 Classification des réseaux de neurones suivant leur apprentissage

I.6 Réseaux de neurones à apprentissage non supervisé

Depuis 1960 ces réseaux ont bénéficié de plusieurs travaux de recherche, et malgré le développement que connaissaient certains algorithmes de ce type tel que le Learning Matrix (LM) en 1963 et le Linear Associative Memory(LAM) en 1977, ils restaient limités puisqu'ils n'avaient pas des modèles mathématiques exacts.

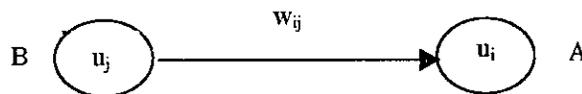
Ce n'est qu'à partir des années 80 et grâce aux travaux de Hopfield et ceux de M.Cohen et S.Grossberg que le comportement mathématique de ces réseaux a été modélisé ce qui a donné un grand élan à la recherche dans ce domaine.

I.6.1 Fondement de l'apprentissage non supervisé

Cet apprentissage est basé sur la loi de J.Hebb, il consiste à séparer l'espace d'entrée selon les classes constitutives de ce dernier.

L'énoncé de cette loi est le suivant :

Si un neurone A est stimulé à maintes reprises par un autre neurone B pendant qu'il est actif, le neurone A devient plus sensible aux stimulations en provenance de B ; le lien synaptique de B vers A devient plus significatif. Ainsi B pourra stimuler A plus facilement à l'avenir.



L'aspect mathématique de cette règle est inspiré de la méthode de corrélation, ainsi la réadaptation des poids w_{ij} entre le neurone A et le neurone B peut s'écrire sous la forme discrète suivante :

$$\Delta w_{ij} = s(u_i) s(u_j) \quad (I.4)$$

où u_i représente l'activité et $s(u_i)$ est sa sortie représentée par une fonction non linéaire appliquée à son activité.

Ce même apprentissage peut se manifester sous une deuxième forme qui traduit le fonctionnement compétitif-coopératif des réseaux de neurones, son principe réside dans la compétition entre les neurones du réseau afin de trouver le neurone gagnant qui représente le groupe dans lequel l'entrée sera classée.

I.6.2 Mémoires associatives

En général, la mémoire est un système de stockage et de récupération d'informations[5], et tous les systèmes de mémoire, qu'ils soient naturels ou artificiels, passent par trois étapes. Il faut tout d'abord alimenter le système en informations, processus habituellement appelé *encodage*; il faut ensuite un moyen de *stockage* pour conserver ces informations dans le temps et pour prévenir leur oubli; enfin il faut pouvoir accéder aux informations stockées(*récupération*). Bien que ces trois processus soient conceptuellement différents, ils sont étroitement liés, si bien qu'une modification touchant l'un aura tendance à influencer sur les autres. Le fonctionnement des réseaux de neurones n'échappe pas à cette règle. En effet, la mise en mémoire des informations à l'intérieur d'un réseau se fait grâce à un apprentissage non supervisé suivant lequel, le

calcul des poids présente un codage de l'information à stocker. Après cela, le réseau peut reconnaître les informations incomplètes ou modifiées qui lui sont présentées.

Suivant le mode de fonctionnement, on distingue deux types de mémoires associatives.

Les mémoires **auto-associatives** sont destinées à stocker des vecteurs prototypes :

$P^k = p_1^k, p_2^k, \dots, p_n^k$ tel que $k=1, 2, \dots, m$, représente les différentes classes.

Les mémoires **hétéro-associatives**, par contre, stockent les couples prototypes (X^k, Y^k) , où :

$Y^k = y_1^k, \dots, y_m^k$ tel que $k=1, \dots, m$, représente les différentes classes.

A chaque vecteur X^k en entrée, la mémoire associe le vecteur Y^k en sortie et vice versa. Dans ce qui suit, nous allons présenter les principaux types de mémoires associatives qui existent .

1.6.3 Réseaux de Hopfield

En 1982 Hopfield a étudié l'évolution de la dynamique des réseaux de neurones vers l'équilibre. En effet tout système physique évolue suivant la fonction d'énergie de Lyapunov vers un état où son énergie interne sera minimale, ainsi cette fonction a été introduite à l'intérieur du réseau pour assurer sa stabilité.

Ce réseau de Hopfield est une mémoire auto-associative constituée par une couche de neurones entièrement interconnectés (fig.I.7).

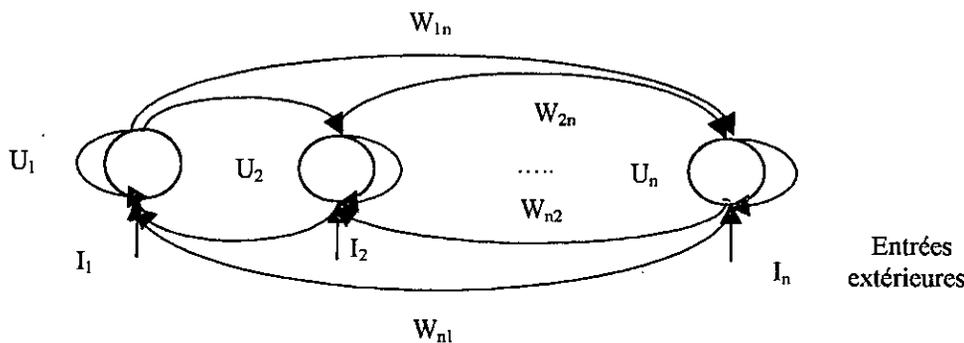


Fig.I.7 Architecture du réseau de Hopfield.

L'origine du réseau de Hopfield réside dans son modèle électrique très connu [6].

La fonction différentielle suivante traduit le fonctionnement dynamique de ce modèle :

$$u_i = \sum_{j=1}^n w_{ij} s_j(u_j) - b_i u_i + I_i \quad i=1, \dots, n \quad (I.5)$$

où w_{ij} sont les paramètres définissant les forces de liens synaptiques représentées par les conductances des connexions, et b_i représentent les conductances équivalentes à l'entrée des neurones.

La sortie des neurones est binaire comme son entrée, mais des modèles de ces réseaux avec une sortie continue utilisant la fonction sigmoïde existent, ces réseaux constituent le modèle analogique de Hopfield.

I.6.4 Mémoires associatives bidirectionnelles (BAM)

Ces réseaux constituent une extension des réseaux de Hopfield vers une mémoire hétéro-associative. En fait, ils sont composés de deux couches où l'information peut se propager dans les deux sens (fig.I.8).

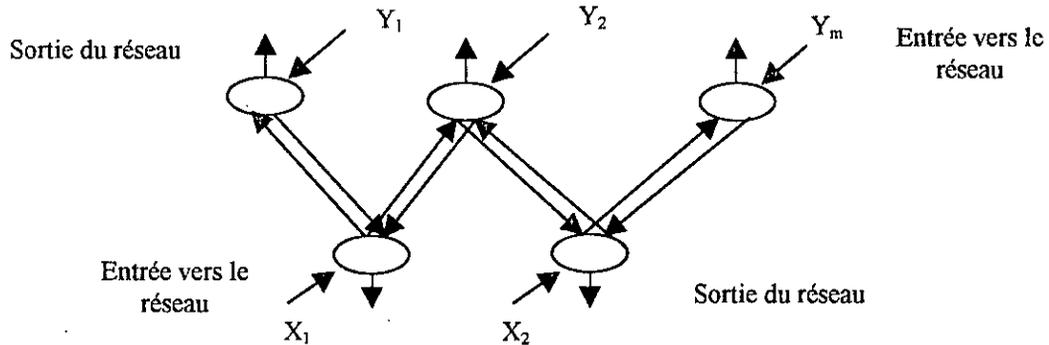


Fig.I.8 Architecture du BAM

En mémorisant les couples d'exemples (X^k, Y^k) , $k=1, \dots, m$, ce réseau associe au vecteur des entrées $X^k(X_1, \dots, X_m)$, le vecteur des sorties $Y^k(Y_1, \dots, Y_m)$ et vice versa.

Le codage de l'information est basé sur la règle de corrélation suivante :

$$w_{ij} = \sum_{k=1}^m x_i^k y_j^k \quad \text{avec } i=1, \dots, m \text{ et } j=1, \dots, m. \quad (I.6)$$

I.6.5 Mémoires associatives linéaires optimisées (OLAM)

Le réseau de Hopfield ne garantit un bon fonctionnement d'une mémoire associative que si les exemples ne sont pas corrélés c'est à dire mutuellement orthogonaux, mais cela n'est pas toujours vérifié en pratique. Alors les OLAM constituent une alternative pour le problème d'orthogonalité. ce réseau utilise la règle de projection pour l'apprentissage. cette dernière a été introduite par T.Kohonen. l'apprentissage de ces réseaux consiste à introduire une entrée égale à un des prototypes d'apprentissage stockés ou proche de lui et on doit retrouver ce prototype en sortie, ceci revient à résoudre l'équation suivante :

$$WX^k = X^k \quad k=1, \dots, m \quad (I.7)$$

Une solution de cette équation est

$$W = XX^+ \quad (I.8)$$

Où X est une matrice regroupant tous les vecteurs prototypes à mémoriser et X^+ représente la pseudo-inverse de cette matrice.

I.6.6 Réseau de Hamming

Ce réseau constitue une mémoire auto-associative. Il est surtout utilisé pour les problèmes de bruit comme dans le domaine de transmission et de communication. Il est basé sur le principe de distance de Hamming qui mesure la distance entre vecteurs dans l'espace. A chaque fois qu'un exemple est injecté, on calcul sa distance par rapport aux états mémorisés, l'état le plus proche de cet exemple sera délivré en sortie.

La mémoire de ce réseau réside dans les poids synaptiques où chaque prototype est mémorisé :

$$W_{ij} = \frac{p_j^i}{2} \tag{I.9}$$

Où p_j^i $i=1, \dots, m$ et $j=1, \dots, n$, représentent les prototypes à mémoriser.

De cette manière, la sortie de chaque neurone de cette couche est la corrélation de l'entrée avec le prototype représenté par ce neurone. Chaque neurone de la première couche calcule sa sortie qui représente son état :

$$s_i(k) = f\left(\sum_{j=1}^n w_{ij} I_j - \theta_i\right) \quad i=1, \dots, M \tag{I.10}$$

où f est la fonction Signum et θ_i le seuil du neurone i .

la couche de sortie détermine le neurone qui a l'activité la plus forte, le gagnant sera activé tandis que tous les autres inhibés. Cette couche est dynamique et chaque unité représente une classe, les poids sont calculés par la méthode *d'inhibition latérale*, où chaque neurone renforce son activité par un retour excitateur et inhibe les autres neurones autour de lui.

Le calcul des poids est donc :

$$t_{jl} = \begin{cases} 1 & j=l \\ -\varepsilon & j \neq l \end{cases} \quad j, l= 1, \dots, M \tag{I.11}$$

avec $\varepsilon < \frac{1}{M}$.

la sortie de chaque neurone de cette couche est donc :

$$s_i(t+1) = f(s_i(t) - \varepsilon \sum_{j \neq i} s_j(t)) \tag{I.12}$$

A l'équilibre cette couche délivre le gagnant qui est le seul neurone à rester activé.

I.6.7 Réseaux compétitifs

Ces réseaux présentent l'avantage que leurs poids ne sont pas fixes à l'équilibre, ils s'adaptent au fur et à mesure au cours de leur utilisation. Ils contiennent une couche compétitive en plus de la couche d'entrée où les neurones entrent en compétition entre eux afin de déterminer lequel est le plus représentatif de la classe d'entrée.

Nous présentons ci après le modèle classique de couches de neurones compétitives.

a. Couche de Grossenberg

Cette couche est constituée de neurones entièrement interconnectés, sa dynamique est régit par l'équation différentielle suivante :

$$\dot{U}_i = \mu U_i + \delta \sum_{j=1}^n S(U_j) W_{ij} + I_i \quad i=1, \dots, n, \tag{I.13}$$

où U_i est l'activité du neurone i .

I_i est l'entrée extérieure.

μ est une constante positive contrôlant la décroissance du neurone.

δ est une constante positive contrôlant les interconnexions latérales entre les neurones.

$S(U_j)$ est la sortie du neurone.

Le fonctionnement de cette couche compétitive est basé sur deux techniques importantes :

La **Self-Excitation**, où chaque neurone renforce son activité à travers le feed-back qu'il se renvoie vers son entrée, et la **Neighbor Inhibition**, où chaque neurone inhibe les autres neurones du réseau.

Lors de la présentation d'un exemple X^k au réseau, la dynamique de l'équation (I.13) fait que le neurone de la classe à laquelle cet exemple est le plus proche est activé au maximum. A l'équilibre il est le seul qui reste activé. Ce neurone gagnant est le seul à voir ses poids réadaptés. L'apprentissage se fait de la manière suivante :

$$w_{ij} = s(u_i) [-\alpha w_{ij} + \beta s(u_i)] \tag{I.14}$$

où α est un paramètre contrôlant le régime transitoire et β est le taux d'apprentissage. Dans ce qui suit nous allons faire l'étude des principaux réseaux compétitifs.

b. Adaptive résonance théorie(ART)

Ce type de réseaux (ART) est basé sur l'idée initiale introduite par Grossberg et qui a été reprise par ce dernier et Carpenter en 1987 pour former ce réseau. Le fonctionnement du ART constitue une analogie avec la résonance physique. En effet, à partir d'un faible signal, l'information fait des « aller-retour » entre la couche d'entrée et celle de sortie, constituant des revibrations, qui continuent jusqu'à ce que celle ci entre en résonance soit en adhérant à une classe déjà existante, soit en créant une nouvelle classe, suivant la finesse de la classification.

L'architecture de ce modèle de réseaux est représenté à la figure (I.9).

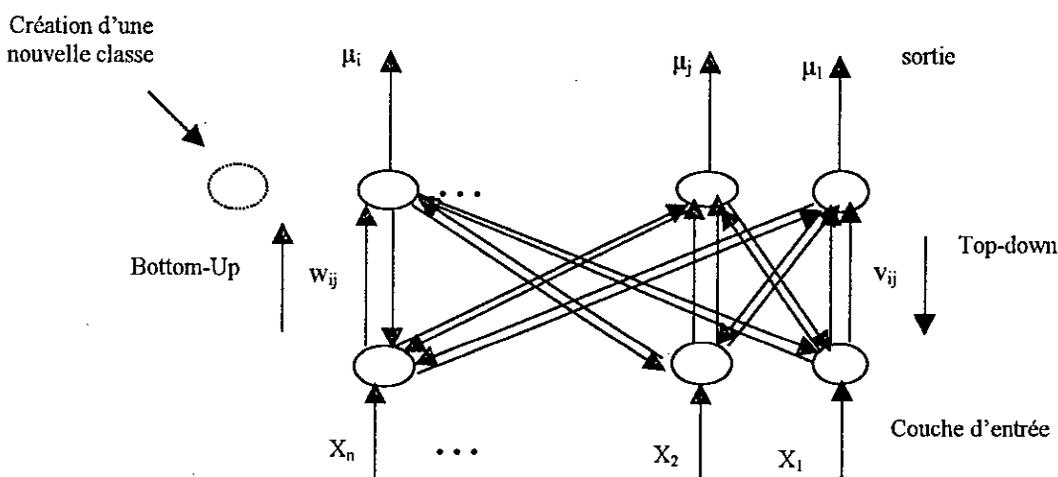


Fig.I.9 Architecture d'un réseau de neurones ART

Ce réseau est constitué de deux couches: une couche d'entée et une couche de sortie appelée Maxnet. Ce réseau a deux sens, un sens direct appelé **Bottom-up**, et un sens inverse appelé **Top-down**. Pour l'implémentation du réseau, deux cas différents ont été introduits, le premier est appelé **ART1**, il est utilisé pour des entées binaires. Le second, appelé **ART2**, est utilisé pour la classification des exemples quelconques.

c. Self organized feature Map de Kohonen.

L'idée de base de ce réseau est l'introduction dans l'apprentissage compétitif de paramètres, rendant chaque neurone dépendant des neurones qui lui sont voisins. De

cette manière, les neurones les plus proches topologiquement, sont sensibles aux mêmes caractéristiques.

Ce modèle de réseaux a été présenté par T.Kohonen en 1982, et avec le développement des algorithmes de contification vectorielle, ce réseau a connu un regain d'intérêt, et de nombreux chercheurs ont proposé des raffinements pour son apprentissage.

Architecture et fonctionnement du réseau

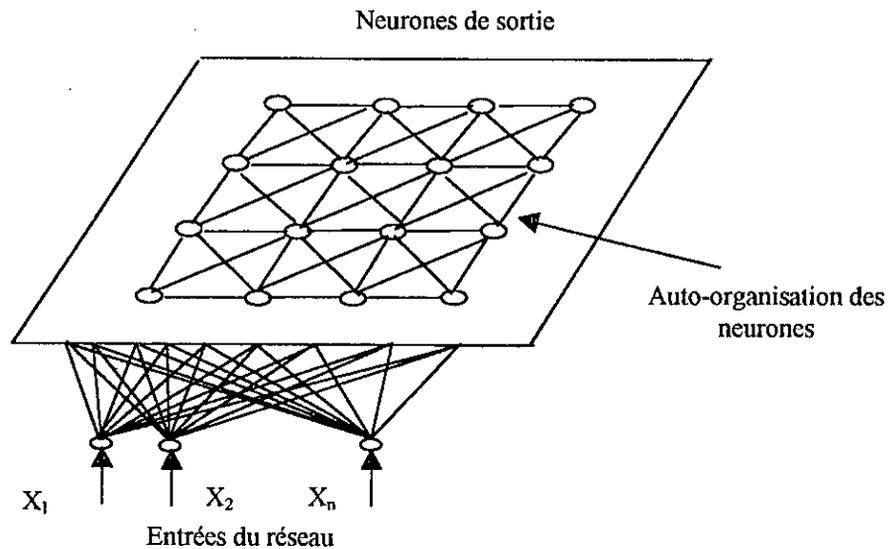


Fig.I.10 Architecture des SOM de Kohonen

L'entrée de ce réseau est un vecteur à n composants connectés tous aux k neurones de la couche de sortie. L'organisation de cette couche de neurones peut être présentée comme une grille à une ou deux dimensions dont tout les nœuds reçoivent les n composants des entrées du réseau. A l'intérieur de cette grille, chaque neurone à des voisins avec lesquels il est relié par des connexions récurrentes selon un noyau de convolution décroissant, de forme généralement schématisée par un « chapeau Mexicain ». ainsi, plus le voisin est éloigné, moins les connexions sont fortes, jusqu'à devenir inhibitrices. Les poids reliant les entrées aux sorties, formeront des centres de groupes de vecteurs d'entrées. Ces poids doivent, en fait, échantillonner l'espace d'entrée en plusieurs classes.

Ce réseau se caractérise par rapport aux autres par son aptitude à faire apparaître les caractéristiques liant les exemples en entrée. Cette faculté est due à son mode d'apprentissage compétitif qui est spécifique.

I.7 Réseaux de neurones à apprentissage supervisé

L'intérêt de cet apprentissage est de trouver des poids synaptiques optimaux afin de ramener au réseau un comportement désiré à l'aide des échantillons aléatoires $(X_i, Y_i), i=1 \dots n$.

On distingue deux approches. La première est basée seulement sur la validité de la décision. Le superviseur n'a qu'à juger la décision du réseau si elle est correcte ou pas. Cette approche permet la possibilité de classification de chaque exemple de l'espace d'entrée. La deuxième approche concerne l'optimisation d'un critère appelé fonction

coût. Elle traite en général les problèmes d'approximation de fonctions et d'optimisation.

I.7.1 Réseaux de neurones basés sur la décision

a. Le Perceptron.

Introduit pour la première fois par Rosenblat en 1958, c'est un réseau statique capable de modifier ses poids en fonction des exemples que l'on veut lui faire apprendre, l'information circule dans un sens unique de la couche d'entrée vers la couche de sortie[7].

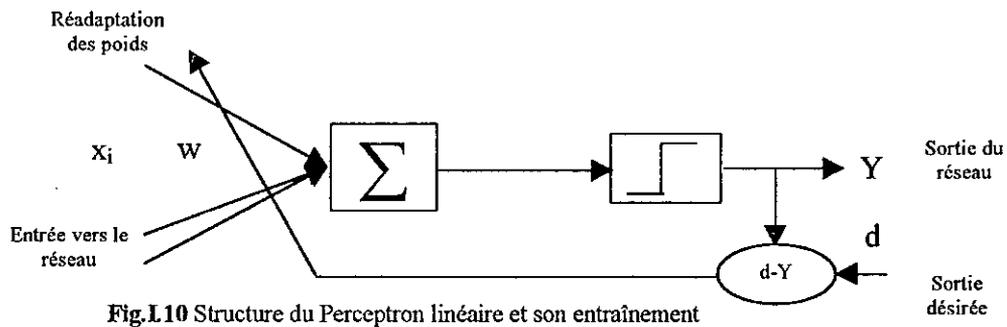


Fig.I.10 Structure du Perceptron linéaire et son entraînement

Tout les réseaux qui traitent la décision sont basés sur le perceptron, dans le cas de perceptron linéaire, l'activité du neurone est définie par l'équation linéaire suivante :

$$net_i = \sum_{j=1}^n x_j w_{ij} + \theta_i \quad i=1 \dots m. \tag{I.15}$$

où les x_j représentent les entrées du neurones j , w_{ij} le poids synaptique en provenance du $j^{ème}$ neurone de la couche d'entrée vers le $i^{ème}$ neurone de sortie et θ_i un paramètre qui agit sur le seuil en sortie.

La sortie est définie par la fonction seuil :

$$s_i(net_i) = \begin{cases} 1 & net_i > 0 \\ 0 & net_i < 0 \end{cases} \tag{I.16}$$

L'adaptation des poids de ce réseau n'est autre qu'un cas particulier de l'application de la méthode de descente du gradient que nous étudierons plus loin :

$$w(t+1) = w(t) + \Delta w(t) \tag{I.17}$$

Dans le cas de perceptron linéaire on a :

$$\Delta w(t) = \eta e(t) X(t) \tag{I.18}$$

$$= \eta (d(t) - Y(t)) X(t) \tag{I.19}$$

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

$X(t)$ et $Y(t)$ sont l'entrée et la sortie du réseau, $d(t)$ est la sortie que nous désirons obtenir et $e(t)$ l'erreur en sortie à l'instant t .

L'entraînement du réseau se poursuit jusqu'à ce que chaque exemple soit correctement classé.

L'inconvénient du perceptron est qu'il n'est efficace que pour les problèmes linéairement séparables. Dans le cas contraire, la combinaison de plusieurs perceptrons linéaires peut être utilisée pour former des zones de décision complexes.

Par ailleurs, pour avoir des espaces de classifications non linéaires, on change carrément la fonction d'activation en une fonction non linéaire telle que les fonctions de base radiale (RBF) ou elliptiques (EBF).

I.7.2 Adaline, Madaline

Introduit par Widrow et M.Hoff en 1960. L'Adaline à la même architecture que le perceptron, une couche d'entrée est connectée à une ou plusieurs sorties, dans ce dernier cas il est appelé Madaline. La différence entre ce réseau et le perceptron, est qu'il est destiné à l'approximation des fonctions. Sa sortie est linéaire(I.20).

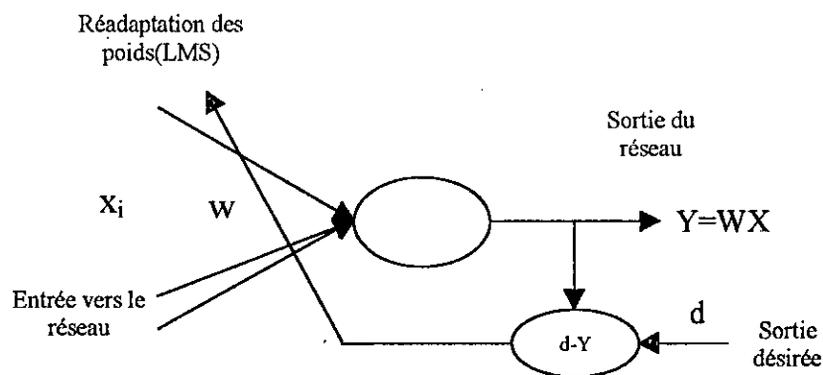


Fig.L11 Structure d'un réseau Adaline et son principe d'apprentissage

• **Apprentissage :**

A chaque fois X^k est présentée, la sortie correspondante est telle que :

$$y_i = \sum_{j=1}^n x_j w_{ij} + \theta_i \quad i=1 \dots m. \tag{I.20}$$

on définit l'erreur instantanée relative à chaque sortie par :

$$e_j^k = d_j^k - y_j^k$$

le problème donc est d'ajuster les poids W de telle sorte que l'écart entre la sortie du réseau et celle désirée soit nulle pour tout les exemples. On peut utiliser également la méthode des moindres carrés.

a. Méthode des moindres carrés(Least Mean Square Error)

Elle consiste à minimiser l'erreur quadratique définie par :

$$(e^k)^2 = (d^k - W^T X^k)^2 \tag{I.21}$$

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

$$\frac{\partial}{\partial W} E[(e^k)^2] = 0 \quad (I.22)$$

Or

$$E[(e^k)^2] = E[d^k]^2 - 2E[d^k X^k] w^T + WE[X^{(k)T} X^k] w^T \quad (I.23)$$

en posant R la matrice moyenne des auto-corrélations entre les exemples d'entrées et P la matrice des intercorrélations entre les entrées et leurs sorties désirées :

$$R = E[X^{(k)T} X^k] \quad (I.24)$$

$$P = E[d^k X^k] \quad (I.25)$$

on obtient :

$$\frac{\partial}{\partial W} E[(Y^k)^2] = -2P + 2WR \quad (I.26)$$

D'où l'on tire le vecteur poids optimal

$$W^* = PR^{-1} \quad (I.27)$$

b. Méthode de descente de gradient

Cette méthode est destinée à déterminer les adaptations qu'il faut apporter aux poids synaptiques à chaque présentation d'un exemple. On calcule le gradient de la fonction objectif (I.28) par rapport aux poids W afin d'évaluer la variation de l'erreur. Pour la faire diminuer, il suffit de parcourir la surface d'erreur dans la direction opposée du gradient. B. Widrow a adopté pour cela la méthode de Delta, dont on estime le gradient en fonction des poids synaptiques, et cela se fait à chaque fois qu'on présente un vecteur à l'entrée. Dans ce cas l'adaptation des poids fait tendre le gradient vers un minimum instantané.

L'erreur quadratique instantanée est [3] :

$$\zeta^k = \frac{1}{2}(e^k)^2 = \frac{1}{2}(d^k - y^k)^2 \quad (I.28)$$

A fin de progresser dans le sens opposé du gradient, on a donc :

$$\Delta W^k = -\eta_k \frac{\partial \zeta^k}{\partial W^k} = e^k X^k \quad (I.29)$$

A partir des équations

$$w_{ij}^{k+1} = w_{ij}^k + \eta_k (d_i^k - y_i^k) x_i^k \quad (I.30)$$

où η_k est un paramètre appelé 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'apporte rien de nouveau sur l'amélioration de ses performances. Pour le rendre sensible à cette augmentation, on doit introduire des fonction non linéarité.

I.7.3 Réseaux multicouches statiques.

Le fait que ces réseaux sont non linéaires, leurs architectures multicouches leur permettent de traiter des problèmes très complexes surtout après la mise au point de l'algorithme d'apprentissage Back-propagation en 1986.

- **Architecture du réseau.**

La topologie de ce type de réseaux est telle qu'avant d'arriver à une sortie, chaque unité d'entrée est suivie par une succession de couches cachées et chaque unité d'une couche donnée reçoit son entrée à partir de la couche précédente seulement(fig.I.12).

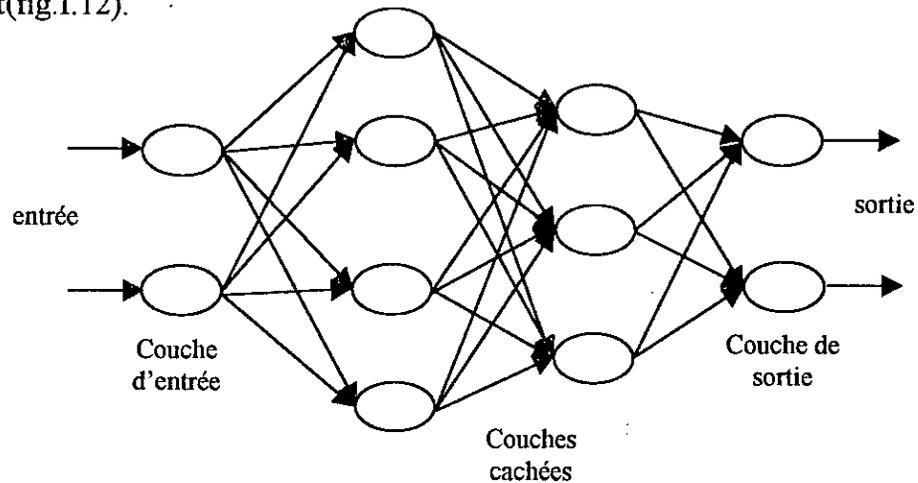


Fig.I.12 Architecture d'un réseau de neurones multicouches statique a deux couches cachées

I.7.4 Réseaux à fonction de base linéaire(LBF).

Ce type de réseaux est basé sur l'algorithme de back-propagation, il est régit par l'équation dynamique suivante :

$y_i^l(t) = f(u_i^l(t))$ et

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

(I.31)

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

$y_j^{l-1}(t)$ la sortie du $j^{\text{ème}}$ neurone dans la $l^{\text{ème}}$ couche.

l : nombre de couches.

N_l : nombre de neurones dans la $l^{\text{ème}}$ couche.

La fonction d'activation $f(u)$ est monotone non décroissante, elle peut être une sigmoïde, une tangente hyperbolique ou une Log-sigmoïde. A la sortie du réseau, la fonction linéaire est la plus utilisée pour pouvoir atteindre toutes les valeurs désirées. Pour la couche d'entrée, cette fonction n'est autre que la fonction identité.

Le terme θ_i^l est appelé *bias*, il est différent pour chaque neurone, il a pour rôle de rendre la fonction de sortie plus souple. Sa réadaptation est identique à celle des autres poids à la seule différence qu'il est considéré comme un poids relié à une entrée constante.

L'apprentissage de ce réseau consiste à minimiser un critère quadratique définissant l'écart entre la sortie du réseau et celle désirée :

$$E_p = \sum_{i=1}^{n_i} e_i^p = \sum_{i=1}^{n_i} (d^p - y_i^l)^2 \tag{I.32}$$

où d^p représente la sortie désirée pour la $p^{\text{ème}}$ entrée.

Nous présentons ici les principaux algorithmes d'apprentissage utilisés.

a. Back-propagation.

Mis en œuvre en 1972 par Werbos puis, Rumelhart, Hinton et Williams l'ont adapté aux réseaux multicouches [3].

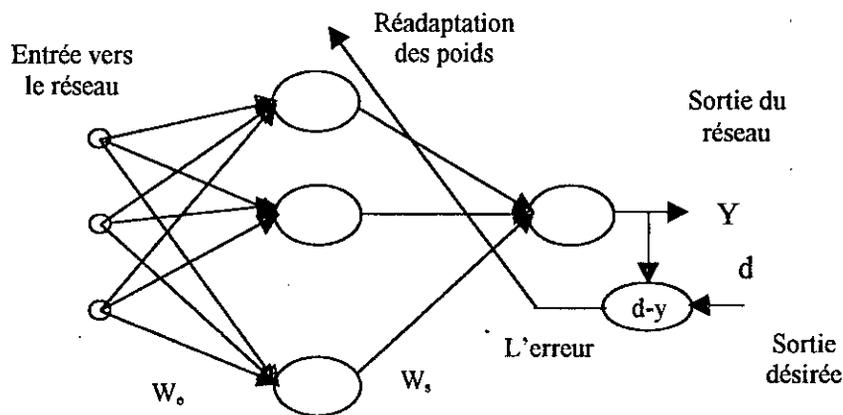


Fig.I.13 Principe de l'entraînement du réseau par rétro-propagation de l'erreur

W_0 , sont les poids synaptiques liant l'entrée avec la couche cachée, et W_s sont les poids synaptiques liant la couche cachée avec la sortie.

Cet algorithme applique la méthode de la descente du gradient mais contrairement au Madaline les poids directement liés à la sortie ont un lien directe 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.

1. Présenter un nouvel exemple X^p à l'entrée du réseau.
2. Calculer la sortie et l'erreur en sortie :

$$E_p = \sum_{i=1}^{n_i} (d_i^l - y_i^l)^2 \tag{I.33}$$

3. calculer les valeurs de réadaptation des poids :

$$\left[\Delta w'_{ij} \right]_p = \eta f'(\hat{u}_i(t)) \delta_i^l y_j^{l-1} \quad (I.34)$$

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

avec :

$$\delta_i^l = \begin{cases} (d_i^p - y_i^l) & \text{Pour les neurones de la couche de sortie.} \\ \sum_{k=1}^{n_{l+1}} \delta_k^{l+1} w_{ki}^{l+1} & \text{Pour les neurones des autres couches.} \end{cases} \quad (I.35)$$

4.

- s'il s'agit d'un data Learning (FFN Pattern), réadapter les poids :

- poser : $\Delta w'_{ij}(t) = \left[\Delta w'_{ij}(t) \right]_p$

- Aller à 5.

- s'il s'agit d'un Block Learning (FFN Batch) :

- si $p \neq M$ (M étant le nombre d'exemple d'entraînement), retourné à 1.

- sinon : $\Delta w'_{ij}(t) = \sum_{p=1}^{p=M} \left[\Delta w'_{ij}(t) \right]_p$ (I.36)

5. réadapter les poids

$$w'_{ij}(t+1) = w'_{ij}(t) + \Delta w'_{ij}(t) \quad (I.37)$$

6. répéter de 1 à 5. Autant de fois que nécessaire jusqu'à convergence vers le seuil d'erreur fixé.

Comme le montre l'algorithme ci-dessus, on peut opérer avec cette méthode de deux manières différentes : data adaptive learning, ou block adaptive learning, que nous examinerons successivement.

b. Data adaptive learning (FFN pattern)

Lors de cette méthode, la réadaptation des poids se fait au passage de chaque exemple ce qui rend le processus sensible à chaque exemple individuel, et par suite sensible aux bruits, cette technique n'est donc utilisée que pour un apprentissage en temps réel.

c. Block adaptive learning (batch learning)

La réadaptation des poids se fait après le passage de tous les exemples, le réajustement se fait donc suivant la moyenne de tous les exemples ce qui le rend insensible aux bruits, et son application ne nécessite pas un apprentissage en temps réel.

I.7.5 Etude du taux d'apprentissage

L'algorithme de backpropagation impose une variation infinitésimale des poids à chaque itération. Cette variation est contrôlée par le taux d'apprentissage; s'il est trop petit, il assure la convergence mais très lentement; par contre s'il est grand, la surface

d'erreur sera parcourue d'une façon « téméraire » ce qui engendrera une instabilité dans la recherche du minimum.

I.7.6 Variantes de la backpropagation

a. Fast backpropagation

Le plus important pendant les premières itérations est la réduction de la fonction erreur. Dès qu'on arrive à cette étape, la convergence de l'algorithme devient de plus en plus lente. Afin d'accélérer la convergence de nouveau, N.B.Karayannis et A.N.Venetsanopoulos proposent d'appliquer l'algorithme de backpropagation avec une fonction objective dite généralisée. Le critère proposé est décrit par [8]:

$$G(\lambda) = \lambda E_1 + (1-\lambda)E_2 \quad (I.38)$$

$$= \lambda \sum_{k=1}^M \sum_{i=1}^{n_i} \frac{1}{2} (e_i^k)^2 + (1-\lambda) \sum_{k=1}^M \sum_{i=1}^{n_i} \Theta(e_i^k)$$

pour un réseau LBF, on prend :

$$\Theta(d_i^k - y_i^k) = d_i^k (d_i^k - y_i^k) \quad (I.39)$$

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 :

$$W_{ij}^{k+1} = W_{ij}^k - \eta \frac{\partial G(\lambda)}{\partial W_{ij}^k} \quad (I.40)$$

au début de l'apprentissage, on prend $\lambda=1$, puis on le fait diminuer jusqu'à ce qu'il s'annule à la fin de l'apprentissage en fonction de l'erreur selon la formule suivante :

$$\lambda = \lambda(E) = \exp\left(-\frac{\mu}{E^n}\right) \quad (I.41)$$

avec μ un nombre positif réel, et n un entier positif.

b. Robust Backpropagation

Elle est caractérisée par son insensibilité aux perturbations affectant les échantillons grâce à des estimateurs robustes. Dans le domaine de la statistique, on trouve des lois de densité réputées, telles que la loi de Gauss ou celle de Cauchy, ou la densité de probabilité basée sur la fonction Logistique.

$$s(E_p) = \tanh\left(\frac{E_p}{2}\right) \quad (I.42)$$

ou sur la fonction de Cauchy :

$$S(E_p) = \frac{2E_p}{1 + E_p^2} \quad (I.43)$$

c. Backpropagation avec Momentum

C'est la version la plus répandue de la backpropagation. D.E.Rumelhart a proposé une solution qui consiste à utiliser les changements précédents des poids pour la réadaptation des poids actuels, donc l'équation d'adaptation devient :

$$w'_{ij}(t+1) = w'_{ij}(t) + \eta \Delta w'_{ij}(t-1) + \alpha w'_{ij}(t-1) \quad (I.44)$$

Le paramètre α est utilisé pour pondérer l'effet du terme ajouté qui est appelé momentum, il est compris entre 0.8 et 0.9.

Cette méthode permet de faire sortir les poids des minima locaux afin d'aboutir à un minimum global.

d. Méthode d'optimisation aléatoire(ROM)

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, ces méthodes ne dépendent nullement de la nature mathématique du système sur lequel elles sont appliquées, leur grand avantage est qu'elle ne nécessite pas la détermination du gradient du critère à minimiser. D'autre part, elle permet d'échapper aux minima locaux.

En 1965, J.Mathias a proposé une méthode aléatoire d'optimisation, et qui a été développée ensuite par F.J.Sollis et J.B.Wetts en 1981, qui ont montré la convergence vers un minimum global.

Nous présentons ci dessous l'algorithme de cette méthode.

- **Algorithme de la ROM**

Cette méthode utilise un bruit Gaussien pour générer les séquences aléatoires nécessaires pour la réadaptation des poids du réseau.

1. Initialiser tous les poids du réseau aléatoirement. Initialiser, la variance v et la moyenne de la séquence $b(0)=0$.
2. Générer une nouvelle séquence aléatoire $\delta(k)$ de moyenne $b(k)$ et de variance $v(k)$.

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

$$E(W(k)), E(W(k)+\delta(k)) \text{ et } E(W(k)-\delta(k))$$

4.

- Si $E(W(k)+\delta(k)) < E(W(k))$, alors

$$W(k+1) = W(k) + \delta(k) \text{ et}$$

$$B(k+1) = 0.4\delta(k) + 0.2b(k) \quad (I.45)$$

- Si $E(W(k)+\delta(k)) \geq E(W(k))$ et $E(W(k)-\delta(k)) < E(W(k))$, alors

$$W(k+1) = W(k) - \delta(k) \text{ et}$$

$$B(k+1) = b(k) - 0.4\delta(k) \quad (I.46)$$

- Sinon,

$$W(k+1) = W(k)$$

$$b(k+1) = 0.5b(k)$$

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

Le grand avantage de cette méthode est qu'elle est indépendante du critère d'optimisation contrairement à la méthode de backpropagation où le calcul du jacobien est nécessaire. L'utilisation de la ROM dépend de la variance du vecteur Gaussien. Une séquence avec une grande variance génère de grands changements dans l'erreur mais jamais des oscillations. Le cas échéant, elle marque des stagnations sur la surface d'erreur jusqu'à l'arrivée d'une séquence favorable. 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.

I.7.7 Réseaux LBF et approximation de fonction

L'approximation de fonction constitue l'un des problèmes les plus importants en mathématique. Les plus anciens problèmes connus est l'approximation d'une fonction continue à plusieurs variables par une combinaison de fonctions continues à une seule variable. En 1965 Spretcher a énoncé un théorème qui est en fait un développement du théorème de Kolmogorov de 1957, stipulant que toute fonction continue sur l'intervalle $[0, 1]^n$ peut être présentée, en utilisant seulement une combinaison linéaire de fonctions non linéaires continues et croissantes, à une seule variable :

$$\forall n \geq 2, \exists \varphi: [0, 1] \rightarrow [0, 1], \varphi \text{ croissante}$$

$$\forall \eta \geq 0, \exists \varepsilon \ 0 < \varepsilon < \eta \ \forall f \text{ sur } I^n$$

$$f(x) = \sum_{j=1}^{2n+1} \chi_j \left(\sum_{i=1}^n \lambda_i \varphi(x_i + \varepsilon(j-1)) + j - 1 \right) \tag{I.47}$$

où χ est une fonction continue et φ est une fonction continue et monotone. Pour un réseau de neurones, l'équation (I.47) contenant des doubles sommations peut être interprétée comme un réseau ayant deux couches cachées possédant respectivement $n(2n+1)$ et $(2n+1)$ neurones. A la base de cela, H.Nielson a démontré en 1987 que n'importe quelle fonction peut être approximée par un réseau à quatre couches. Mais la nature des fonctions d'activation de ces neurones était encore ambiguë. C'est Kreinovich, qui en se basant sur le théorème de Weishtrass sur l'approximation de fonctions multivariées linéaires par des régressions polynomiales, a prouvé que ces fonctions doivent être continues, monotones et dérivable. Ceci a orienté la plupart des chercheurs vers la fonction sigmoïde qui, en plus, est en parfait accord avec le modèle de Mc.Culloch et Pitts. C'est la classe de réseaux basés sur la fonction discriminant linéaire LBF.

Enfin Cybenko et Funahashi ont émis un théorème en 1991 selon lequel, si $f(x)$ est une fonction continue croissante et réelle, et $t(x)$ une fonction continue non constante dans un sous espace S dans \mathcal{R} alors :

$$\forall \varepsilon > 0, \exists K \in \mathbb{N}, w_{ij} \in \mathcal{R}, \theta_i \in \mathcal{R} \text{ tel que :}$$

$$\phi = \sum c_i f \left(\sum w_{ij} x_j + \theta_i \right) \tag{I.48}$$

satisfaisant : $|\phi(X, W) - t(X)| < \varepsilon$ pour $t, X \in S$

Ce dernier théorème prouve que les LBF sont capables d'approximer n'importe quelle fonction continue dans l'espace.

I.7.8 dimension du réseau, l'entraînement et le dilemme Précision-Généralisation

la dimension d'un réseau permettant d'obtenir un meilleur résultat est impossible à fixer. Il a été établi qu'un réseau à une seule couche cachée, peut faire l'approximation de n'importe quelle fonction, mais le nombre de neurones cachés n'a jamais pu être spécifié pour un problème donné.

Plusieurs travaux ont été menés sur cette question. E.B.Baum et Haussler ont proposé une technique où le nombre de neurones dépend du nombre d'entrées n_x , du nombre de sorties n_y , et de celui d'exemples n_p .

D'où, un intervalle où le nombre de poids n_w nécessaires peut varier, est donné par l'expression suivante :

$$\frac{n_y n_p}{1 + \log_2(n_p)} \leq n_w \leq n_y \left(\frac{n_p}{n_x} + 1 \right) (n_x + n_y + n_p) + n_y \quad (I.49)$$

En appliquant ce résultat, le nombre de neurones peut devenir trop important avec l'augmentation du nombre des exemples d'entraînement.

Il faut cependant savoir qu'un nombre important de neurones, bien que compliquant les calculs, permet d'obtenir une bonne **précision d'entraînement**. Un nombre insuffisant par contre entraînera sa dégradation.

La précision d'entraînement ne concerne, malheureusement, que les exemples avec lesquelles le réseau a été entraîné, or l'apprentissage est surtout destiné à rendre le réseau capable de répondre correctement aux exemples non présents pendant l'entraînement : c'est la **Généralisation**.

Un réseau trop chargé en neurones cachés souffre de manque de généralisation, échouant ainsi devant des exemples autres que ceux présents dans la base d'apprentissage. Un nombre réduit de neurones qui devrait être favorable à la généralisation rend la précision d'entraînement très difficile à obtenir, si ce n'est impossible. Ainsi augmenter la dimension est favorable à la précision d'entraînement et défavorable à la généralisation. Sa diminution en revanche améliore la généralisation mais au détriment de la précision. Ce problème constitue le dilemme **précision-généralisation**. Il faut donc réussir un compromis entre les deux.

D'autre part un apprentissage misant beaucoup sur la précision d'entraînement, visant donc à trop minimiser l'erreur en sortie conduit à un **surentraînement**. Ce phénomène rend le réseau incapable de généraliser. En effet, « en poussant » l'entraînement très loin, le réseau se spécialise sur les exemples d'apprentissage auxquels une précision très importante lui a été demandée, lui retirant ainsi son pouvoir d'interpolation.

Un autre problème important dont souffre les réseaux LBF est la difficulté de subir l'apprentissage avec de nouveaux exemples sans perdre les caractéristiques déjà acquises. En effet, la fonction d'activation des neurones qui est non locale favorise ce phénomène d'oubli appelé *effet de non entraînement* et rend ces réseaux souvent incapables d'apprendre de nouveaux exemples après avoir déjà subi un premier apprentissage. Ceci les rend mal adaptés pour un apprentissage en temps réel pendant le déroulement de processus, notamment dans le cas de variation de leurs paramètres.

I.7.9 Réseaux à fonction de base radiale RBF

a. Principe de base

Ce sont des réseaux à une seule couche cachée, ils sont utilisés pour la classification et l'approximation de fonction, à travers une combinaison linéaire des

fonctions non linéaires de base radiale. Le fonctionnement de ces réseaux repose sur le principe des estimateurs à noyau, et le généralise du monovarié vers le multivarié [9] [2].

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

$$\phi_i(x) = \phi_i(\|x - \xi_i\|) \tag{I.50}$$

l'idée principale est introduite par M.J.D.Powell. elle est basé sur l'interpolation. Toute fonction $f(x)$ d'une variable $x \in R^n$ peut être approchée par une interpolation composée par la somme de p fonctions noyaux de forme fixée $\phi(x)$:

$$f(x) = \sum_{j=1}^p \lambda_j \phi(\|x - \zeta_j\|) \tag{I.51}$$

où ζ_i représentant les nœuds de l'interpolation pour $i=1, \dots, n$.

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

$$f(x_i) = y_i \quad i=1, \dots, n. \tag{I.52}$$

et $\phi(x)$ est une fonction assurant la continuité aux nœuds et la dérivabilité d'ordre supérieur en ces points. Dans le cas général d'une interpolation, une fonction polynomiale peut, par exemple, être choisie.

b. Architecture et fonctionnement des réseaux RBF

Broomhead et Lowell puis Che et al, ont conçu un réseau de neurones à une seule couche cachée dont le fonctionnement est basé sur l'idée des approximateurs à noyaux introduite ci- dessus. Ainsi à partir de l'équation (I.51) et en la considérant dans le cas multivarié, la sortie que doit délivrer ce réseau est de la forme (fig.I.14) :

$$f_k(x) = \sum_{j=1}^n \lambda_{kj} \phi(\|x - c_j\|) \quad k=1, \dots, n ; i=1, \dots, m \tag{I.51}$$

Où c_i sont des valeurs, que nous attribuons à chaque neurone de la couche cachée et appelées centres des classes, λ_{kj} sont les poids synaptiques reliant ces neurones aux sorties.

La fonction noyau ϕ qui doit toujours assurer la continuité et la dérivabilité aux points de jonction est radialement symétrique. Ainsi, la contribution de cette dernière change en fonction de la position de l'entrée par rapport aux centres. Par ailleurs, elle doit produire des réponses localisées, dont les valeurs ne sont significatives que dans un certain intervalle de l'espace des entrées.

Parmi ces fonction, on peut trouver les formes suivantes :

Forme Cubique	$\phi(r) = r^3$	
Forme Multiquadratique	$\phi(r) = (r^2 + k^2)^{\frac{1}{2}}$	
Forme Logarithmique décalée	$\phi(r) = \log(r^2 + k^2)$	(I.54)
Forme Gaussienne	$\phi(r) = \exp(-r^2 / \beta^2)$	

Où r représente un réel quelconque.

Parmi ces formes, la fonction Gaussienne est la plus utilisée en raison de ses caractéristiques, que nous avons déjà présentées.

C'est cette fonction que nous utiliserons dans notre travail.

La figure(I.14) représente l'architecture d'un réseau RBF.

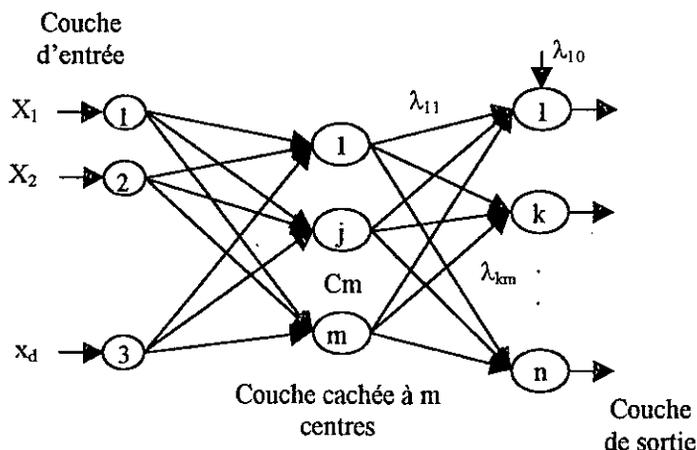


Fig.I.14 Architecture d'un réseau RBF Multi-sorties

Celui-ci est composé de trois couches (fig.I.14). La première couche reçoit les entrées injectées au réseau. La couche cachée est constituée de neurones, dont le discriminant est à fonction radiale de base. Chacun de ces neurones est doté d'un vecteur c_i appelé centre, chaque entrée du réseau est rangée dans l'une des classes représentées par les neurones de la couche cachée. Les valeurs des centres doivent donc, représenter chacune le barycentre de la classe qui lui correspond. Plus le nombre d'exemples d'entraînement est important, plus cette couche est chargée de neurones.

Lorsqu'un vecteur X est présenté à l'entrée du réseau, chaque neurone calcule, à travers son discriminant, la distance euclidienne de cet exemple par rapport au centre de la classe qu'il représente :

$$d_i = \|X - c_i\| \tag{I.55}$$

La sortie de chaque neurone de cette couche est calculée en appliquant la fonction noyau :

$$s_i(X) = \phi(\|X - c_i\|) \quad i=1, \dots, m, \tag{I.56}$$

m étant le nombre de neurones cachés.

Cette fonction est généralement Gaussienne. D'où l'on a :

$$s_i(X) = \exp\left[-\frac{(X - c_i)^T (X - c_i)}{2\sigma_i^2}\right] \tag{I.57}$$

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 cette fonction 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 ainsi l'espace d'entrée, les aptitudes de ce réseau se trouvent en approximation 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, qui est constituée de neurones linéaires effectue l'approximation définie par la somme pondérée suivante :

$$y_k(X) = \sum_{i=1}^m \lambda_{ki} \phi(\|X - c_i\|) \quad k=1, \dots, n, i=1, \dots, m \quad (I.58)$$

n nombre de neurones de sorties.

c. Apprentissage des réseaux RBF [10] [11]

L'entraînement d'un réseau RBF comprend deux étapes :

- L'apprentissage de la couche cachée : sélection des centres
- 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 constitue la partie la plus simple de l'apprentissage. En effet la sortie étant linéaire, n'importe quel algorithme d'optimisation linéaire (LMS, par exemple) peut être aisément utilisé.

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

- Apprentissage non supervisé basé sur un algorithme de regroupement.
- Apprentissage supervisé, utilisant les algorithmes d'entraînement basé sur l'optimisation.
- Apprentissage supervisé de regroupement (*clustering*) hiérarchique.

Nous présentons ci-dessous un algorithme pour chacun des deux cas.

d. Méthode de Centrage Adaptative (Adaptive Centering Method)

Comme pour les paramètres de pondération λ_{ij} , à toute valeur attribuée aux centres c_i correspondra une certaine erreur en sortie. Il est aisé de remarquer, à partir de l'équation (I.58), que cette fonction est aussi bien dérivable par rapport aux centres qu'aux poids synaptiques.

Ainsi tous ces paramètres peuvent donc être ajustés en utilisant la méthode de descente du gradient jusqu'à obtention d'un minimum satisfaisant.

Nous présentons ci-dessous les étapes de cette technique d'apprentissage supervisé.

1. Initialiser les centres $c_i(0)$, les paramètres de la sortie Gaussienne de chaque neurone de la couche cachée $\sigma_i(0)$, $i=1, \dots, m$, et les poids synaptiques $\lambda_{ij}(0)$, $j=1, \dots, n$.
2. Choisir un tau d'entraînement $\alpha < 1$ variable ou fixe.
3. Présentation successive de P exemples d'entraînement.

Calcul de l'erreur quadratique en sorties E entre la réponse désirée et celle fournie par le réseau.

$$E = \sum_{p=1}^M \sum_{j=1}^n (d_j^p - y_j^p)^2 \quad (I.59)$$

avec M représentant le nombre d'exemples d'entraînement. y_j^p et d_j^p représentent la j^{ème} sortie du réseau et celle désirée pour le p^{ème} vecteur d'entrée.

4. réajuster les centres et éventuellement les paramètres des fonctions noyau :

$$\begin{aligned} C_i(t+1) &= C_i(t) + \alpha \Delta C_i(t) \\ \sigma_i(t+1) &= \sigma_i(t) + \alpha \Delta \sigma_i(t) \end{aligned} \quad (I.60)$$

avec :

$$\begin{aligned} \Delta C_i(t) &= -\frac{\partial E}{\partial C_i(t)} \\ \Delta \sigma_i(t) &= -\frac{\partial E}{\partial \sigma_i(t)} \end{aligned} \quad i=1, \dots, m$$

5. réajuster les poids synaptiques.

$$\lambda_{ij}(t+1) = \lambda_{ij}(t) + \alpha \Delta \lambda_{ij}(t) \quad (\alpha : \text{taux d'apprentissage}) \quad (I.61)$$

avec

$$\Delta \lambda_{ij}(t) = -\frac{\partial E}{\partial \lambda_{ij}} \quad j=1, \dots, m, \quad i=1, \dots, m.$$

6. répéter les étapes 3 à 5, un nombre de fois nécessaire jusqu'à obtention de la précision désirée en sortie.

Cet algorithme simplifie l'apprentissage, dans le sens où les poids et les centres sont réadaptés directement ensembles, et ceci en dépit du calcul des dérivées partielles de l'erreur par rapport à chaque centre. Ce calcul ne pose pas de problème puisque le réseau ne contient qu'une seule couche cachée. Les dérivées se calculent d'une manière automatique, contrairement aux réseaux LBF multicouches.

Une autre technique, ne nécessitant pas le calcul des dérivées par rapport aux centres, peut être utilisée. Elle repose sur le partage de l'espace des exemples en sous espaces, représentant chacun une classe différente.

Nous présentons ci-dessous cette méthode.

e. Méthode basée sur l'algorithme de regroupement (clustering algorithm)

Cet algorithme comprend un apprentissage non supervisé de classification, basé sur l'approche du **Nearest Neighborhood**.

Les étapes de l'algorithme d'apprentissage du RBF dans ce cas, sont les suivantes :

1. Initialiser les centres $C_i(0)$, $i=1, \dots, m$, avec des valeurs aléatoires et choisir un taux d'apprentissage initial $\alpha(0) < 1$.

2. Calculer la distance euclidienne de l'exemple X^p par rapport à chaque classe.

$$\text{Dist}_i(p) = \|X^p - C_i(p-1)\| \quad i=1, \dots, m \quad (I.62)$$

3. noter l'argument k du centre pour lequel la distance est minimale :

$$k = \arg[\min(\text{dist}_i(p))] \quad i=1, \dots, m \quad (I.63)$$

tel que $\arg[.]$ identifie le rang du neurone (i) dans la couche cachée.

4. réadapter les centres :

$$\begin{aligned} C_i(p) &= C_i(p-1) & i=1, \dots, m \text{ et } i \neq k \\ C_i(p) &= C_i(p-1) + \alpha(p)[X^p - C_i(p-1)] & i=k \end{aligned} \quad (I.64)$$

5. réduire le taux d'apprentissage :

$$\alpha(p) = \frac{\alpha(p-1)}{1 + \text{int}[\frac{P}{M}]^{1/2}} \quad (I.65)$$

($\text{int}[.]$: partie entière). Avec M le nombre d'exemples que contient la base d'apprentissage.

6. refaire les étapes de 1 à 5 jusqu'à ce que chaque exemple soit classé et que les classes ne changent plus.

Pour les poids synaptiques, la méthode des moindres carrés peut être aisément utilisée après la détermination des centres.

Contrairement au premier, cet algorithme détermine d'abord les centres, puis les poids synaptiques. L'apprentissage non supervisé utilisé pour la détermination des centres, peut limiter le nombre de neurones, suivant le nombre de la classes auquel cet apprentissage a abouti. L'algorithme supervisé, par contre, nous oblige à fixer le nombre de neurones dès le début. Cette méthode nous permet donc, d'appliquer les techniques d'apprentissage compétitif.

Il existe une autre famille d'algorithmes d'entraînement plus rapides et plus économiques en nombre de neurones, ils sont basés sur la classification, appelée *Algorithmes de regroupement hiérarchique*.

f. Apprentissage supervisé de Regroupement Hiérarchique

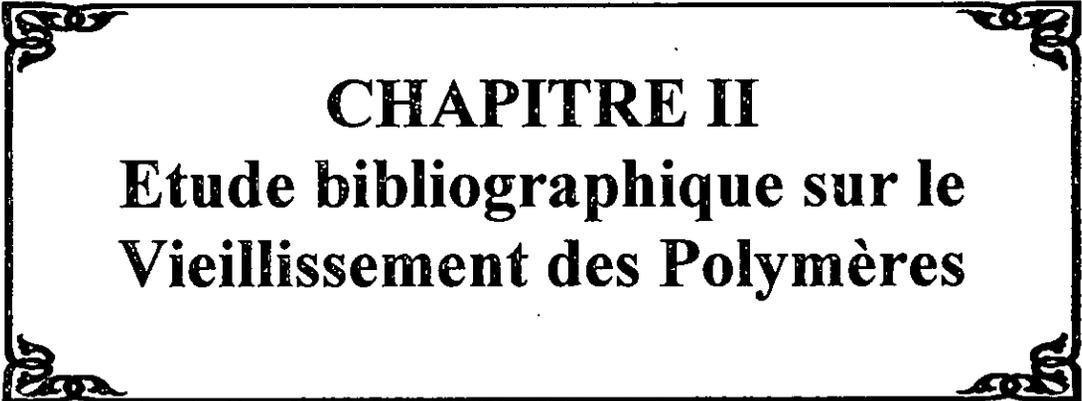
Ces méthodes, basées sur la classification des exemples d'entraînement, nécessitent la présence des couples d'entrées et leurs sorties désirées. Elles permettent, tout en procédant avec un apprentissage supervisé, de rendre l'architecture du réseau évolutive ; ce qui permet de limiter le nombre de neurones dans la couche cachée et d'accélérer l'apprentissage.

Il existe deux techniques différentes pour ce genre d'apprentissage, la première consiste à commencer l'apprentissage avec un seule neurone dans la couche cachée, et augmenter leur nombre par la suite, pendant qu'avec la seconde, l'apprentissage commence avec un nombre important de neurones, pour le faire diminuer pendant l'apprentissage.

g. Réseaux RBF et Approximation de fonctions

Les réseaux RBF constituent des modèles des réseaux très efficaces notamment pour l'approximation de fonctions. L'utilisation de la fonction Gaussienne permet de bénéficier de sa caractéristique locale pour faciliter l'apprentissage et améliorer l'approximation. Par ailleurs, la procédure d'entraînement, basée sur le fonctionnement de ces réseaux, qui consiste à une classification suivie d'une optimisation rend l'apprentissage beaucoup moins difficile et plus rapide que celui des réseaux LBF. Par ailleurs, ces réseaux sont toujours utilisés avec une seule couche cachée. Ceci libère l'utilisateur du choix du nombre de couches.

En outre, afin d'alléger l'apprentissage, il est nécessaire de fixer les paramètres de dispersion dès le début de l'apprentissage à une valeur fixe commune pour tous les neurones. Dans ce cas, il faut savoir qu'une valeur très petite de ce paramètre nécessite beaucoup de neurones dans la couche cachée, ce qui peut diminuer les capacités de généralisation du réseau. Un paramètre de valeur très importante, par contre, peut entraîner un chevauchement entre les classes rendant ainsi l'apprentissage impossible.



CHAPITRE II
Etude bibliographique sur le
Vieillissement des Polymères

II.1. INTRODUCTION.

Les câbles isolés souterrains sont principalement employés, pour le transport et la distribution de l'énergie électrique dans les zones fortement urbanisées aux abords ou à l'intérieur des grandes villes. Toutefois les câbles souterrains sont de plus en plus utilisés en moyenne tension même en zones rurales ou semi-rurales. De plus des progrès récents en haute tension faciliteront le transport en souterrains dans un avenir proche[14].

Au cours des dernières années, l'isolation électrique des câbles avec du papier imprégné d'huile a laissé peu à peu la place à une isolation faisant appel à des polymères(PE, PRC, PRS, PVC, EPDM, etc.). Les câbles à isolation synthétique parfois appelée plastique, présentent des avantages non négligeables, à la fois sur le plan technique et sur le plan économique. Parmi ces avantages nous pouvons citer la grande facilité de mise en œuvre de ces câbles, leur volume réduit, leur exploitation aisée et la non nécessité de leur entretien du fait que contrairement aux câbles isolés au papier imprégné, il n'existe aucun problème de fuite d'huile diélectrique. Ajoutées à ces qualités, de meilleures propriétés électriques et mécaniques ainsi qu'une durée de vie largement supérieure à celles des isolants naturels font des isolants polymères ceux qui répondent le mieux aux exigences actuelles.

Toutefois, bien que l'isolation aux polymères présente des avantages sérieux, ils n'en demeure pas moins que ces matériaux présentent certains inconvénients parmi lesquels nous pouvons citer les pertes diélectriques importantes à partir d'une certaine valeur de la tension d'utilisation. La chaleur, le champ électrique ainsi que les radiations peuvent avoir, à long terme, un effet néfaste sur les isolations en polymère. Ainsi au cours de leur utilisation ces isolants peuvent être amenés à perdre leurs qualités. Ce phénomène est connu sous le nom de vieillissement du matériau[15].

On appelle vieillissement tout phénomène se traduisant par une évolution des propriétés du matériau. Cette évolution peut résulter de la modification de la structure des macromolécules. L'étude du phénomène de vieillissement thermique d'un isolant solide a deux buts principaux. Le premier est d'observer l'évolution de la durée de vie de ces isolants sous l'effet des contraintes thermique. Le second but est de trouver une éventuelle corrélation entre le processus de vieillissement et les contraintes qui le provoquent [16] tel que l'effet de la chaleur. Ce dernier peut se manifester de deux manières différentes; la première concerne l'effet instantané sur le matériau dû à une forte élévation de la température; la seconde concerne quant à elle l'effet d'une température pas trop élevée sur le matériau lorsque ce dernier est exposé à la chaleur pendant un temps assez long.

Les conséquences qui peuvent se manifester à cause de ce vieillissement sont en général : un rétrécissement du matériau et un changement de couleur, ainsi qu'une perte de masse. Lorsque les temps d'exposition à la chaleur deviennent assez longs, nous pouvons assister à la variation de la composition chimique du matériau. Cette variation est due généralement à la libération de certains constituants ou produits de dégradation. Ces effets peuvent également être accompagnés d'une augmentation des pertes diélectriques, ainsi que d'une réduction des propriétés mécaniques du matériau[15].

II.2. POLARISATION DES DIELECTRIQUES.

II.2.1.Principe de la polarisation.

Selon le modèle des bandes d'énergie, la matière devient diélectrique, c'est à dire mauvaise conductrice de l'électricité, quand les bandes de conduction et de valence sont séparées par une énergie supérieure à 5 eV. Alors, un très petit nombre d'électrons reçoivent, à température ordinaire, l'énergie thermique nécessaire pour effectuer une transition dans la bande de conduction. Quand la température s'élève la probabilité de transition augmente. Le fait que les charges présentes dans un diélectrique ne soient pas libres ne signifie nullement qu'elles sont liées de façon absolument rigides les unes aux autres. En particulier, un champ électrique appliqué E déplace légèrement les charges positives et négatives les unes par rapport aux autres, provoquant l'apparition des dipôles électriques. Plusieurs mécanismes concourent à l'apparition de ces dipôles. Sous l'action d'un champ électrique alternatif, cette polarisation ne suit pas le champ électrique d'une manière instantanée (déphasage entre le champ électrique et l'établissement des dipôles). Cette relaxation provoque la dissipation d'énergie responsable des pertes diélectriques[14].

1. Echelle macroscopique.

Dans un espace contenant des charges libres, le champ électrique E et l'induction D sont liées par la relation II.1 :

$$D = \epsilon_0 E \quad (\text{II.1})$$

Où ϵ_0 est la permittivité du vide.

Si l'on introduit un diélectrique dans un volume d'espace, il faut tenir compte également des charges dues aux phénomènes de polarisation. L'induction dans le diélectrique satisfait à la relation II.2 :

$$D = \epsilon_0 E + P \quad (\text{II.2})$$

Où P est la polarisation.

Dans un milieu linéaire, homogène et isotrope,

$$P = \epsilon_0 \chi E \quad (\text{II.3})$$

En tenant compte de l'expression II.2 on en déduit :

$$D = \epsilon_0 (1 + \chi) E \quad (\text{II.4})$$

Où χ est la susceptibilité électrique.

et en posant :

$$1 + \chi = \epsilon_r \quad \text{et} \quad \epsilon_0 \epsilon_r = \epsilon$$

on obtient :

$$D = \epsilon E \quad (\text{II.5})$$

ϵ_r est la permittivité relative et ϵ permittivité absolue.

La polarisation P s'écrit aussi :

$$P = \epsilon_0 (\epsilon_r - 1) E \quad (\text{II.6})$$

2. Echelle microscopique.

Dans un diélectrique polarisé, il existe un champ électrique local E_1 ; il induit un moment dipolaire P au niveau de chaque molécule :

$$P = N \alpha E_1 \quad (\text{II.7})$$

Où α est une constante représentant la polarisabilité de la molécule.

N est le nombre de dipôles par unité de volume.

Le champ local E_1 peut être considéré comme la source de deux composants :

$$E_1 = E + E_d \quad (\text{II.8})$$

Où :

E est le champ imposé par la source.

E_d est le champ résultant de la présence de dipôles dans tout le volume du diélectrique.

Le champ local agissant sur un dipôle est donné par l'expression :

$$E_1 = E + \frac{P}{3\epsilon} \quad (\text{II.9})$$

En tenant compte de la relation II.6 et en la combinant avec II.9, on trouve :

$$E_1 = \frac{\epsilon_r + 2}{3} E \quad (\text{II.10})$$

On établit la relation de Clausius-Mossotti :

$$\frac{\epsilon_r - 1}{\epsilon_r + 2} = \frac{\chi}{\epsilon_r + 2} = \frac{N\alpha}{3\epsilon_0} \quad (\text{II.11})$$

dans la mesure où plusieurs mécanismes de polarisation différents existent, leurs effets s'additionnent :

$$\frac{\epsilon_r - 1}{\epsilon_r + 2} = \frac{1}{3\epsilon_0} \sum_i^n N_i \alpha_i \quad (\text{II.12})$$

Cette équation relie les propriétés microscopiques α_i à la grandeur macroscopique qu'est la permittivité.

II.2.2. Différents types de polarisation.

Avec les polymères couramment utilisés, les dissipations en courant alternatif proviennent principalement de la relaxation des différentes polarisations. On distingue quatre types de polarisation :

- Polarisation électronique
- Polarisation ionique
- Polarisation dipolaire ou d'orientation
- Polarisation par charges d'espace

Chacune de ces polarisations contribue à la valeur de la permittivité relative et donne une relaxation qui se traduit par un pic pour l'indice de pertes diélectriques $\epsilon_r \text{tg}\delta$.

A) Polarisation électronique.

A l'échelle atomique du diélectrique, sous l'effet du champ électrique E , le nuage électronique de chaque atome est déplacé légèrement par rapport au noyau. Il se crée ainsi des moments dipolaires induits, causes de la polarisation électronique dont le temps d'établissement est très court (de l'ordre de 10^{-15} s).

$$P_e = N\alpha_e E_1 \quad (\text{II.13})$$

Où α_e est la polarisabilité électronique, $\alpha_e = 4\pi\epsilon_0 r^3$ (r : rayon de l'atome)

B) Polarisation ionique.

Lorsque différents atomes sont présents dans une molécule (cas des polymères), les électrons participant aux liaisons de covalence se déplacent préférentiellement vers

l'atome le plus électronégatif créant ainsi, sous contrainte extérieure, un moment de liaison. Si l'on superpose un champ électrique, les atomes se déplacent les uns par rapport aux autres donnant naissance à une polarisation ionique qui met en jeu le déplacement des ions dans les réseaux cristallins. Son temps d'établissement est de l'ordre de 10^{-13} à 10^{-12} s.

$$P_i = N\alpha_i E_1 \quad (\text{II.14})$$

α_i est la polarisabilité ionique.

C) Polarisation dipolaire ou d'orientation.

En général, les molécules hétérogènes présentent un moment dipolaire dit spontané. Il est omniprésent même en l'absence d'un champ électrique. L'orientation de ces dipôles est alors aléatoire. La polarisation d'orientation a été étudiée par Debye. Elle concerne l'orientation des molécules dipolaires sous l'action du champ électrique. Elle dépend de la température et apparaît dans les liquides et les corps amorphes. Elle donne lieu à des pertes diélectriques. Dans le cas des polymères, le moment dipolaire résultant de chaque unité manométrique est obtenu par la somme vectorielle des différents moments de liaison. Si on superpose un champ électrique constant, chaque dipôle est soumis à un couple créant une polarisation résultante dans le sens du champ électrique. La valeur de la polarisation est alors donnée par la relation :

$$P_d = N\alpha_d E_1 \quad (\text{II.15})$$

Avec :

$$\alpha_d = \mu^2 / (3KT)$$

μ : moment dipolaire de la molécule.

K : constante de Boltzman.

T : température absolue.

Quand un champ électrique est appliqué, la polarisation d'orientation n'atteint pas immédiatement sa valeur maximale et de même, à la coupure du champ, la polarisation décroît au bout d'un temps t déterminé. La polarisation s'exprime généralement par une équation de la forme

$$P = P_d e^{-t/\tau} \quad (\text{II.16})$$

Avec τ : temps de relaxation.

Deux conditions sont à vérifier pour observer cette polarisation. La première exige une dissymétrie structurale, la seconde exige que la molécule ou le groupement atomique puisse s'orienter dans le champ électrique. La polarisation d'orientation confère aux matériaux une valeur relativement élevée de la permittivité relative, une tangente de l'angle de pertes diélectriques élevée et une forte dépendance de la constante diélectrique avec la température.

D) Polarisation par charge d'espace (Effet Maxwell-Wagner-Sillars).

Ce type de polarisation intervient dans le cas de matériaux hétérogènes, soit constitués de composés différents (charges, impuretés), soit constitués de conglomerats de différents éléments de volume macroscopique ayant des conductivités et des constantes diélectriques différentes. C'est le cas du polyéthylène qui contient une charge de conductivité appréciable telle que le noir de carbone. Dans ce cas, des charges d'espace peuvent alors s'accumuler aux interfaces créant des distorsions macroscopiques du champ appliqué. Ces charges d'espace proviennent du fait que les porteurs de charges qui, habituellement, peuvent migrer jusqu'aux électrodes dans un

diélectrique homogène, sont alors piégés aux interfaces. Le temps d'établissement de la polarisation interfaciale est de quelques 10^{-2} à 10^{-1} s.

$$P_c = N\alpha_c E_1 \quad (\text{II.17})$$

II.2.3. Conduction des isolants polymères.

Aux champs électriques faibles (de l'ordre de 10^6 V/m), la conduction des isolants présente généralement un caractère ohmique. Pour les solides, il est admis qu'elle est due au déplacement d'ions. En ce qui concerne les polymères, de nombreux chercheurs insistent sur le rôle joué par les ions résultant de la dissociation d'impuretés qui ont été introduites dans le matériau lors des opérations de fabrication et de mise en œuvre.

En générale, la conductivité d'un matériau peut s'exprimer selon la formule II.18 :

$$\sigma = \sum_i^n n_i q_i \mu_i \quad (\text{II.18})$$

Où : σ : conductivité
 n_i : nombre de porteurs i
 q_i : charge d'un porteur i
 μ_i : mobilité d'un porteur i

dans la plupart des cas, la conductivité σ varie avec la température suivant la formule II.19 :

$$\sigma = \sigma_0 \exp(-w/RT) \quad (\text{II.19})$$

Où : w : énergie d'activation
 R : constante des gaz parfaits
 T : température absolue

Cette relation a été vérifiée dans le cas de nombreux isolants liquides et solides. Néanmoins, la nature des charges transitant dans les isolants solides (ions ou électrons) reste controversée. Les modèles théoriques ne permettent pas de représenter les phénomènes physiques qui restent complexes compte tenu des structures des matériaux.

Aux champs électriques élevés, la conduction des isolants perd son caractère ohmique. Dans les isolants solides, les phénomènes de conduction ne sont pas encore complètement élucidés[14].

II.3 MECANISME DE RUPTURE DIELECTRIQUE DES ISOLANTS SOLIDES.

Dans les isolants solides, la rupture obéit à plusieurs processus physiques différents et dépendant des conditions expérimentales.

Les isolants solides utilisés en pratique, renferment des vacuoles de gaz quel que soit le soin apporté à leur mise en œuvre. La rupture diélectrique se produit à une valeur de champ électrique relativement faible. Le rôle de l'ionisation du gaz occlus suivie d'une érosion chimique joue ici un rôle déterminant. Les différents processus de rupture diélectrique des isolants solides peuvent être classés suivant cinq types :

- Rupture intrinsèque
- Rupture par avalanche
- Rupture thermique
- Rupture par décharge partielles

- Rupture par arborescences

II.3.1. Rupture intrinsèque.

Lorsqu'un diélectrique est soumis à un champ électrique, on constate expérimentalement que les temps de rupture sont d'autant plus courts que le champ est plus élevé. Dans ce cas la valeur du champ de claquage est une caractéristique dépendant uniquement du solide isolant.

Compte tenu de la rapidité avec laquelle se développe cette rupture diélectrique, on peut penser qu'elle est d'origine électronique.

En appliquant un champ électrique E à un isolant, il s'ensuit une densité de courant J donnant un accroissement d'énergie de la forme : $A = JE$. Cette énergie dépend du champ E, de la température T_0 et d'un paramètre α servant à décrire les électrons de conduction.

En 1947 Frohlich proposa un modèle pour quantifier la quantité d'énergie prélevée du champ électrique par les électrons libres. Cette énergie est de la forme :

$$A(E, T_0, T) = n_c \frac{e\tau_v}{m} E^2. \tag{II.20}$$

avec $n_c = n_s \text{Exp}\left(-\frac{\Delta W}{KT}\right)$

où

- n_c : nombre d'électrons dans la bande de conduction.
- n_s : nombre de pièges au niveau ΔV .
- ΔW : énergie du niveau des pièges.
- τ_v : temps de relaxation des électrons.
- K : constante de Boltzmann.
- e : charge de l'électron.

II.3.2. Rupture diélectrique par collisions ionisantes.

Dans ce mécanisme on suppose que les électrons sont injectés par la cathode et traversent l'isolant en créant des paires électrons-trous par collision avec les atomes. Les trous pièges forment une charge d'espace positive qui tend à extraire d'avantage d'électrons de la cathode. Dans le cas où le champ électrique dépasse une certaine valeur (champ critique E_c), on a une forte augmentation du courant et une rupture diélectrique s'ensuit. Le processus est gouverné par les quatre équations suivantes :

$$J = AE^2 \text{Exp}(-B/E) \quad \text{le courant d'électron émis par la cathode} \tag{II.21}$$

$$\frac{\partial E}{\partial X} = \frac{qp}{\epsilon_0 \epsilon_r} \quad \text{Equation de poisson} \tag{II.22}$$

$$\frac{\partial P}{\partial T} = \left(\frac{J}{q}\right)\alpha - \left(\frac{J}{q}\right)\delta_p \quad \text{Equation de création des trous par ionisation} \tag{II.23}$$

δ : section de rupture.

$$\alpha = \alpha_0 \text{Exp}\left(\frac{-H}{E}\right) \quad \text{Expression du coefficient d'ionisation} \tag{II.24}$$

où $H = \frac{E_i}{qL}$

E_i : potentiel d'ionisation

L : libre parcours moyen des électrons.

q : charge de l'électron.

A : quantité d'énergie prélevée du champ électrique par les électrons libres.

P : polarisation diélectrique.

B : constante.

Ces équations conduisent au champ de rupture :

$$E_c = E_m \left(1 + \frac{E_m}{H} \right) \quad (II.25)$$

Où E_m est donné par la relation :

$$\frac{\text{Exp} \left(\frac{H}{E_m} \right)}{\left(\frac{H}{E_m} \right)^2} = \frac{q \alpha_0 e}{H \epsilon_0 \epsilon_r} \quad (II.25)$$

avec e : épaisseur de l'échantillon.

II.3.3. La rupture thermique.

En présence d'un champ électrique, l'existence de porteurs de charges dans un matériau isolant implique celle d'un courant de conduction qui peut conduire à un échauffement de cet isolant par effet Joule. Cependant ce courant n'est pas la seule source de chaleur. Nous savons que les diélectriques présentent des phénomènes de relaxation, quand ils sont soumis à un champ électrique alternatif. Ce qui engendre des mécanismes de pertes qui contribuent à leur tour à cet échauffement.

Le claquage thermique a lieu alors, quand une accumulation de chaleur se produit et que la température du matériau atteint une valeur critique T_c , au delà de laquelle le courant croit rapidement entraînant sa rupture totale.

Si un champ électrique statique est appliqué à un isolant, la chaleur produite par effet Joule est : σE^2 ; en courant alternatif il faut tenir compte des pertes diélectriques.

$$E^2 \omega \epsilon_0 \epsilon_r'' = E^2 \omega \epsilon_0 \epsilon_r' \text{tg} \quad (II.26)$$

La chaleur stockée dans l'isolant vaut : $C_v dT/dt$ et celle dissipée vaut : $\text{div}(\lambda \Delta T)$.

D'où la relation : $\sigma E^2 = C_v dT/dt - \text{div}(\lambda \Delta T)$

Avec :

C_v : chaleur spécifique.

λ : conductivité thermique.

T : température.

σ : conductivité électrique.

L'équation de continuité est également nécessaire puisqu'il ne doit pas se produire d'accumulation de charge : $\text{div}(\sigma E) = 0$.

Considérant un échantillon d'épaisseur d et de grande surface, les électrodes ont une résistance thermique faible et le flux de chaleur progresse dans la direction z qui leur est perpendiculaire. En régime permanent on a : $dT/dt = 0$.

L'équation de la chaleur devient :

$$\frac{\partial}{\partial z} \left(\lambda \frac{\partial T}{\partial z} \right) + \sigma \left(\frac{\partial V}{\partial z} \right) = 0 \quad (II.26)$$

$$\sigma \left(\frac{\partial V}{\partial z} \right) = -J$$

$$V = E/d$$

Pour le calcul de la tension critique V_{0c} pour une température critique T_{mc} , on prend en considération les conditions suivantes :

- au centre du diélectrique : $z=0, V=0, T = T_m$.
- à la surface $z = \frac{d}{2}$: $V = \frac{V_0}{2}$ et la tension V_{0c} est donnée par :

$$V^2_{0c} = 8 \int_{T_0}^{T_m} \left(\frac{\lambda}{\sigma} \right) dT$$

la conductivité σ est donnée par :

$$\sigma = \sigma_0 \text{Exp}(-W/KT) \quad \text{où}$$

W : est l'énergie d'activation.

σ_0 : une constante.

K : constante de Boltzmann.

$$V^2_{0c} = 8 \int_{T_0}^{T_{mc}} \frac{\lambda}{\sigma_0} \text{Exp} \left(\frac{W}{KT} \right) dT. \quad (\text{II.27})$$

comme $W \gg KT$ et $T_{mc} > T_0$, alors V_{0c} vaut :

$$V_{0c} \approx \left(\frac{8K\lambda T_0^2}{\sigma_0 W} \right)^{1/2} \text{Exp} \left(\frac{W}{2KT_0} \right)$$

V_{0c} représente la tension « thermique maximum » qui est la tension de rupture diélectrique maximum.

II.3.4. Rupture diélectrique provoquée par les décharges partielles.

La rupture diélectrique causée par les décharges partielles qui se forment au sein de l'isolant au cours de sa mise en œuvre est caractérisée par les éléments suivants :

- le processus de dégradation est très lent, la durée de vie dépend de l'amplitude du champ électrique appliqué et de sa fréquence.
- la mise sous tension de l'isolation conduit à des perturbations radiophoniques importantes.

L'existence de vacuoles à l'intérieur d'un isolant a pour conséquence une augmentation non négligeable du champ électrique qui y règne. Lorsqu'on atteint la tension seuil d'ionisation du gaz enserré dans la vacuole, des décharges par alternance apparaissent et ont pour conséquences les effets suivants :

- une augmentation locale de la température.
- une érosion mécanique des parois de la vacuole par le bombardement d'ions et d'électrons.
- des réactions chimiques provoquées par des atomes excités par les décharges.
- Des pertes diélectriques appelées pertes par ionisation.

II.3.5. Rupture diélectrique provoquée par les arborescences.

Le développement de l'arborescence est favorisé par la présence de défaut : hétérogénéité à l'interface des parties isolantes et conductrices, impuretés ou occlusions gazeuses au sein de l'isolant. Ce n'est qu'après un temps de gestion appelé aussi temps

d'incubation plus au moins long qu'elle se manifeste. Le temps d'incubation dépend de la nature du défaut d'où elle est issue.

L'arborescence n'apparaît qu'aux points où des conditions critiques sont atteintes. Révélatrice d'un stade très avancé de la dégradation du matériau, elle est considérée comme un phénomène précurseur du préclaquage.

Il est d'usage de classer les arborescences suivant deux catégories :

- les arborescences électriques.
- les arborescences électrochimiques, dans ce cas, on suppose que leur formation est due à un phénomène de migration d'ions.

on distingue trois phases dans l'évolution du phénomène d'arborescence :

- la phase d'initiation qui est la période qui sépare le début de l'application de la tension du moment d'apparition des premières arborescences.
- la phase de propagation qui est toujours accompagnée de décharges partielles.
- la phase de claquage qui se produit rapidement lorsque les branches de l'arbre traversent complètement le mur isolant[17].

II.4. POLYMERES ET POLYMERISATION.

Un polymère est une macromolécule obtenue par la répétition d'une unité constitutive, comportant un groupe d'atomes, appelée monomère. Les polymères sont synthétisés en reliant des molécules des monomères entre elles par des liaisons chimiques. Cette réactivité chimique des monomères résulte de la présence des groupes fonctionnels capables de former des liaisons chimiques avec les groupes fonctionnels d'autres molécules de monomère.

Pour démarrer une réaction de polymérisation on a besoin généralement d'un amorceur de polymérisation qui est une substance qui possède ou moins une liaison chimique susceptible de générer par rupture homolytique des radicaux libres. Il existe deux types d'amorceurs : les peroxydes et les dérivés azoïques.

II.4.1. Classification des polymères.

Cette classification peut se faire en se basant sur l'origine du matériau, sa structure chimique, la géométrie de la molécule et le type de polymérisation utilisé.

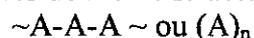
II.4.1.1. Origine du matériau.

- polymères naturels tels que : le bois, le coton, le cuir, la laine, ..., etc
- polymères synthétiques : appelés aussi résines synthétiques, du fait des similarités observées entre leur structure et celle de la résine naturelle.
- polymères artificiels : qui résultent de la transformation chimique des polymères naturels.

II.4.1.2. Structure chimique et fonctionnalité des motifs.

Suivant que le motif principal du polymère comprenne un ou plusieurs types de monomères, on distingue : les homopolymères et les copolymères.

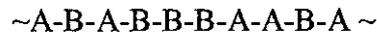
a/ les homopolymères sont des polymères constitués par des liaisons covalentes entre un nombre de monomères de même structure (même monomère).



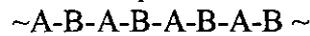
avec A : le monomère, n : le nombre de monomères.

b/ les copolymères : les monomères sont de structure différentes suivant leur architecture moléculaire; en distingue plusieurs types :

- copolymère statistique : les monomères A et B sont réparties aléatoirement dans la chaîne.



- copolymère alterné : A et B réparties alternativement



- copolymère séquencé :



II.4.1.3. Géométrie de la macromolécule :

La géométrie de la macromolécule peut être caractérisée par l'ordre atomique et la distribution des liaisons.

a) Ordre atomique.

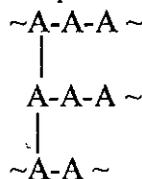
On distingue selon cette caractéristique plusieurs types :

- polymère amorphe : où les atomes sont désordonnés le long de la chaîne du polymère.
- polymère cristallin : suite à certaines conditions (refroidissement lent après fusion, ...) les polymères peuvent posséder une structure semi-ordonnée, et donner lieu à une cristallisation partielle.

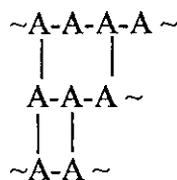
b) Distribution spatiale des liaisons.

Dans ce cas, on parle des polymères linéaires, ramifiés ou tridimensionnels.

- polymères linéaires : ils peuvent être des homopolymères ou des copolymères.
- polymères ramifiés : leur structure est toujours planaire mais leurs macromolécules sont constituées de branches liées chimiquement.



- polymères tridimensionnels : pour des raisons de stabilité des liaisons relie les divers branches.



Aux structures linéaires et faiblement ramifiées correspondent les thermoplastiques, et aux structures faiblement réticulées (réseau à mailles larges) correspondent les élastomères, et quand les mailles du réseau deviennent serrées le matériau est appelé thermodurcissable.

II.4.1.4. Type de polymérisation utilisé.

Il existe deux types de polymérisation, réaction de polycondensation et réaction de polyaddition.

a) Réaction de polycondensation.

La réaction se produit par simple chauffage en présence d'un simple catalyseur (amorceur) convenable : on passe des molécules monomères puis aux molécules polymères formées, et enfin aux macromolécules

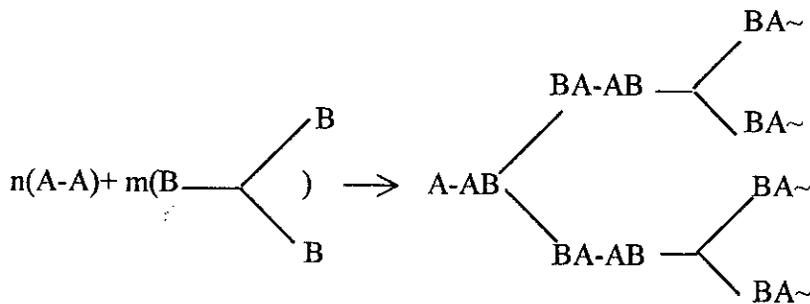
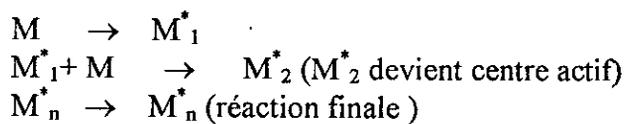


Fig II.2 : polymérisation par réaction de polycondensation.

b) Réaction de polyaddition.

Elle est appelée aussi polymérisation proprement dite : une molécule monomère M s'ajoute à un centre actif de la chaîne macromoléculaire en cours de formation.



II.4.1.5. Techniques de polymérisation.

a) Polymérisation en masse.

L'amorçage de la polymérisation s'effectue dans le monomère liquide à l'aide d'amorceur radicalaire conventionnel. Cette méthode permet d'obtenir des masses moléculaires élevées indispensables aux bonnes propriétés mécaniques.

b) Polymérisation en suspension.

Dans ce cas, le polymère est insoluble dans un milieu dispersant tel que l'eau, l'amorceur n'est soluble que dans le polymère. Chaque gouttelette de la suspension peut être considérée comme un système isolé de polymérisation en masse.

A la fin de la réaction, le polymère se trouve sous forme de perles si le polymère et le monomère sont solubles l'un dans l'autre, ou sous forme de granulés, s'ils ne le sont pas.

c) Polymérisation en solution.

Elle est effectuée cette fois en présence d'un solvant, dont le rôle essentiel est de permettre un contrôle de dégagement thermique lors de la réaction.

d) Polymérisation en émulsion.

Le monomère presque insoluble dans le milieu dispersant est maintenu sous forme d'émulsion, grâce à un agent tensioactif et à une agitation énergétique. Cette fois le système catalytique est soluble dans l'eau et non plus dans le polymère[16].

II.5. CARACTERISTIQUES DES POLYMERES.**II.5.1. Comportement chimique des polymères.**

Au contact des liquides ou des gaz, le comportement des polymères varie suivant leur nature et leur composition; il peut y avoir diffusion, absorption avec gonflement ou gélification, dissolution ou distinction selon l'agressivité des produits en présence. Ceci peut donc causer d'importantes contraintes qui peuvent être catastrophiques pour le matériau.

II.5.2. Propriétés mécaniques des polymères.

Vu leurs excellentes propriétés mécaniques, leur aptitude aux déformations et leur grande sensibilité aux variations de la température, les polymères se distinguent généralement par :

- Une masse volumique relativement faible.
- Un pouvoir d'isolement calorifique élevé.
- Une résistance élevée aux agressions atmosphériques.
- Une importante résistance aux efforts de compression et de traction.
- Une bonne résistance à l'usure.

II.5.3. Propriétés électriques des polymères.

Ces propriétés sont : la rigidité diélectrique, la résistance à la décharge superficielle et la résistance transversale.

D'une façon générale les propriétés électriques sont liées aux structures moléculaires du polymère, mais il faut quand même tenir compte de :

- la présence de charge.
- la présence d'adjuvants.
- l'absorption éventuelle d'humidité.
- la morphologie du polymère[17].

II.6. PRESENTATION DU MATERIAU .

Dans notre projet, nous nous sommes basés sur les résultats des travaux de recherche concernant le PVC. A cet effet nous ne présentons que ce matériau.

II.6.1. Présentation du PVC.

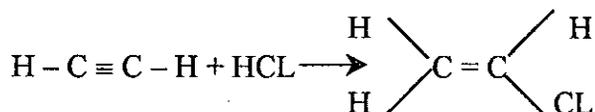
Le PVC est très utilisé dans l'isolation des câbles de haute tension à cause de son faible prix de revient. A l'état pur, le PVC est cassant d'où la nécessité de lui ajouter des produits auxiliaires appelés adjuvants. Le type et la quantité de chaque adjuvant est le grand problème des producteurs de câbles électriques.

II.6.2. formulation et mise on œuvre du PVC.

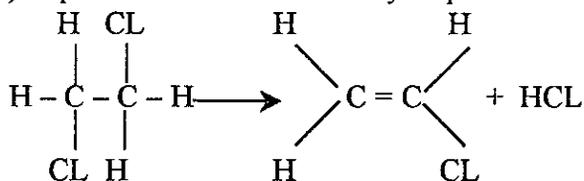
II.6.2.1. Préparation du chlorure de vinyle.

Le chlorure de vinyle monomère (CVM) $CH_2 = CH CL$ est préparé suivant deux procédés principaux.

a) Addition du gaz chlorhydrique à l'acétylène :



b) Séparation de l'acide chlorhydrique du dichloréthane :

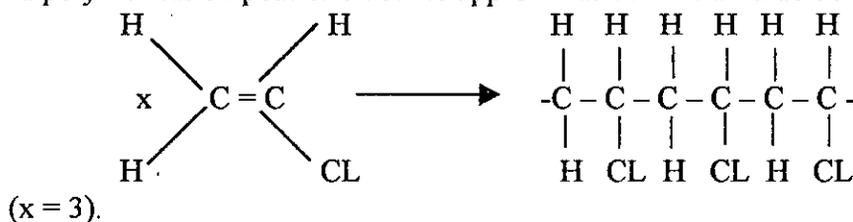


II.6.2.2. Molécule de PVC.

La polymérisation du chlorure de vinyle conduit à la formation de la molécule de PVC. Cette polymérisation peut être conduite de différentes façons et permet d'obtenir des résines particulières. Selon la firme PLASTIMER, la polymérisation dépend :

- du procédé d'obtention : polymérisation en suspension, en masse ou en émulsion.
- De la longueur de la chaîne c'est à dire le nombre n de motifs monomères qui figurent dans cette dernière.
- De l'insertion dans la chaîne de motifs monomères autres que le chlorure de vinyle pour modifier certaines caractéristiques.

La polymérisation peut être décrite approximativement à l'aide de la formule suivante :



II.6.2.3. Mélange de PVC prêt à l'emploi.

La difficulté majeure dans la fabrication du PVC destiné à l'isolation diélectrique réside dans la stabilisation des composés du matériau qui renferme en général de la résine de PVC et des additifs tels que les charges, les stabilisants, les lubrifiants, les plastifiants, ... etc.

II.6.2.4. Les adjuvants.**a) les stabilisants.**

Sous l'action de la chaleur, le PVC peut libérer de l'acide chlorhydrique. L'effet des stabilisants est d'éviter ce dégagement d'acide et à le capter s'il se produit. Pour cela ils empêchent l'oxydation et la rupture des liaisons doubles.

b) Les plastifiants.

Dans le but de rendre le matériau moins cassant, il est généralement fait appel à des additifs appelés plastifiants. Le moment dipolaire élevé des liaisons polaires C – Cl de la molécule de PVC engendre de puissantes chaînes moléculaires. Ces liaisons élèvent la température de transition vitreuse à environ 78°C. Au dessous de cette température, le PVC est rigide et ne résiste que peu aux chocs, étant donné que la mobilité des segments de chaînes est diminuée. Cet état de rigidité peut être supprimé par des substances qui affaiblissent les forces de liaisons intermoléculaires : les plastifiants.

Le plastifiant intervient entre les macromolécules de PVC à haute température parce que les liaisons intermoléculaires sont moins fortes, ce qui facilite la pénétration du plastifiant, et permet par suite aux macromolécules des mouvements de chaînes les unes par rapports aux autres. Les plastifiants ont pour effet de diminuer la résistance à la traction et d'augmenter l'allongement à la rupture et cela selon le pourcentage utilisé.

c) Les charges.

L'emploi des charges n'a généralement pour but qu'un abaissement du prix de revient ou une amélioration des propriétés diélectriques mais non des propriétés mécaniques [13].

Les principales charges utilisées sont : le carbonate de calcium, le kaolin calciné, la silice et l'amiante.

d) Les lubrifiants.

L'emploi du lubrifiant a pour but de faciliter l'écoulement du PVC lors de la mise en œuvre. Ils ont pour effet d'atténuer la tendance qu'a la résine de base à adhérer aux surfaces environnantes. Ceci est réalisé à une température de collage supérieure à la température de travail.

Les lubrifiants peuvent être classés en deux groupes d'après la firme ULTRYL PVC.

- lubrifiants externes : leur rôle est de réduire les frictions entre le polymère en fusion et l'outillage c'est à dire d'empêcher le PVC fondu d'adhérer à la vis du cylindre ou ailleurs.
- Lubrifiants internes : leur rôle est de réduire les frictions entre les molécules du polymère et de faciliter ainsi l'écoulement de la matière en limitant l'auto-échauffement.

e) Les pigments et colorants.

Ils permettent de réaliser la coloration des mélanges. Parmi les colorants et pigments, il faut choisir ceux qui sont stables aux températures de travail et éventuellement en présence de traces d'acide chlorhydrique. Il faut éviter, sauf dans des cas très particuliers, l'emploi de colorants solubles dans les plastifiants qui donneraient lieu à des migrations ou déteintes en surface.

II.6.2.5. Propriétés du PVC.

a. Propriétés physico-chimiques.

La masse volumique des compositions du PVC varie entre 1.0 et 2.0 g/cm³ à la température ambiante (1.4 g/cm³ pour les mélanges non plastifiés et 1.2 à 1.3 g/cm³ pour les formules plastifiées).

- Le PVC transmet mal la flamme mais sa combustion entraîne le dégagement de gaz corrosifs et de fumées. Ce dégagement peut être réduit par l'emploi d'un mélange de formulation spéciale.

- Le PVC a un bon vieillissement thermique. Sous l'action prolongée de la chaleur, il devient jaune puis brun, puis rouge et enfin noire et forme en général une intumescence

- Le PVC a une résistance exceptionnelle à l'ozone. il possède une résistance moyenne aux hydrocarbures aliphatiques, aux acides et bases .

- Le PVC résiste au fluage à température élevée (absence de point de fusion franc) mais il est fragile aux basses températures (demande de précautions pour la pose à temps très froid).

- Le PVC rigide résiste bien aux agents chimiques inorganiques. Il est soluble dans les cétones, le cyclohexanone et le dichloroéthylène .

- La température de transition vitreuse du PVC est élevée, elle se situe entre 75 et 80°C, ce qui le rend impropre à de nombreuses applications (câbles par exemple) sans l'adjonction de plastifiants. Avec l'emploi de plastifiants, la température de transition vitreuse se situe entre -10°C et -5°C .

b. Propriétés électriques.

La rigidité diélectrique est de 37 kV/mm pour une éprouvette de 3 mm d'épaisseur à 20°C.

- Le facteur de pertes diélectriques est de l'ordre de : $\text{tg } \delta$ (à 50 Hz) = 0.015 à 20°C .

- La constante diélectrique est de l'ordre de : ϵ_r (à 50 Hz) = 3.4 à 20°C .

- La résistance transversale est de l'ordre de : $\rho_t \approx 10^{16} \Omega \cdot \text{cm}$ à 20°C pour les produits non plastifiés, $\rho_t \approx (10^{10} \text{ à } 10^{15}) \Omega \cdot \text{cm}$ à 20°C pour les produits plastifiés.

c. Propriétés mécaniques.

- Le PVC a une bonne résistance à la traction de l'ordre de : 5.5 à 6 kgf/mm² à 20°C
- L'allongement à la rupture est de l'ordre de : 100% pour les produits non plastifiés, et 200 à 400% pour les produits plastifiés.
- Au dessus de 140 à 150°C le PVC perd toutes ses propriétés mécaniques[15].

II.6.3. Vieillissement thermique du PVC.

On convient d'appeler vieillissement thermique tout phénomènes traduisant une évolution qui ne fait intervenir que :

- le matériau,
- l'atmosphère,
- la température,

le vieillissement thermique peut faire intervenir des phénomènes purement physiques (dégazage, migration de plastifiants, évolution de la morphologie) ou des phénomènes chimiques parmi lesquels le plus important est l'oxydation. Ces phénomènes peuvent affecter la structure chimique du polymère, ou, dans certains cas, concerner seulement les additifs (problèmes d'incompatibilité d'adjuvants). Le vieillissement est d'autant plus rapide que la température est élevée.

Le vieillissement thermique peut être continu ou cyclique. Dans le premier cas, le polymère est exposé continuellement à la chaleur, le second cas consiste à faire subir au matériau des cycles de chauffage et de refroidissement.

II.6.3.1. Processus de dégradation des polymères.

a) Oxydation.

En présence d'air, il se produit un processus d'oxydation. Cette oxydation a pour conséquence la formation de produits acides qui augmentent la conductivité ainsi que la rupture de chaînes qui est responsable de l'augmentation de la charge à la rupture.

b) Réticulation.

Le processus d'oxydation s'accompagne généralement d'un effet de réticulation ou de pontage entre des chaînes voisines qui réduit la souplesse de l'isolant, le rendant ainsi cassant.

c) Réaction de dépolymérisation.

Des réactions de dépolymérisation peuvent prendre naissance même en l'absence d'oxygène. L'évaporation de produits de scission de plus bas poids moléculaire a une double conséquence : d'une part un rétrécissement du matériau et d'autre part la formation de produits volatils susceptibles d'être à l'origine de décharges partielles dans les isolations à haute tension.

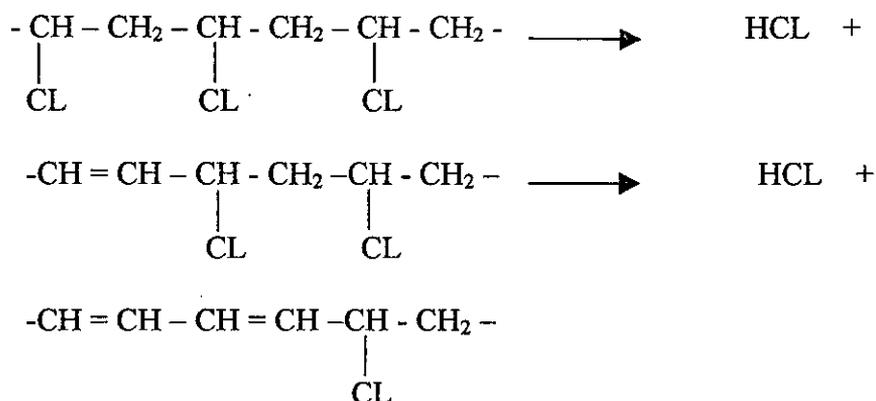
II.6.3.2. dégradation thermique du PVC.

a) Dégagement d'acide chlorhydrique.

Les propriétés toxiques et corrosives de l'acide chlorhydrique sont bien connues. Or il apparaît d'après les travaux de HAVEY (1959), STEPANEK et DOLEZEL (1963)[], TALAMINI et PEZZIN (1960) que l'action de la chaleur sur le PVC peut mener à la production de cet acide. D'autres auteurs ont proposé un schéma de formation de ce produit.

b) Déroulement de la réaction.

Plusieurs auteurs ont donné une explication au déroulement de la réaction. La firme BAYER a montré que les groupements C - CL instables dus à des irrégularités de structures, sont responsables de la libération d'acide chlorhydrique. Ce dégagement commence au dessus de 100°C et se poursuit à une vitesse accrue aux températures de mise en œuvre de l'ordre de 170°C. Il s'agit notamment de groupes insaturés qui résultent d'un transfert ou d'une rupture de chaînes pendant la polymérisation. Ces groupes se manifestent surtout à des température avoisinant 150°C. A des températures plus élevées de l'ordre de 190°C, des ramifications tertiaires peuvent également intervenir. Dès que la réaction a démarré au centre en question, elle progresse d'unité de chaîne en unité de chaîne compte tenu du fait que chaque liaison insaturée disloque l'atome de chlore en position alpha par activation « allyle » et induit une nouvelle libération d'acide chlorhydrique. Le schéma de la réaction est le suivant :



Ce processus qui se déroule le long des macromolécules peut donner lieu à la formation de chaînes polyéniques conjuguées.

c) conséquence de la dégradation.

- Rétrécissement de la molécule de PVC :

La déshydrochloruration et l'auto-oxydation modifient la taille de la molécule de PVC par l'intervention simultanée d'une rupture de chaînes de réaction de réticulation.

- Changement de couleur :

La dégradation thermique du PVC se traduit par une modification de la teinte de la résine .

- Perte de plastifiant :

le vieillissement thermique du PVC peut conduire à une perte de plastifiant qui est la cause dominante de la dégradation du polymère[15].

II.6.4. Endurance thermique du PVC.

II.6.4.1. Position du problème.

La caractéristique de stabilité à long terme des isolants solides doit permettre de répondre à l'une des questions simples suivantes :

- Pendant quel temps t , peut-on maintenir un matériau isolant sous une contrainte C_0 ?
- Si l'on désire une durée de maintien ou durée de vie donnée t_0 , quelle contrainte maximale sera autorisée ?

pour y répondre, il suffit de connaître la relation contrainte – temps $C = f(t)$ (fig. II.3).

C : peut être la température, la tension ou le champ électrique.

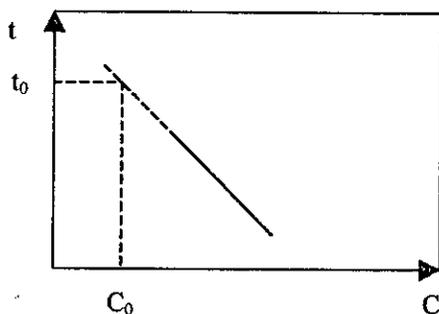


Fig.II.3 : variation du temps de vieillissement en fonction de la contrainte appliquée

Généralement on détermine les variations d'une propriété du matériau en fonction du temps de vieillissement pour différentes valeurs C_1, C_2, \dots, C_n de la contrainte appliquée C (fig. II.4).

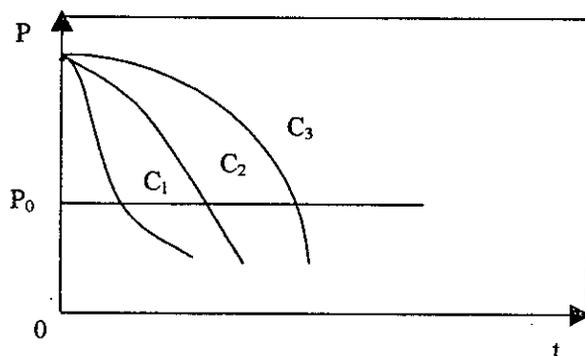


Fig.II.4 : variation de la propriété en fonction du temps de vieillissement pour différentes valeurs de la contrainte.

Si on fixe une valeur de la propriété P_0 choisie en fonction des critères d'emploi, on obtient la relation $C(t)$ relative à la valeur de la propriété P_0 . Cette courbe est appelée courbe de durée de vie.

P : peut être une propriété mécanique (charge à la rupture par exemple) ou propriété électrique (rigidité diélectrique par exemple).

II.6.4.2. Principe de la détermination de l'endurance thermique des isolants solides.

Les travaux relatifs à l'endurance thermique des matériaux isolants solides ont pour objet la mise au point des méthodes générales de détermination de leur stabilité. Le but de ces essais est de simplifier les résultats et de les exploiter sous forme d'indices de températures attribués aux matériaux pour leurs propriétés électriques d'autres part.

En ce qui concerne les essais, on soumet les matériaux à l'action prolongée de la chaleur puis on détermine l'évolution d'une ou plusieurs de leurs propriétés en fonction du temps de vieillissement, ce qui permet de tracer les caractéristiques $P = f(t)$ (fig. II.5)

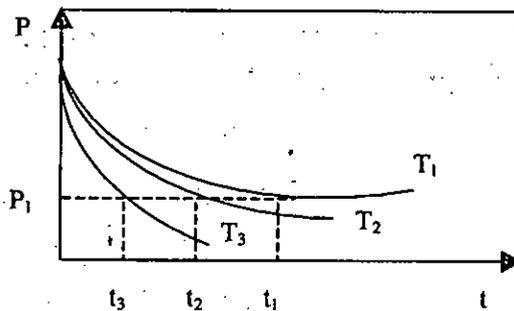


Fig.II.5 : variation de la propriété P en fonction du temps de vieillissement pour différentes températures.

Le choix d'une valeur limite de la propriété P_1 permet de déterminer les temps pour lesquels cette limite est atteinte ou durée de vie et ceci pour plusieurs températures de vieillissement. En répétant les essais, on peut tracer la courbe d'endurance thermique ou durée de vie : $t = f(T)$ (fig. II.6).

Généralement on s'intéresse à l'indice de température qui est, par définition, le nombre égal à la température exprimée en degrés Celcius qui correspond, sur le graphique d'endurance thermique, à un temps donné normalement égal à 20.000 h.

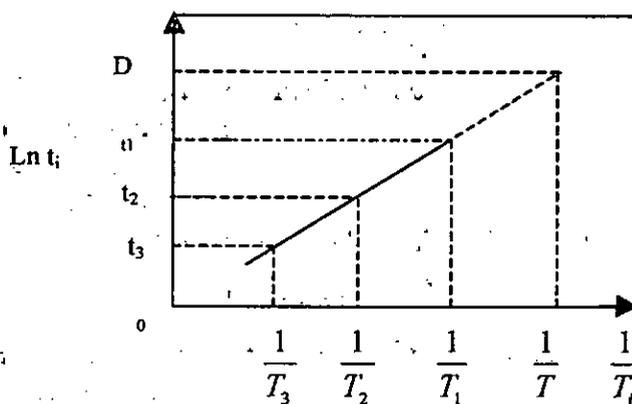


Fig.II.6 : variation du logarithme de la durée de vie des polymères en fonction de l'inverse de la température.

II.6.4.3. choix des températures de vieillissement.

La C.E.I. 216-1 fixe un minimum de trois températures différant chacune de la plus voisine de 20°C et telles que le critère de dégradation soit atteint.

propriétés mécaniques: allongement à la rupture et résistance à la traction. Nous présentons également l'indice de température, correspondant aux deux propriétés. Les conséquences de la dégradation du matériau ont été mises en valeur: perte de masse, contraction et changement de couleur.

Les essais de vieillissement thermique ont été effectués par M. NEDJAR [] sur des éprouvettes, soumises à un champ électrique sous haute tension, à différentes températures: 80°C, 100°C, 120°C and 140°C.

II.7.1. variation de l'allongement à la rupture en fonction du temps de vieillissement

Pour les températures de 80°C et 100°C, le PVC présente une augmentation, puis une diminution de l'allongement à la rupture, en fonction du temps de vieillissement. Pour les autres températures, nous n'avons constaté qu'une diminution de ce facteur (figure.II.7)

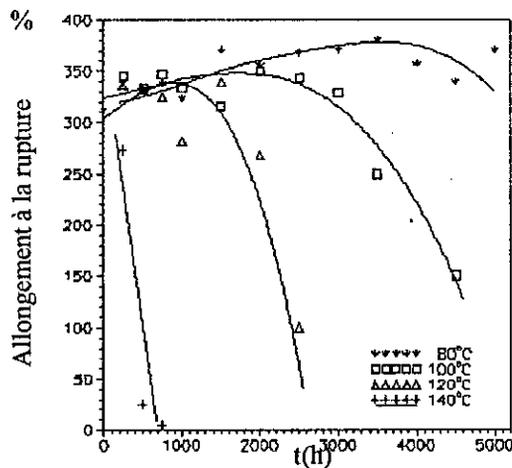


Figure.II.7: variation de l'allongement à la rupture en fonction du temps de vieillissement.

l'augmentation de l'allongement à la rupture du polymère pourrait être interprétée par un effet de réticulation, et sa diminution par une rupture de chaînes carbonées et une perte de plastifiant.

II.7.2. résistance à la traction

Pour ces caractéristiques, on constate un effet inverse par rapport à l'allongement à la rupture: lorsque ce dernier décroît, la résistance à la traction croît (figure.II.8). Pour les températures de 120°C et 140°C, la résistance à la traction croît rapidement en fonction du temps de vieillissement (figure.II.8).

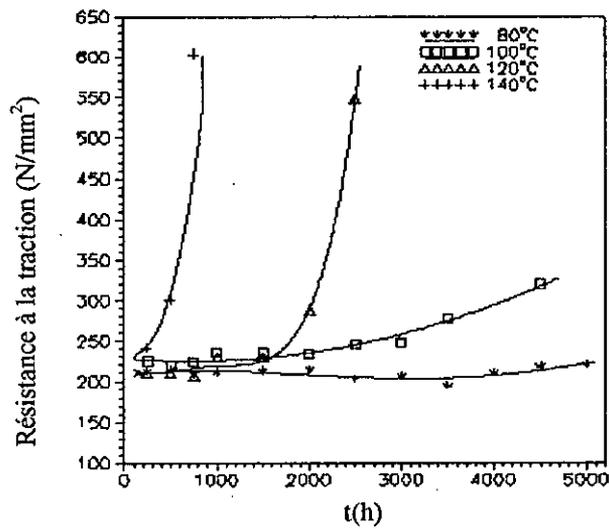


Figure.II.8. variation de la résistance à la traction en fonction du temps de vieillissement.

II.7.3. Endurance thermique

En extrapolant les caractéristiques de l'allongement à la rupture, les durées de vie du polymère relativement aux différentes températures de vieillissement peuvent être déterminées en supposant un critère de fin de vie de 50% de la valeur initiale de la propriété considérée (figure.II.9 et figure.II.10).

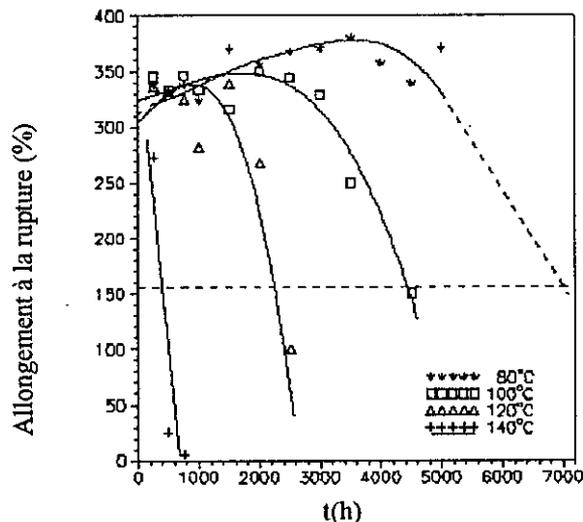


Figure.II.9. détermination de la durée de vie à partir des caractéristiques de l'allongement à la rupture.

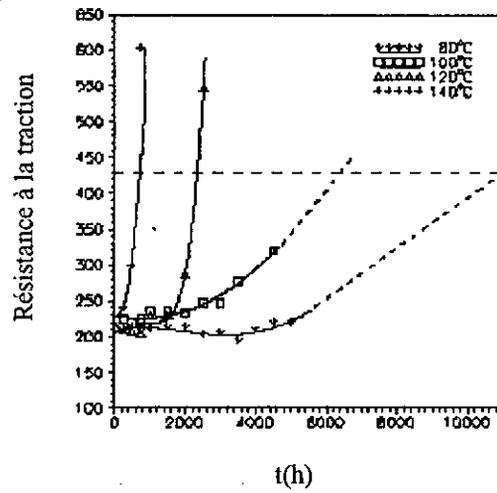


Figure.II.10. détermination de la durée de vie à partir des caractéristiques de la

La figure.II.11 montre le graphe de l'endurance thermique correspondant simultanément aux deux propriétés :

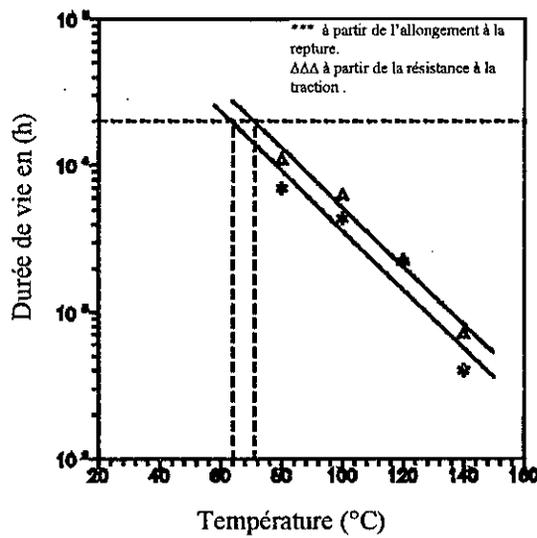


Figure.II.11. Graphe d'endurance thermique.

II.7.4. Perte de masse.

D'après M. NEDJAR, lorsque le temps de vieillissement augmente, la perte de masse augmente. Elle peut atteindre 30% après 4000h de vieillissement (figure.II.12).

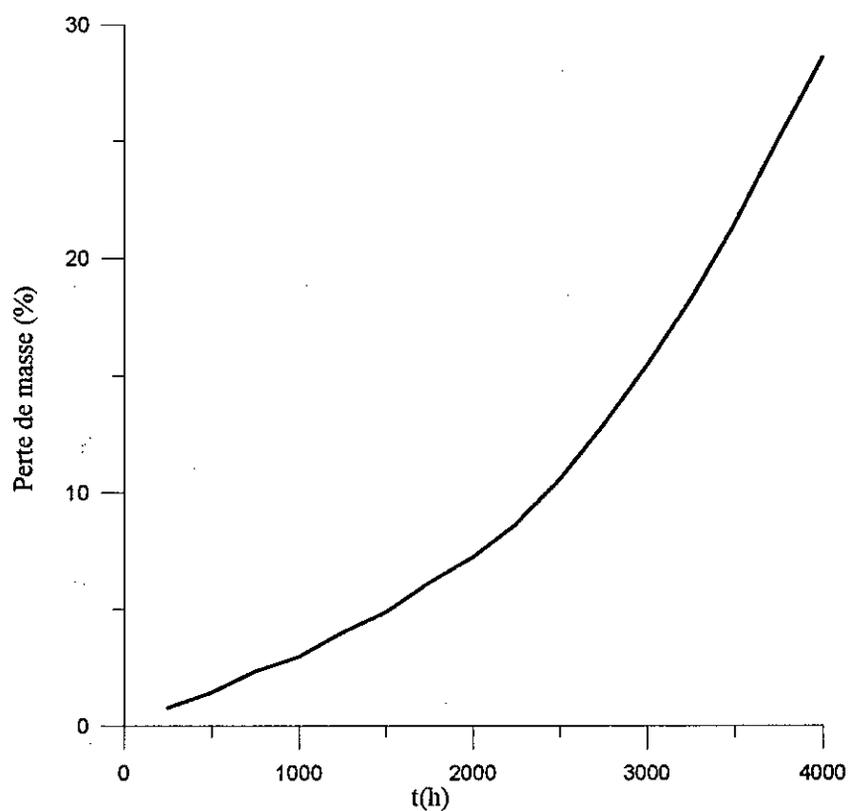
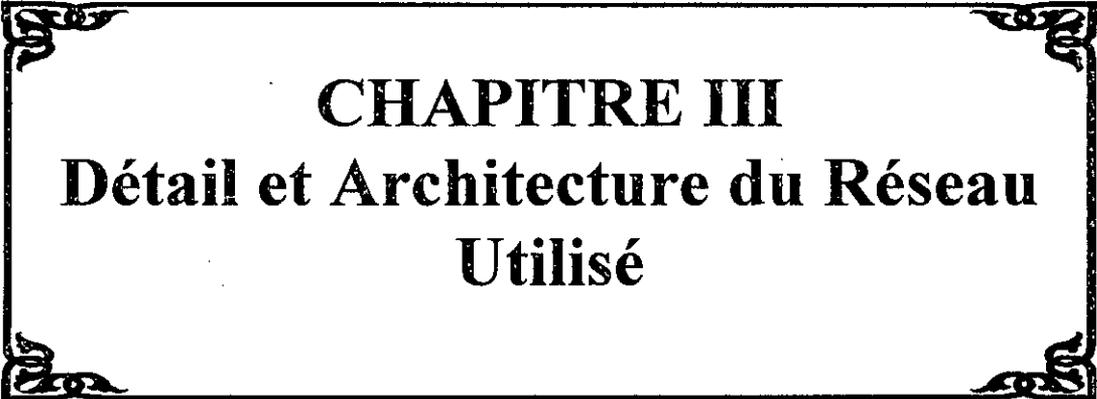


Figure.II.12. variation de la perte de masse surfacique en fonction du temps de vieillissement.



CHAPITRE III
Détail et Architecture du Réseau
Utilisé

III.1. Nature du problème étudié.

Nous allons nous intéresser à l'étude des variations de quelques propriétés diélectriques et mécaniques du PVC soumis au vieillissement thermique à différentes températures.

Notons que les allures traitées représentant les variations des propriétés en fonction du temps de vieillissement sont toutes monotones. Donc notre travail se résume comme suit.

A partir des essais déjà faits représentant la variation d'une propriété du PVC sur une période de quelques milliers d'heures sous une contrainte donnée, nous allons nous servir du début de la variation qui s'étale sur un intervalle de temps de quelques centaines d'heures seulement pour effectuer l'apprentissage d'un réseau de neurones artificiels tout en considérant le reste de l'allure comme complètement inconnu pour nous. Après l'apprentissage, nous utilisons ce réseau dont tout les paramètres sont adaptés pour prédire le reste de la courbe point par point. Nous présentons cette opération en détail dans les paragraphes suivants. En dernier lieu, et afin de vérifier la qualité des résultats obtenus, nous les comparons avec les points constituant la partie restante de la courbe réelle.

III.2. Procédure d'apprentissage suivie dans un problème de prédiction.

En général, quand il s'agit de faire une approximation d'une fonction donnée à des fins d'interpolation et lors de la phase d'apprentissage, on doit entraîner le réseau avec des exemples de type (X_p, Y_p) . C'est à dire, à chaque fois on présente à l'entrée du réseau la $p^{\text{ième}}$ abscisse X_p et on calcule l'erreur sur l'image correspondante à la sortie du réseau $Y_p - Y_p^*$ telles que Y_p est la sortie désirée et Y_p^* représente la sortie du réseau. Mais quand il s'agit d'une prédiction ou d'une extrapolation, il est nécessaire de faire l'apprentissage seulement avec les valeurs de la fonction en éliminant les abscisses. Pour justifier cela, nous considérons que l'axe des abscisses est régulièrement espacé - l'axe des heures dans notre application - De cette façon, nous pouvons savoir a priori que les valeurs Y_1, Y_2, \dots, Y_n , sont les images des abscisses $X_1, X_1 + \Delta X, \dots, X_1 + n\Delta X$. Donc on effectue l'apprentissage avec des exemples de type (Y_p, Y_{p+1}) . Par exemple on commence par présenter à l'entrée du réseau Y_1 puis on calcule à la sortie l'erreur $Y_2 - Y_2^*$, en suite on présente Y_2 et on calcule l'erreur $Y_3 - Y_3^*$, et on continue cette opération jusqu'à la $n^{\text{ième}}$ valeur. A la fin de l'apprentissage, les paramètres sont tous adaptés de telle manière que lorsque le réseau reçoit une valeur actuelle, et grâce aux quelques améliorations convenables introduites au niveau des poids, il fournira une valeur future. Donc si on entraîne le réseau avec les valeurs Y_1, Y_2, \dots, Y_{n-1} , après l'apprentissage, lorsqu'on lui présente la valeur Y_n , ce dernier nous donne Y_{n+1} que l'on n'avait pas au début. On peut dire que le réseau a prédit la valeur Y_{n+1} .

Dans ce cas, le pas de prédiction est de 1 puisque le réseau a prédit la valeur Y_{n+1} à partir de Y_n .

Ce pas peut être supérieur à 1, par exemple pour un pas de 2, il suffit juste pendant l'apprentissage, d'utiliser des exemples de la forme (Y_p, Y_{p+2}) au lieu des exemples du type (Y_p, Y_{p+1}) . Cependant, l'erreur sera plus importante lorsque le pas augmente.

Selon le nombre des entrées du réseau, et avec l'ensemble des valeurs Y_1, Y_2, \dots, Y_n , on peut effectuer l'apprentissage de plusieurs façons :

Si le réseau ne contient qu'une seule entrée, l'apprentissage sera identique à celui décrit précédemment.

Si le nombre d'entrées est supérieur à 1, alors il faut introduire un décalage entre les valeurs égal au nombre des entrées. Par exemple, si ce dernier est de 2, il faut introduire le couple $\begin{pmatrix} Y_2 \\ Y_1 \end{pmatrix}$ pour que le réseau nous donne Y_3 , puis $\begin{pmatrix} Y_3 \\ Y_2 \end{pmatrix}$ pour avoir Y_4 , et continuer la présentation des exemples $\begin{pmatrix} Y_p \\ Y_{p-1} \end{pmatrix}$ pour obtenir Y_{p+1} . A la fin de l'apprentissage, lorsqu'on présente $\begin{pmatrix} Y_n \\ Y_{n-1} \end{pmatrix}$ le réseau nous fournit Y_{n+1} qui est la valeur correspondant à l'instant $t_n + \Delta t$.

En ajoutant cette nouvelle valeur aux anciennes Y_1, Y_2, \dots, Y_n , et en refaisant l'apprentissage dès le début, on peut obtenir la valeur Y_{n+2} qui correspond à l'instant $t_{n+1} + \Delta t$.

De cette façon, on peut faire la prédiction de tous les points constituant le reste de la courbe représentant la propriété étudiée.

III.3 Architecture du réseau utilisé.

En utilisant les réseaux de neurones artificiels pour le développement d'une méthode de prédiction, nous faisons appel aux capacités d'approximation et de généralisation.

La plupart des applications utilisant des architectures multicouches, dont les fonctions d'activation des neurones, sont de type sigmoïde[9][12].

Néanmoins, ce type d'architecture pour une application pareille entraîne des difficultés d'apprentissage dont les plus importantes sont :

- La présence de minima locaux dans la fonction coût –fonction de l'erreur-entraînant un apprentissage sous optimal.
- La possibilité de paralysie durant l'apprentissage à cause de la saturation introduite par les sigmoïdes.
- La possibilité de 'désapprentissage' du fait du caractère non local introduit par les sigmoïdes.

Cependant, on peut penser à adapter un autre type d'architecture pour éviter ces problèmes en proposant une structure de réseaux à base radiale (figure III.1).

Ce type de réseaux, - rappelons le - peut aisément et naturellement satisfaire aux contraintes imposées au comportement de ces réseaux telles que[9] :

- Contrainte de capacité d'approximation universelle.
- Contrainte de couverture : mathématiquement, elle s'exprime par le fait que, pour chaque sortie Y_i associée à un vecteur d'entrée u_i , il existe au moins un paramètre w_j tel que la valeur $\left| \frac{\partial Y_i(u_i, w)}{\partial w_j} \right|$ (appelée fonction de sensibilité) soit suffisamment grande dans le voisinage du vecteur d'entrée u_i .
- Contrainte de généralisation locale : si la valeur de $\left| \frac{\partial Y_i(u_i, w)}{\partial w_j} \right|$ est grande dans le voisinage de u_i , elle doit être relativement proche de zéro une fois que l'on s'éloigne de ce voisinage.

Outre les couches d'entrée et de sortie, ce type de réseaux comprend une couche cachée de n unités qui ne réagissent significativement qu'à une partie restreinte de l'espace d'entrée suivant une fonction d'activation de type gaussien.

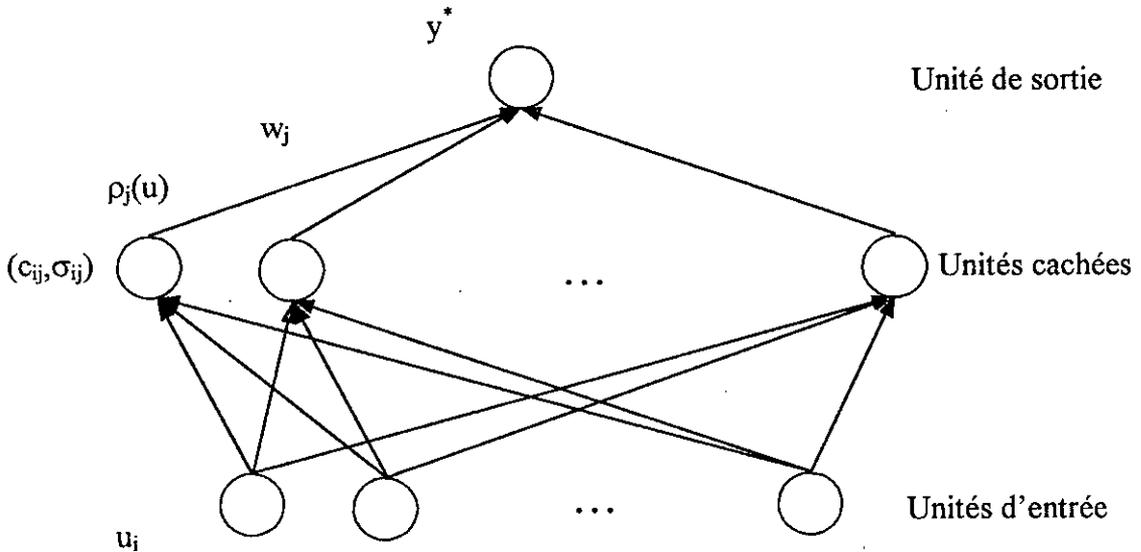


Figure III.1 réseau de neurones (RBF) à une seule sortie

u_i : la $i^{\text{ème}}$ entrée, c_{ij} : le centre du $j^{\text{ème}}$ nœud de la couche cachée correspondant à la $i^{\text{ème}}$ entrée, σ_{ij} : écart type du $j^{\text{ème}}$ nœud de la couche cachée correspondant à la $i^{\text{ème}}$ entrée, $\rho_j(u)$: est le terme de normalisation du $j^{\text{ème}}$ neurone de la couche cachée, w_j : poids synaptique du $j^{\text{ème}}$ neurone de la couche cachée, liant l'unité cachée avec l'unité de sortie, y^* : la sortie du réseau.

Ainsi, en utilisant un réseau à base radiale normalisé à une seule sortie, la valeur de cette dernière est donnée par :

$$y^* = \frac{\sum_{j=1}^n w_j \rho_j(u)}{\sum_{j=1}^n \rho_j(u)} \tag{III.1}$$

- $\left(\sum_{j=1}^n \rho_j(u) \right)$ est le terme de normalisation avec :

$$\rho_j(u) = \exp\left(-\frac{1}{2} \sum_{i=1}^m \frac{(u_i - c_{ij})^2}{\sigma_{ij}^2} \right) \tag{III.2}$$

- n est le nombre de nœuds cachés, et m est le nombre de nœuds dans la couche d'entrée.

Les vecteurs c_j ($j=1...n$) sont appelés 'centre des gaussiennes', ils ont la même dimension que le vecteur d'entrée, c'est-à-dire que chaque vecteur centre c_j à m composantes. Donc, si on regarde la globalité du réseau, ces centres c_j à m composantes chacun, constituent les n lignes de la matrice $\zeta(n,m)$ des centres du réseau entier telle que :

$$\zeta = \begin{bmatrix} c_{11} & \cdot & \cdot & c_{1m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ c_{n1} & \cdot & \cdot & c_{nm} \end{bmatrix}$$

Les centres des gaussiennes sont choisis par des techniques de regroupement de données, le but étant de choisir les centres comme points représentatifs de zones de grandes concentrations de données. Mais il existe une autre méthode plus simple qui consiste à disposer les centres en treillis régulier pour couvrir uniformément la partie utile de l'espace d'entrée et respecter ainsi les contraintes de couverture et de généralisation localisée. Les rayons des gaussiennes dans ce cas, sont fixés par les distances entre centres voisins afin d'assurer un certain degré de recouvrement entre les différentes gaussiennes (champs d'influence des différents centres).

La valeur du taux de recouvrement est donnée par [9]:

$$\tau = \exp \left(-\frac{1}{2} \frac{\left(\frac{\delta}{2}\right)^2}{\sigma^2} \right) \tag{III.3}$$

où σ est l'écart type de chaque gaussienne, et δ est une distance entre centres.

C'est cette dernière technique que nous allons utiliser dans notre travail.

Si le réseau ne contient qu'une seule entrée, le recouvrement se fait en treillis régulier, et les rayons de toutes les gaussiennes sont identiques. Dans ce cas, le nombre des unités de la couche cachée est égal au nombre des centres des gaussiennes. Si C est l'ensemble des centres, alors nombre des centres est égal au cardinal de C (fig.III.2).

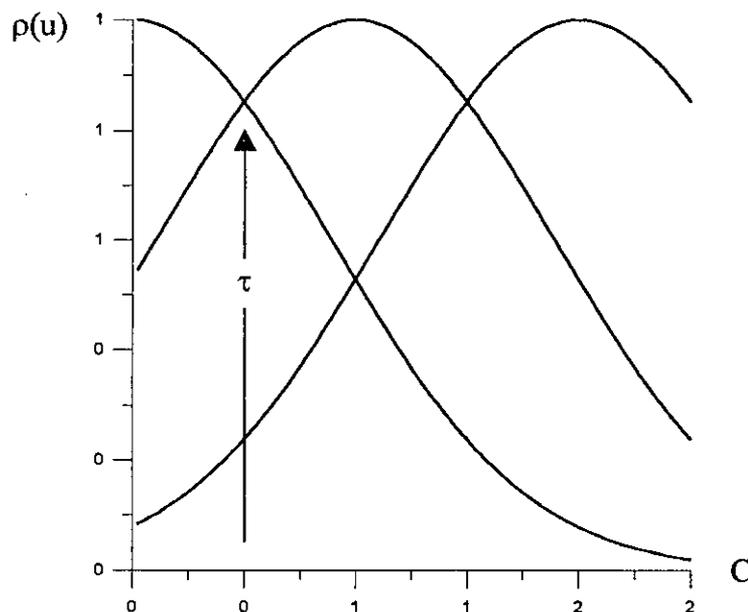


Fig.III.2 : recouvrement de l'espace d'entrée à une seule dimension pour une forme d'activation gaussienne.

Dans ce cas, la matrice des centres du réseau se réduit à un vecteur. Son transposé a la forme :

$$\zeta^t = [c_1, c_2, \dots, c_n]$$

tels que :

$$\begin{aligned} c_1 &= \min(y_1, y_2, \dots, y_n), & c_n &= \max(y_1, y_2, \dots, y_n), \\ c_2 &= c_1 + \delta, \\ & \vdots \\ c_n &= c_{n-1} + \delta \end{aligned}$$

Par contre, si le réseau contient deux entrées, le recouvrement en treillis assure aussi des rayons identiques pour toutes les gaussiennes. Mais le nombre des unités cachées est égal au nombre de couples $(c_1, c_2) \in C^2$, où C est l'ensemble des centres des gaussiennes. En d'autres termes, le nombre d'unités cachées est égal au carré du cardinal de C (fig.III.3).

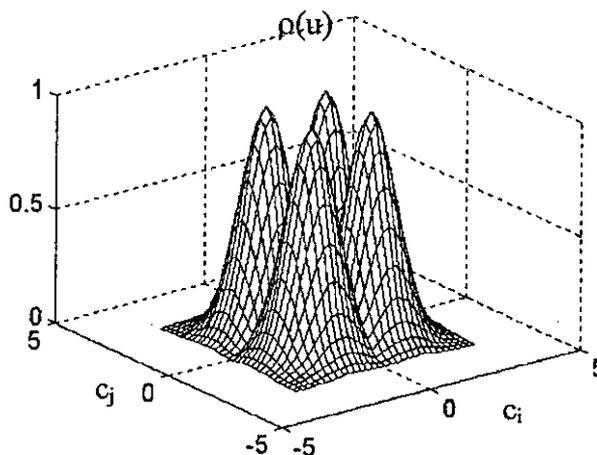


Fig.III.3 : recouvrement de l'espace d'entrée à deux dimensions pour une forme d'activation gaussienne.

Dans ce cas, les combinaisons possibles, sont données par :

$$[(c_1, c_1), \dots, (c_1, c_n), (c_2, c_1), \dots, (c_2, c_n), \dots, (c_n, c_1), \dots, (c_n, c_n)]$$

tels que :

$$c_1 = \min(y_1, y_2, \dots, y_n), \quad c_n = \max(y_1, y_2, \dots, y_n),$$

Le recouvrement de l'espace d'entrées à trois dimensions, et plus, se fait de la même façon, mais sa schématisation relève de l'imaginaire.

En outre, la disposition des centres a priori s'avère quelques fois impraticable. En effet, le nombre de centres requis est parfois trop élevé (dans le cas où l'on déciderait d'adopter une disposition en treillis régulier). Ce qui est conseillé est d'utiliser alors un nombre beaucoup plus restreint de centres et de laisser libre leur disposition, cela revient à choisir un pas temporel variable pour l'intégration numérique des équations différentielles. Dans ce cas, l'algorithme d'apprentissage devra ajuster, outre les poids w_j les positions des centres ainsi que les valeurs des rayons associés. Idéalement, le réseau s'auto-organisera pour placer un nombre élevé de centres là où la

fonction est accidentée, et espacer les centres dans les régions où la fonction est plus lisse. Ce faisant, la sortie n'est plus linéaire en ces paramètres (centres et rayons) et le risque de minima locaux apparaît par la même occasion, ces réseaux seront alors appelés 'réseaux à base radiale généralisés'. Pour l'ajustement des centres et rayons de ces réseaux, on peut utiliser l'algorithme de 'descente du gradient' comme le montre les équations ci-dessous :

Soit $J = \frac{1}{2}(y - y^*)^2$ le critère à minimiser, telle que y^* est donnée par l'équation (III.1).

Posons $R(u) = \sum_{j=1}^n \rho_j(u)$, ($R(u)$: est le facteur de normalisation), où n est le nombre de centres.

$$\begin{aligned} \frac{\partial J}{\partial w_j} &= (y - y^*) \frac{\rho_j(u)}{R(u)} \\ \frac{\partial J}{\partial c_{j,i}} &= \frac{\partial J}{\partial \rho_j} \frac{\partial \rho_j}{\partial c_{j,i}} \\ \frac{\partial J}{\partial \sigma_{j,i}} &= \frac{\partial J}{\partial \rho_j} \frac{\partial \rho_j}{\partial \sigma_{j,i}} \\ \frac{\partial J}{\partial \rho_j} &= (y - y^*) \left[\frac{w_j R(u) - \sum_{i=1}^n \rho_i(u) w_i}{R^2(u)} \right] \\ \frac{\partial \rho_j}{\partial c_{j,i}} &= \rho_j \frac{u_i - c_{j,i}}{\sigma_{j,i}^2} \\ \frac{\partial \rho_j}{\partial \sigma_{j,i}} &= \rho_j \frac{(u_i - c_{j,i})^2}{\sigma_{j,i}^3} \end{aligned} \tag{III.4}$$

les corrections s'obtiennent par :

$$\begin{aligned} \Delta w_j &= -\eta_w \frac{\partial J}{\partial w_j} \\ \Delta c_{j,i} &= -\eta_c \frac{\partial J}{\partial c_{j,i}} \\ \Delta \sigma_{j,i} &= -\eta_\sigma \frac{\partial J}{\partial \sigma_{j,i}} \end{aligned} \tag{III.5}$$

où η_w , η_c , η_σ sont les taux d'apprentissage des poids w_j , des centres c_j et des rayons σ_j respectivement.

Notons qu'il y a maintenant $(1+2m)n$ paramètres ajustables au lieu de n paramètres dans le cas d'un réseau à base radiale normal (m étant le nombre d'entrées).

Les paramètres η_w , η_c , η_σ , sont en premier lieu choisis expérimentalement. s'ils sont trop petits, on avance trop lentement, mais dans une direction optimale. Par contre, s'ils sont trop grands, on peut osciller autour de l'optimum[13].

Dans notre application de prédiction, et comme toutes les fonctions traitées sont lisses, donc un ajustement des poids seulement et une disposition des centres en treillis régulier suffit. Malgré ça, on présente quelques résultats obtenus avec un RBF généralisé.

Après avoir présenté la procédure d'apprentissage concernant un problème de prédiction et le fonctionnement des réseaux à base radiale généralisés, il nous reste maintenant le point le plus important : comment bénéficier de la souplesse de cette architecture pour prédire une valeur future ?

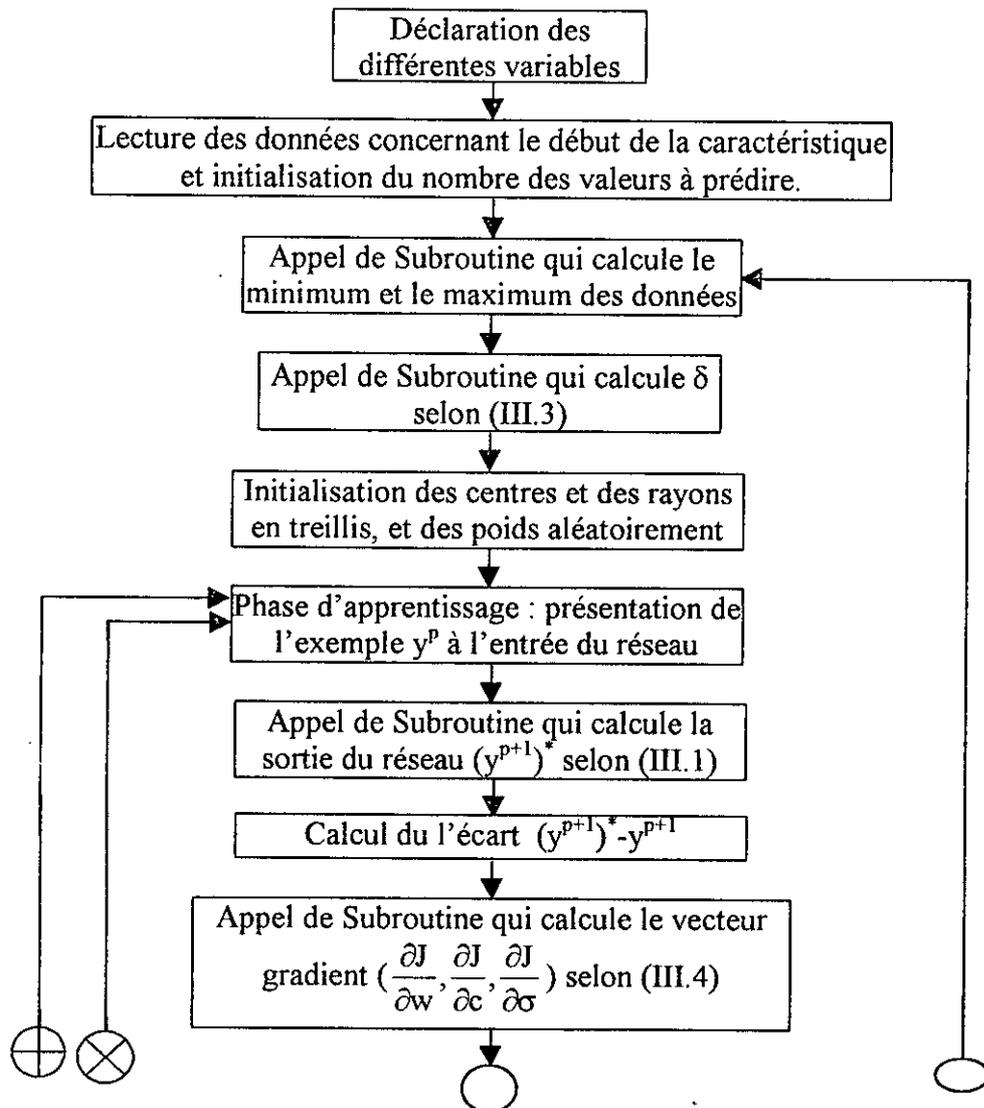
Toute l'astuce est de rendre la base d'apprentissage initiale (y_1, \dots, y_n) dynamique évolutive, c'est à dire après la prédiction de toute nouvelle valeur, on lui crée un centre supplémentaire dans la couche cachée pendant la phase d'apprentissage suivante et on élimine la première valeur y_1 de la base d'apprentissage. De cette façon, on oblige le réseau à se débarrasser de ses premiers souvenirs pour s'adapter avec les nouvelles valeurs prédites. Donc, et puisque le nombre d'exemples est constant pour chaque apprentissage –on élimine la première valeur après la prédiction d'une nouvelle valeur– alors le nombre des centres du réseau l'est aussi.

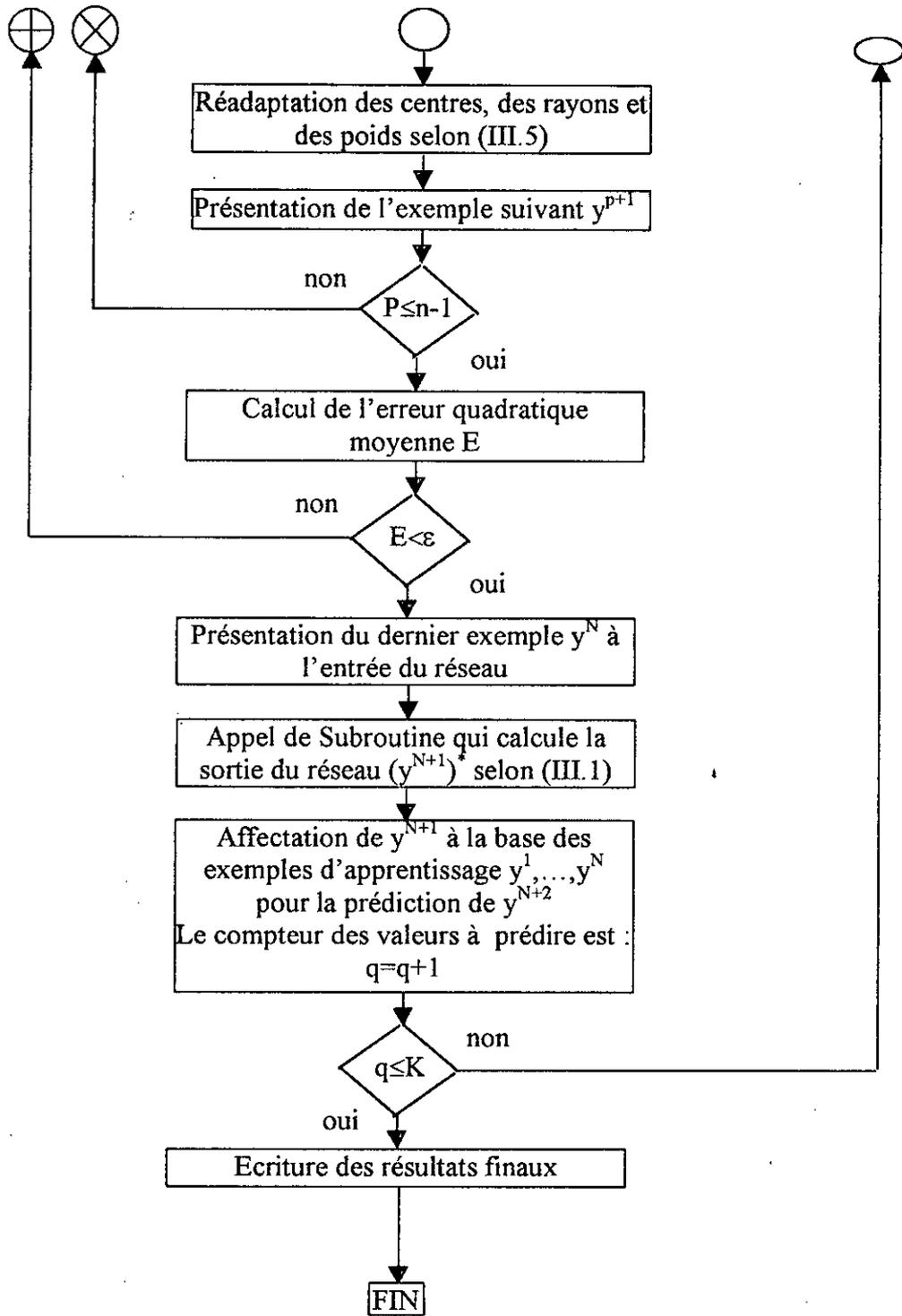
Nous avons utilisé une variété de techniques d'optimisation décrites dans le premier chapitre pour l'adaptation des poids du réseau telles que la méthode de descente du gradient avec un taux fixe ou variable, avec ou sans momentum, mais malheureusement, et comme nous l'avons vu au premier chapitre, toutes ces méthodes dépendent fortement du critère à minimiser et par suite du système étudié. Pour notre application, après l'adaptation des poids du réseau avec une de ces méthodes pour prédire une valeur future, nous avons remarqué que la qualité de prédiction se dégrade énormément si on garde les mêmes paramètres pour extrapoler une autre fonction. Donc, il nous a fallu chercher de nouveaux paramètres ce qui rend très difficile notre travail de prédiction.

Pour pallier tout ces problèmes de paramétrisation, nous avons utilisé la méthode aléatoire d'optimisation ROM dont le principe est basé sur la génération des séquences aléatoires afin de déplacer les valeurs des poids dans l'espace à la recherche de l'optimum, la combinaison fournissant le meilleur résultat est retenue à chaque itération. L'algorithme de la ROM est cité au premier chapitre.

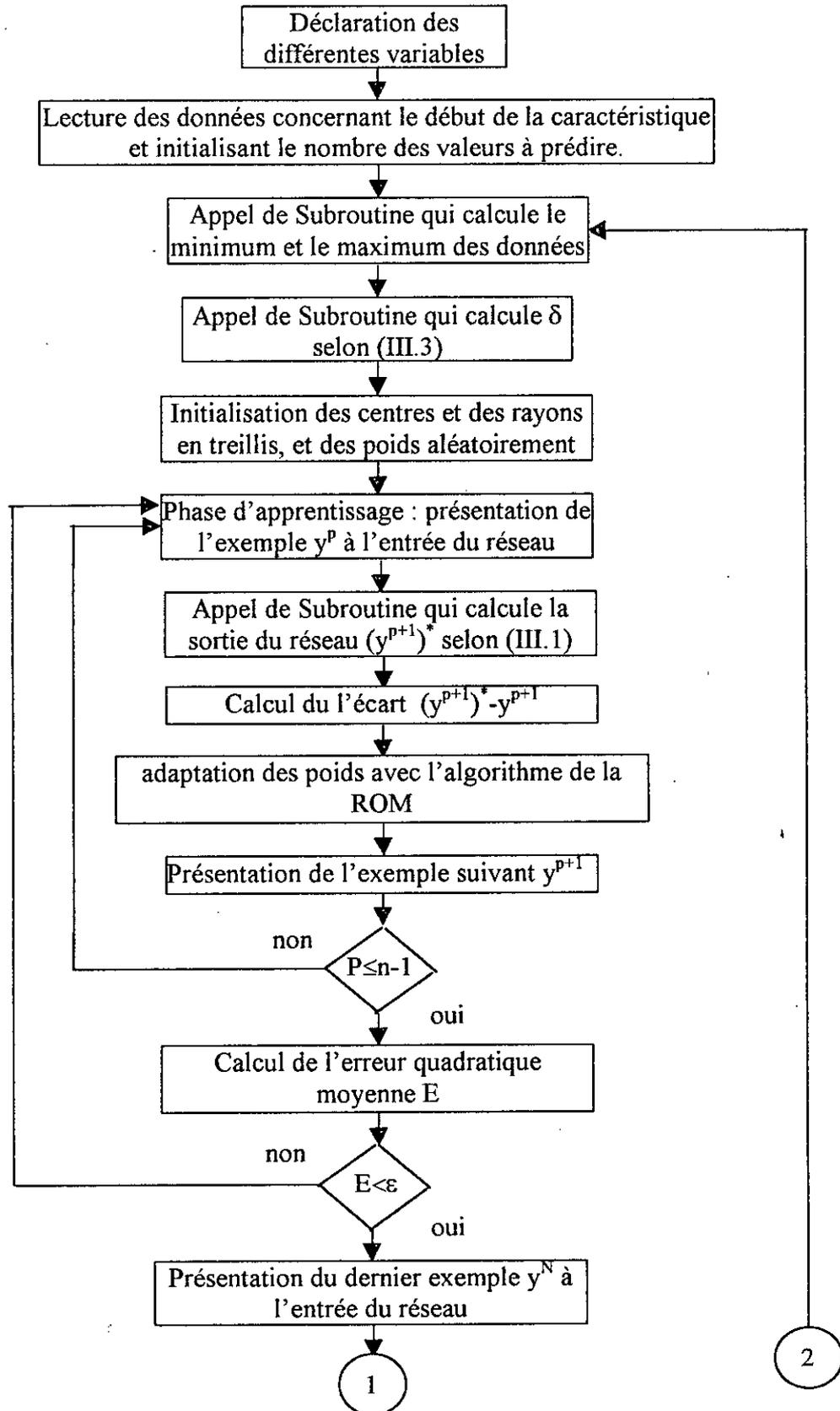
Nous présentons ci-dessous la structure générale de nos programmes :

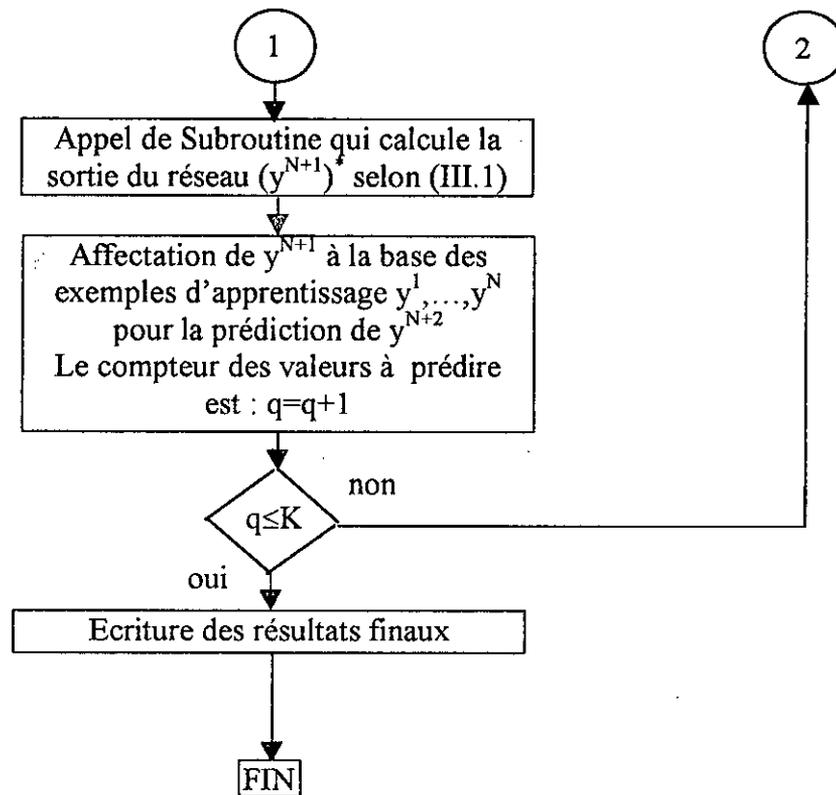
- Organigramme d'un RBF généralisé :

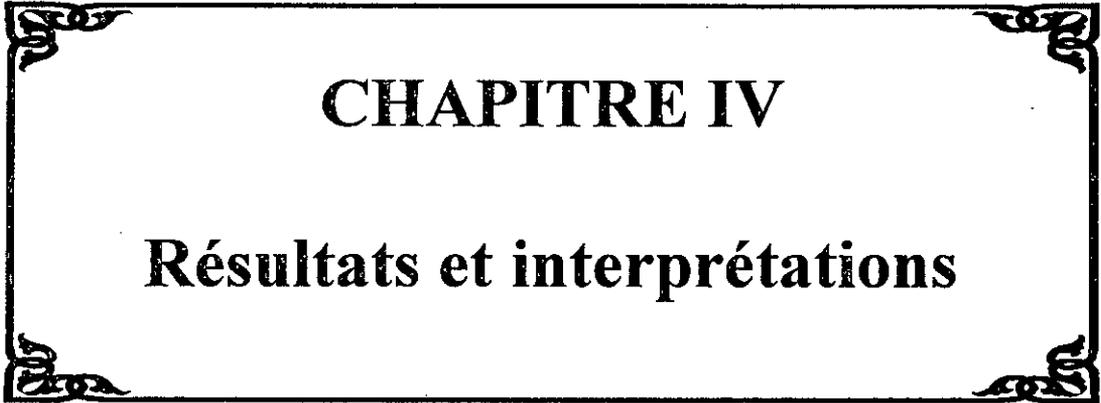




• Organigramme d'un RBF basé sur la ROM







CHAPITRE IV

Résultats et interprétations

Dans ce chapitre nous allons présenter quelques résultats de prédiction obtenus avec un réseau de neurones RBF dont les centres et les rayons sont disposés en treillis régulier, et les poids synaptiques sont ajustés par l'algorithme d'optimisation aléatoire.

Nous allons exposer également quelques résultats de prédiction, mais cette fois avec un RBF généralisé dont tous les paramètres (centres, rayons et poids) sont ajustables avec l'algorithme de la rétro-propagation, mais vu les problèmes du choix des trois taux d'apprentissage η_w , η_c , η_σ , relativement aux centres, aux rayons et aux poids respectivement que nous avons détaillé dans le chapitre précédent, les résultats suivants sont présentés à titre illustratif seulement.

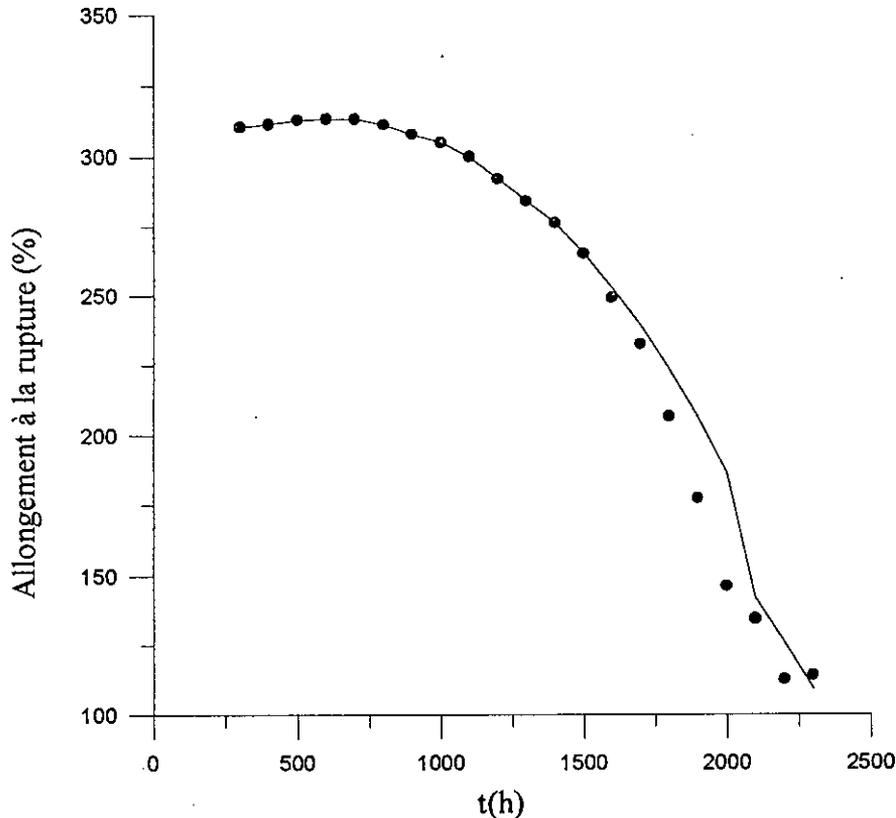


Figure IV.1 exemple d'une courbe représentant la variation de l'allongement à la rupture en fonction du temps de vieillissement, l'extrapolation est obtenue avec un réseau RBF généralisé entraîné avec la méthode de rétro-propagation avec $\eta_w=.18$, $\eta_c=.08$, $\eta_\sigma=.05$.

Maintenant, nous tournons vers l'algorithme d'optimisation aléatoire en commençant par trouver le temps d'apprentissage minimal (le temps minimal pendant lequel nous devons effectuer nos essais tout en garantissant une bonne qualité de prédiction des points constituant le reste de l'allure expérimentale). Pour cela, nous démarrons notre apprentissage sur un intervalle de temps relativement grand, puis nous diminuons progressivement jusqu'à ce que nous obtenons l'intervalle le plus petit garantissant une erreur acceptable. Puis, nous faisons varier le pas d'apprentissage à fin de trouver la meilleure division et par suite le meilleur nombre d'exemples convenable pour notre programme.

Pour la caractéristique de pertes de masse du PVC qui s'étale sur une période de temps de 4000 heures, nous avons démarré l'apprentissage sur un intervalle de 2500 heures puis, nous l'avons diminué à 2250 heures, puis à 2000 heures puis finalement à 1500 heures. Les résultats sont représentés ci dessous :

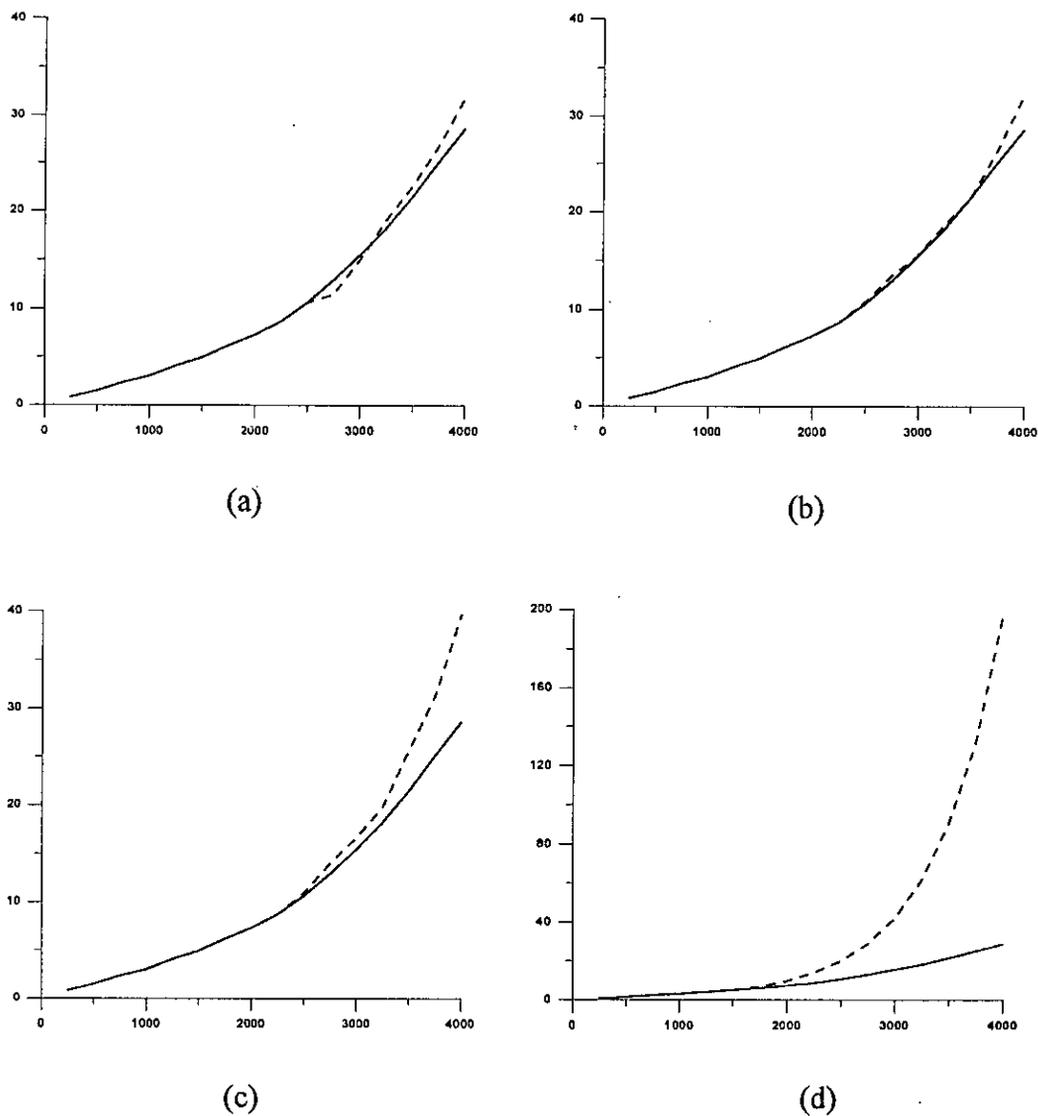
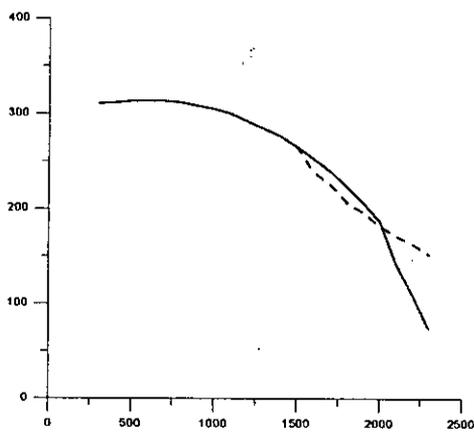


Figure IV.2 : variation de la perte de masse surfacique en fonction du temps de vieillissement :

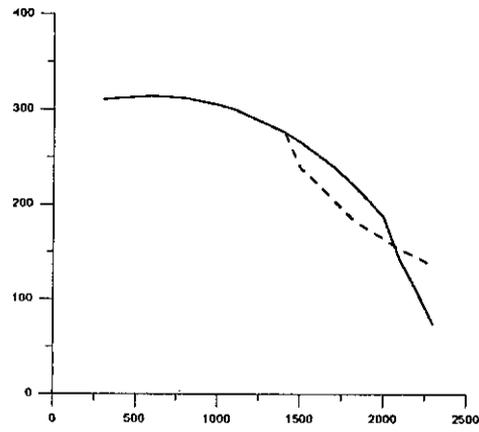
- a- Apprentissage sur un intervalle de 2500h.
- b- Apprentissage sur un intervalle de 2250h.
- c- Apprentissage sur un intervalle de 2000h.
- d- Apprentissage sur un intervalle de 1750h.

On remarque d'après la figure (IV.2) que le temps pour lequel le réseau peut donner des valeurs plus proches des valeurs réelles est celui de 2250 heures (IV.2.b), et si on essaye de le faire diminuer encore, on risque de s'éloigner de la courbe exacte.

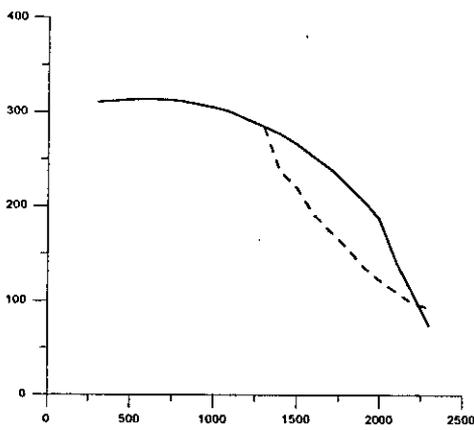
Pour la caractéristique de l'allongement à la rupture et pour la contrainte de température de 120°C qui s'étale sur 2500 heures, nous avons commencé l'apprentissage sur un intervalle de 1600 heures, puis 1500h, en suite 1400h et finalement 1300 heures.



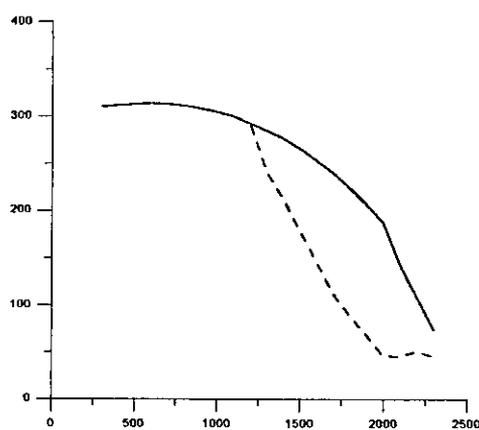
(a)



(b)



(c)



(d)

Figure IV.3 : variation de l'allongement à la rupture en fonction du temps de vieillissement sous une contrainte de 120°C

- a- Apprentissage sur un intervalle de 1600h.
- b- Apprentissage sur un intervalle de 1500h.
- c- Apprentissage sur un intervalle de 1400h.
- d- Apprentissage sur un intervalle de 1300h.

Dans cette caractéristique (figure.IV.3) nous remarquons que l'allure pour laquelle le réseau suit convenablement la courbe réelle est celle de la figure(IV.3.c) où l'apprentissage a été effectué sur un intervalle de 1400 heures, pour ce qui concerne les autres allures nous constatons que la diminution de l'intervalle d'apprentissage provoque la dégradation des performances du réseau.

Pour la caractéristique de l'allongement à la rupture et pour la contrainte de température de 100°C qui s'étale sur 5000 heures, nous avons commencé l'apprentissage sur un intervalle de 3100 heures puis 2900h, en suite 2700h et finalement 2500 heures.

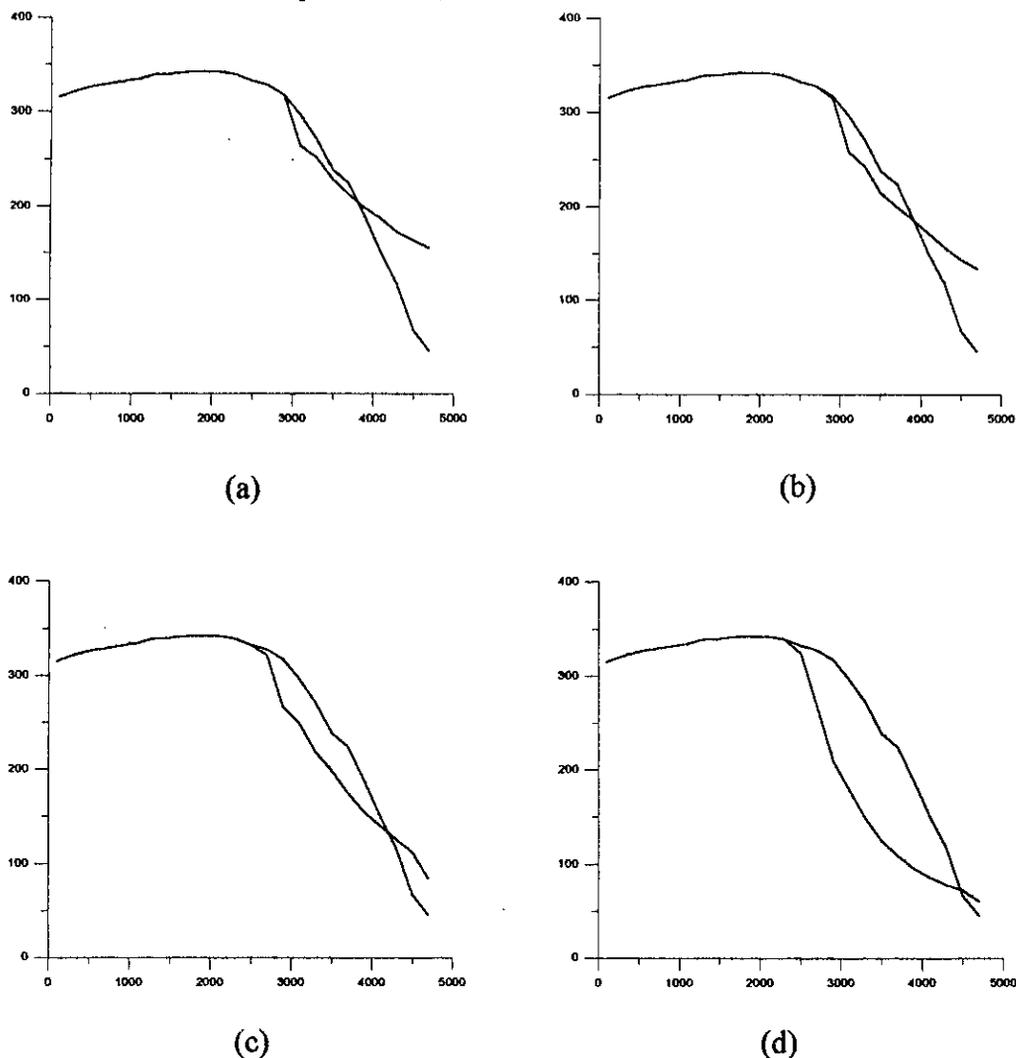


Figure IV.4 : variation de l'allongement à la rupture en fonction du temps de vieillissement sous une contrainte de 100°C
 a- Apprentissage sur un intervalle de 3100h.
 b- Apprentissage sur un intervalle de 2900h.
 c- Apprentissage sur un intervalle de 2700h.
 d- Apprentissage sur un intervalle de 2500h.

Comme pour les contraintes précédentes, nous remarquons que l'intervalle qui nous intéresse mieux est celui qui correspond à la figure(IV.4.c).

Il nous reste maintenant l'allure de la caractéristique pour la contrainte 80°C, qui s'étale sur un intervalle de temps de 7000h. nous avons commencé l'apprentissage sur un intervalle de 4900 heures puis nous l'avons diminué à 4500h

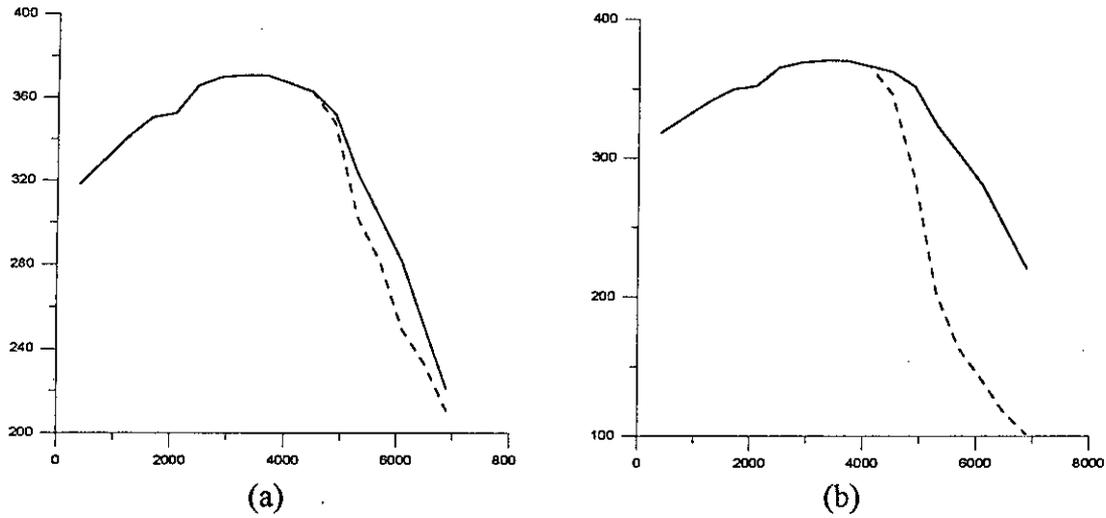


Figure IV.5 : variation de l'allongement à la rupture en fonction du temps de vieillissement sous une contrainte de 80°C
a- Apprentissage sur un intervalle de 4900h.
b- Apprentissage sur un intervalle de 4500h.

Nous remarquons sans doute que l'intervalle de 4900h correspondant à la figure (IV.5.a) est le plus convenable pour cette contrainte.

Maintenant nous allons faire l'apprentissage de chaque courbe sur son intervalle optimal pour les tracer simultanément dans le même graphe à fin de pouvoir exploiter les résultats sous forme d'indice de température.

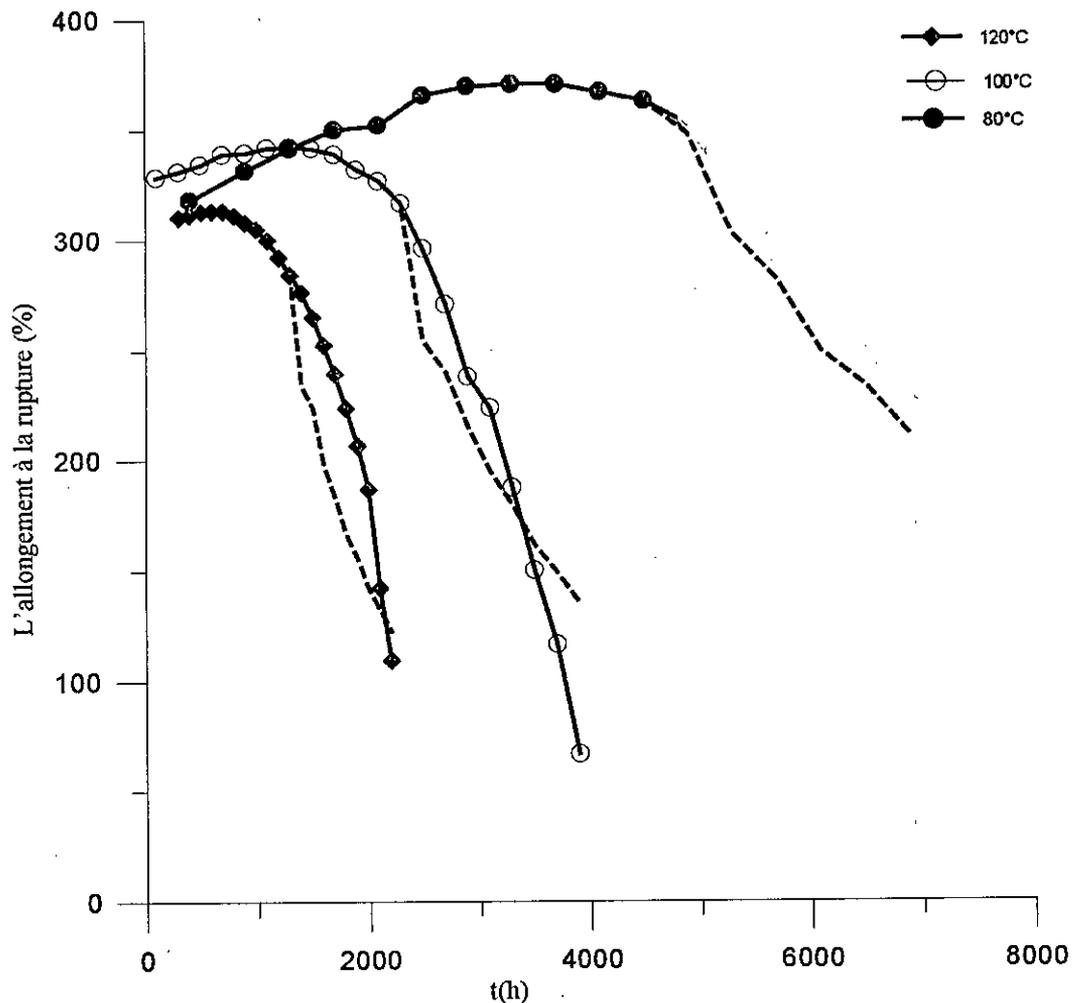


Figure IV. 6. variation de l'allongement à la rupture en fonction du temps de vieillissement.

----- sortie du réseau.
 _____ Courbe réelle.

De même, la caractéristique de la résistance à l'attraction est représentée à la figure(IV.8) :

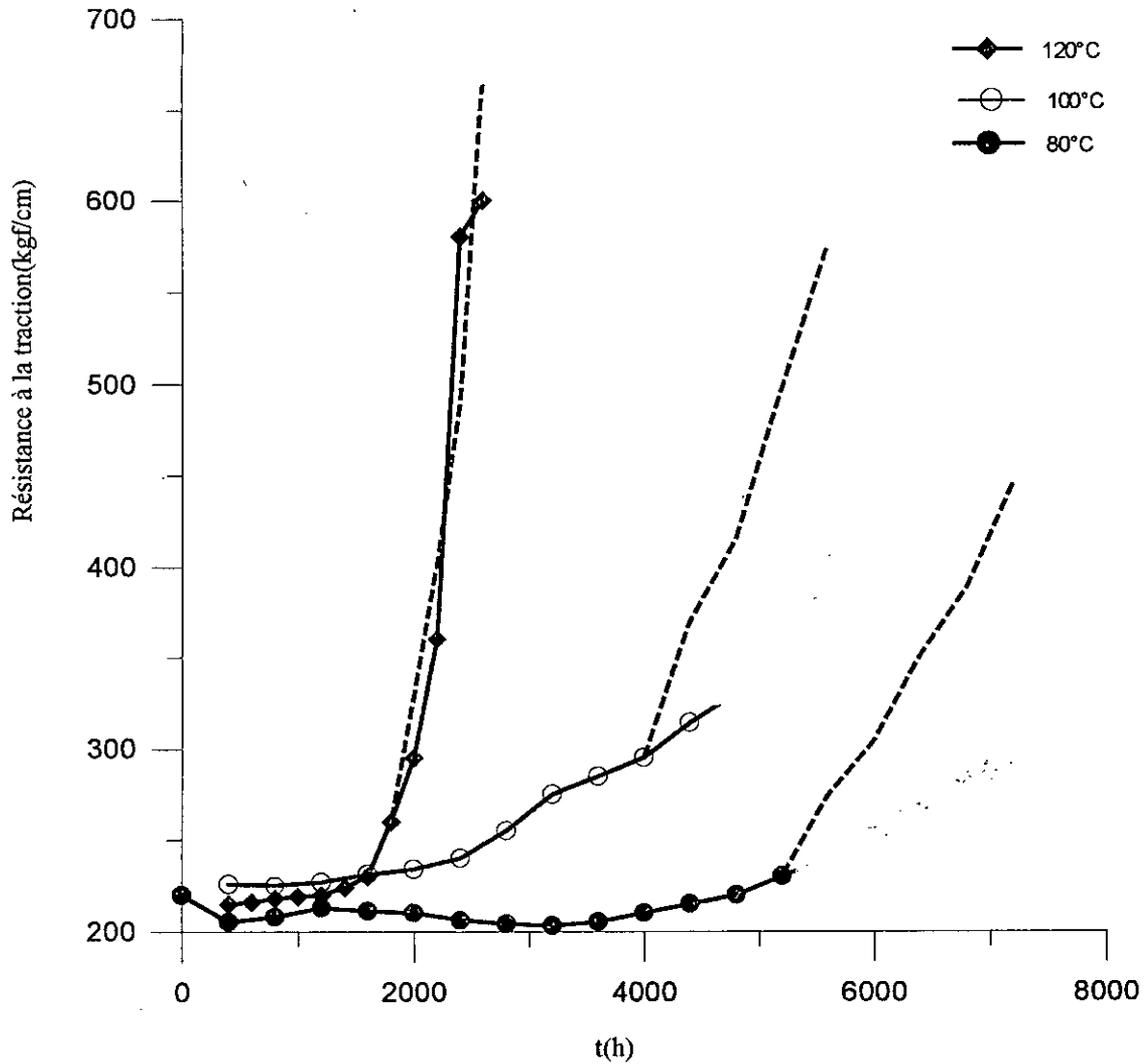


Figure.IV.8. Variation de la résistance à la traction en fonction du temps de vieillissement.
 ----- sortie du réseau.
 ———— Courbe réelle.

L'écart observé entre la courbe réelle et celle du réseau pour les contraintes 80°C et 100°C, résulte du fait que la première a été extrapolée à la main afin d'atteindre le critère de fin de vie. Donc, nous pouvons dire que le réseau a corrigé ces résultats.

Pour tester encore une fois les performances de notre programme, nous allons effectuer l'apprentissage sur le meilleur intervalle qui répond à nos besoins, mais cette fois ci, nous faisons varier le pas de discrétisation. A partir de là, on peut estimer le pas optimal pour lequel notre prédiction est meilleure.

Considérons la courbe représentant la perte de masse surfacique en fonction du temps de vieillissement, l'intervalle d'apprentissage dans ce cas est de 2250h, nous démarrons avec un pas de 250h soit 10% de l'intervalle globale, puis nous le diminuons à 125h soit 5% et finalement à 62.5h soit 2.5% (figureIV.9).

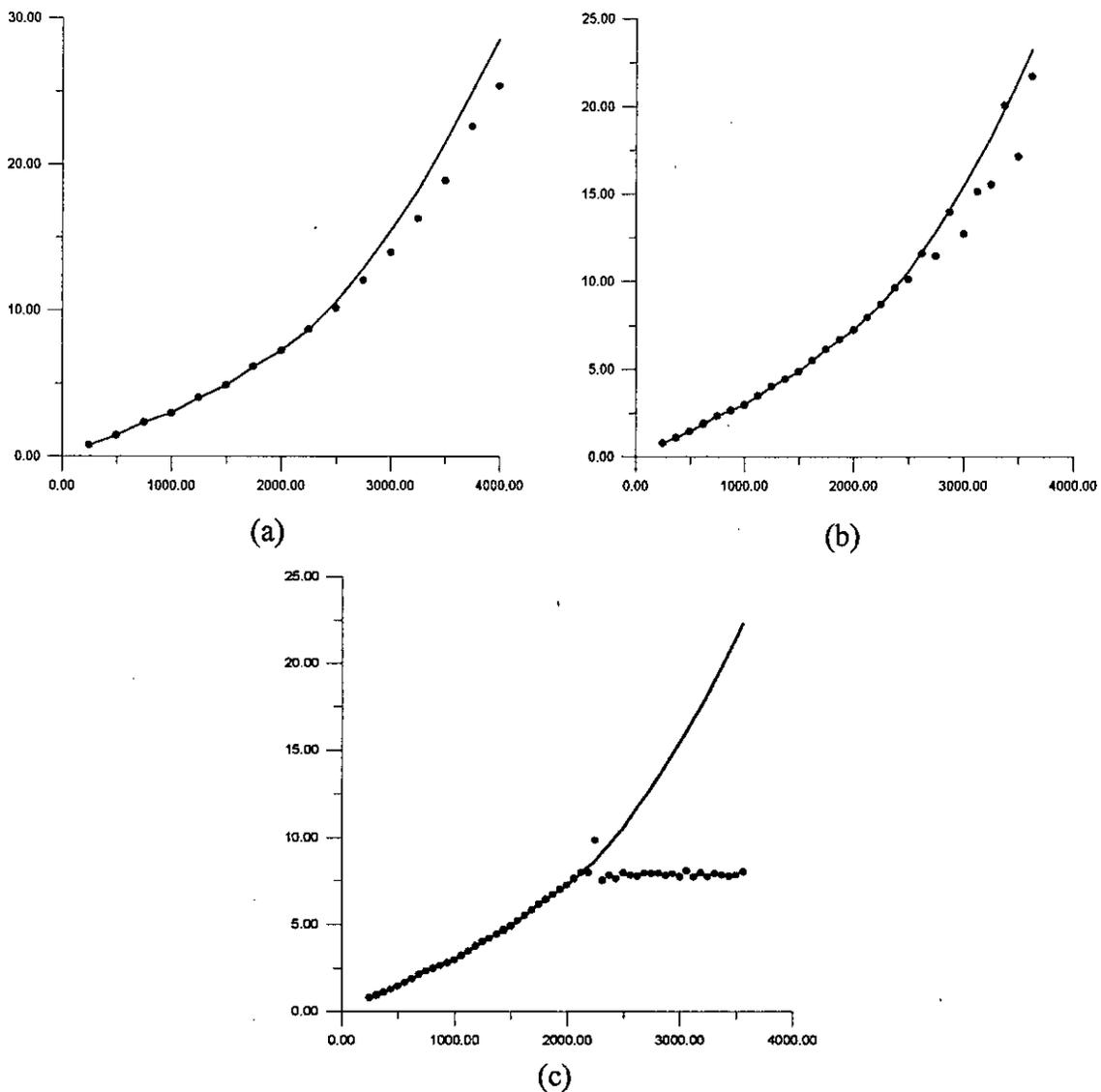


Figure IV.9. variation de la perte de masse surfacique en fonction du temps de vieillissement.

- (a) apprentissage avec un pas de 10% de l'intervalle global.
- (b) apprentissage avec un pas de 5% de l'intervalle global.
- (c) apprentissage avec un pas de 2.5% de l'intervalle global.

D'après la figure(IV.9.a), on remarque que pour la caractéristique de la perte de masse surfacique que le pas optimal qu'il faut prendre représente 10% de l'intervalle d'apprentissage.

Après la caractéristique de la perte de masse surfacique qui est une fonction croissante, nous traitons maintenant un exemple de fonction décroissante, soit par exemple la caractéristique de l'allongement à la rupture sous la contrainte de température 120°C qui s'étale sur 2400h, l'intervalle d'apprentissage étant de 1500h . nous considérons deux cas de figure : le premier est pris pour pas de 100h soit 6.66% tandis que le deuxième est de 50h soit un pourcentage de 3.33% de l'intervalle d'apprentissage (figure IV.10)

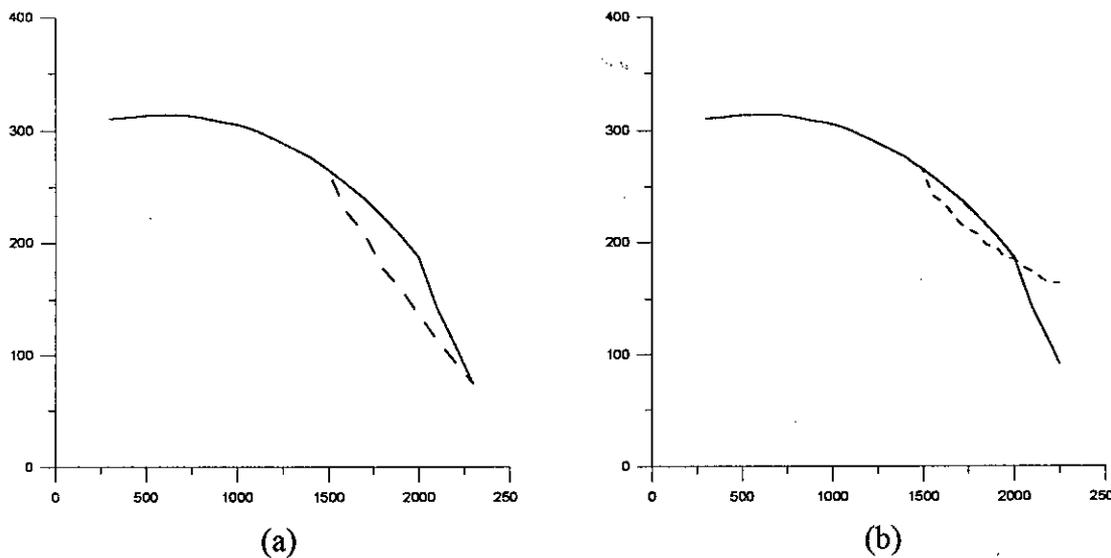
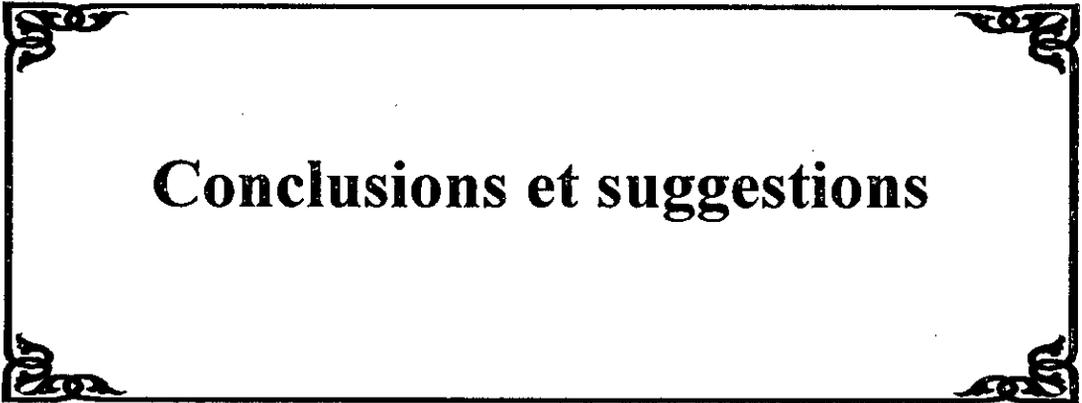


Figure IV.10. variation de l'allongement à la rupture en fonction du temps de vieillissement sous la contrainte 120°C.

- (a) apprentissage avec un pas de 6.66% de l'intervalle global.
- (b) apprentissage avec un pas de 3.33% de l'intervalle global.

On remarque que la qualité de prédiction se dégrade en diminuant le pas d'apprentissage.



Conclusions et suggestions

Conclusions et suggestions

Nous avons commencé notre travail par une large recherche bibliographique qui nous a permis d'explorer les différents types de réseaux de neurones artificiels mis en œuvre jusqu'à ce jour, afin de pouvoir sélectionner la structure de réseau la plus adéquate dont les caractéristiques peuvent nous permettre d'élaborer un modèle de prédiction des différentes propriétés mécaniques et diélectriques des polymères et en particulier du Polychlorure de Vinyle PVC utilisé dans l'isolation des câbles moyenne et haute tension.

Au début, nous avons choisis une architecture multicouche dont les fonctions d'activation des neurones sont de type sigmoïde, mais vus les problèmes qui peuvent se produire tels que la possibilité de paralysie durant l'apprentissage et la divergence du fait de la saturation et le caractère non local introduits par les sigmoïdes, nous avons abandonné ces réseaux pour que notre choix se stabilise sur un autre type d'architecture : les réseaux de neurones à base radiale. Ces derniers, en les dotant d'un apprentissage prédictif que nous avons développé au troisième chapitre, peuvent répondre très bien à tous les critères que nécessite un problème de prédiction. A cet effet nous avons utilisé plusieurs algorithmes d'optimisation paramétrique pour ajuster les différents paramètres de ces réseaux tels que la rétropropagation, la rétropropagation avec momentum, la rétropropagation avec des taux d'apprentissage adaptatifs et la méthode d'optimisation aléatoire. Notre but était bien sûr d'obtenir un réseau universel robuste, dont les performances ne se dégradent pas en passant de la prédiction d'une allure à une autre, mais malheureusement tous les algorithmes basés sur la rétropropagation qui est basée à son tour sur la descente du gradient ne répondaient pas à ce point très important à nos yeux. La raison en est principalement la difficulté de choix des taux d'apprentissage η_w , η_c , η_σ , des poids, des centres et des rayons respectivement, et malgré ça, nous avons obtenus quelques résultats avec ces réseaux pour l'adaptation des poids.

Cependant, le choix de la méthode d'optimisation aléatoire ROM était indispensable pour nous, celle ci permet d'entraîner le système d'une manière plus souple, et elle ne dépend nullement de la nature mathématique du système sur lequel elle est appliquée. En plus de ça, avec cette méthode et contrairement à la méthode de la rétropropagation, l'erreur ne peut faire que diminuer en évitant ainsi toute divergence éventuelle.

- Pour les allures des propriétés traitées, nous avons remarqué que plus la courbe est aplatie plus la qualité de prédiction se dégrade, cela nous oblige à mener l'apprentissage sur des intervalles relativement grands pour que le réseau puisse reconnaître l'allure croissante de l'allure décroissante. Ce qui affecte le but essentiel de notre travail qui consiste à effectuer l'apprentissage sur un intervalle minimale pour avoir une plage de prédiction maximale.
- Le programme que nous avons élaboré peut prédire une durée moyenne de 1500 heures (deux mois environ) avec une marge d'erreur acceptable, cette dernière est dans les environs de 5%.

Conclusions et suggestions

- Pour que notre programme puisse donner de bons résultats, les mesures avec lesquelles on effectue l'apprentissage doivent être espacées d'un intervalle de temps de l'ordre 10% de l'intervalle global d'apprentissage.
- Avant, on était obligé de compléter les différentes courbes de vieillissement à la main. Maintenant, et grâce aux réseaux de neurones artificiels nous disposons d'un outil qui a donné des résultats très encourageants.
- Pour les propriétés qui correspondent à des contraintes de température relativement basse, et qui nécessitent un intervalle de temps très long avant d'atteindre le critère de dégradation, les réseaux de neurones artificiels constituent une excellente solution pour pallier ce problème en garantissant ainsi une réduction considérable des coûts des essais.
- Une autre idée qui semble être très prometteuse, consiste à utiliser des algorithmes génétiques pour optimiser les différents paramètres du réseau ; leur principe consiste à choisir une population de réseaux différents. Chacun de ces réseaux représente un individu. Avec certaines techniques convenables de croisement entre tous les individus, on obtient un réseau qui englobe toutes les performances présentes dans la population initiale.

Les réseaux de neurones artificiels sont des systèmes à traitement parallèle. L'utilisation, par conséquent, de calculateurs à traitement parallèle, au lieu de calculateurs à traitement séquentiel, permettrait une réduction considérable des temps d'exécution

Annexe

I. Le modèle Gaussien

$$g(t) = \frac{A}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(t-m)^2}{2\sigma^2}\right\} B(t) \equiv G(m, \sigma^2)$$

$g(t)$: densité du signal, A : amplitude, σ^2 : variance, σ : écart type, m : abscisse de son axe de symétrie.

$$\text{Forme réduite } G(0,1) = \frac{A}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\}$$

Propriétés :

$$g_m = \frac{A}{\sigma\sqrt{2\pi}}$$

points d'inflexions : $t = \pm\sigma$ et $g_i(t) = \frac{g_m}{\sqrt{e}}$

surface = A , son enveloppe : $\frac{g_m}{\sqrt{e}} \frac{\sigma}{H}$.

II. Variables aléatoires Gaussiennes

II.1. Variable Gaussienne scalaire

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x_0^2}{2\sigma^2}\right) B(x)$$

x_0 : valeur centrée de x : $x_0 = x - \xi(x)$

II.2. Couple aléatoire Gaussien

$$\zeta(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)} \left[\frac{\hat{x}_1^2}{\sigma_1^2} - \frac{2r\hat{x}_1\hat{x}_2}{\sigma_1\sigma_2} + \frac{\hat{x}_2^2}{\sigma_2^2} \right] \right\}$$

II.3. Système aléatoire Gaussien

$$f(x_1, x_2, \dots, x_n) = A_1 \exp\{-Q_1(x_1, x_2, \dots, x_n)\}$$

avec : $Q_1 = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$ et $a_{ij} = a_{ji}$

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^n \sqrt{|\mu|}} \exp\left\{-\frac{1}{2} (\hat{x})^T (\mu)^{-1} (\hat{x})\right\}$$

$$(\hat{x})^T = [x_1 - m_1, x_2 - m_2, \dots, x_n - m_n] \quad |\mu| = \det \mu.$$

Annexe

$$\mu = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \dots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \dots & \dots & \mu_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \mu_{n1} & \mu_{n2} & \dots & \dots & \mu_{nn} \end{bmatrix} \quad \mu_{ij} = \mu_{ji}, \quad \mu_{ii} = \sigma_{ii}^2$$

(8) contient :

les moyennes individuelles $m_i = \xi\{X_i(\varepsilon)\}$

les écarts types : $\sigma_{ii} = \sqrt{\mu_{ii}}$

les coefficients de corrélation : $r_{ij} = \frac{\mu_{ij}}{\sqrt{\mu_{ii}\mu_{jj}}}$, $-1 < r_{ij} < 1$

Il arrive souvent que les variables aléatoires $X_1(\xi), \dots, X_n(\xi)$ ont des moyennes nulles. Des variances égales et sont décorrélées :

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp\left\{-\frac{X_1^2 + X_2^2 + \dots + X_n^2}{2\sigma^2}\right\}$$

REFERENCES BIBLIOGRAPHIQUES

REFERENCES BIBLIOGRAPHIQUES

- [1] DAVALO, P. NAIM, : Des Réseaux de Neurones, Editions Eyrolles (1989).
- [2] Y.M.YEDDOU : Etude de synthèse sur les Réseaux de Neurones et leurs applications, thèse de magistère en Automatique à l'ENP, (1998).
- [3] WERBOS.P.J :Backpropagation Trought Time, what is does and how to do it ?, Proceedings of IEEE, VOL.78, N°10, October (1995).
- [4] HAMZILB et S. LABIOD :Identification et commande des systèmes dynamiques par Réseaux de Neurones, Projet de fin d'étude en Automatique à l'ENP, (1995).
- [5] P.VIDAL-NAQUET : LA MEMOIRE, LA RECHERCHE MENSUEL 267 Juillet-Aout 1994, volume 25.
- [6] FREEMAN. J. D. and M. SKAPURA. : Neural Networks, Algorithms, applications and programming technique. Adison Wesley company (1992).
- [7] JOUDOUIN.P : Les Réseaux de Neurones : principe et définitions. Editions HERMES (1995).
- [8] KARRAYANIS.N.B and A.N. VENET SANOPOULOS, Artificial Neural Networks, Learning, performance, évolution and application. Editions Kluwer Academic Publishers (1993).
- [9] RENDERS.J.M, Algorithmes génétiques et Réseaux de Neurones. Application à la commande des processus. Editions HERMES (1995).
- [10] KHEMAISSIA.S and A.S. MORRIS. Review of Networks and choice of radial basis function Networks for systems identification. Technologies avancés. NG. Pp 55-85.(1994)
- [11] M. HENICHE.F. BOUDJEMA and M. ATTARI , Simulating Nonlinear functions with Artificial Neural Networks. Proceeding of the international conference on signals and systems, ICSS'94. September 24-26 1994.
- [12] A. AZIBI. K. MECCELEM , Réseaux de Neurones : outil de traitement d'information. Application : Prédiction des prix du pétrole, Projet de fin d'étude en Génie Industriel, ENP. (1998).
- [13] C. FOULARD, S. GENTIL, J.P. SANDRAZ, Commande et régulation par calculateur numérique. Editions EYROLLES (1987).
- [14] Y. MECHERI , Influence du vieillissement thermique continu sur les propriétés diélectriques et mécaniques du polyéthylène reticulé chimiquement utilisé dans l'isolation des câbles moyennes tension, Thèse de Magister, ENP, Alger (1998).

REFERENCES BIBLIOGRAPHIQUES

[15] M. NEDJAR , Influence du vieillissement thermique sur les propriétés du Pylchlorure de Vinyle utilisé dans l'isolation des câbles de haute tension, Thèse de Magister, UTO, Tizi-Ouzo (1991).

[16] A. HALLASSA, K. BADACHE, Influence du vieillissement thermique sur les propriétés du PRC utilisé dans l'isolation des câbles moyenne tension, Projet de fin d'étude en électrotechnique, ENP, Alger (1996).

[17] A. MEJDOUB, ``Influence des vieillissements thermique et électrique sur les propriétés du Polyéthylène réticulé chimiquement utilisé dans les câbles de moyenne tension``, Thèse de Magister ENP, Alger (1997).

[18] A. BOUBAKEUR, M. NEDAJR et R. KHALI, Thermal endurance of polyvinil chloride from mechanical properties, Congrès Arabe de Mécanique, Damas (1997).