

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Centre Universitaire D'El-Oued
Institut des Sciences Et Technologie
Département de Génie Electrique Et Hydraulique



MEMOIRE

Pour l'Obtention du Diplôme d'ingénieur d'état en Électrotechnique
Option : Réseaux Electrique

THÈME

**Prédiction de quelques propriétés d'un huile
De transformateur utilisant les réseaux
De neurones artificiels**

Présenté par :

Mr: BENNADJI ABDEN NACER

Mr: GUEMOULA SMAIL

Mr: KHELIEL HAMZA

Encadré par :

Mr: GUEHRAR YUCEF



Promotion 2007

Sommaire

Introduction générale.....	1
----------------------------	---

Chapitre .I. les réseaux de neurones artificiel

I.1 Introduction.....	2
I.2 Les neurones.....	2
I.2.1. Le neurone biologique.....	2
I.2.2. Le neurone artificiel.....	3
I.2.2.1.Definition.....	3
I.2.2.2.L'état des neurones.....	3
I.2.2.3. Les connexions entre neurones.....	4
I.2.2.3.1. Le voisinage.....	4
I.2.2.3.2. Les connexions.....	4
I.2.2.3.3. les poids des connexions.....	4
I.2.2.4. Description d'un neurone formel.....	4
I.3.Modèle mathématique général du neurone.....	5
I.3.1.Fonction discriminant (fonction de base)	5
I.3.1.1.Fonction de base linéaire LBF	6
I.3.1.2.Fonction de base radiale RBF (radial basis function)	6
I.3.1.3.Fonction de base elliptique EBF(elliptic basis function)	6
I.3.2.Fonction dynamique linéaire.....	6
I.3.3.Fonction non dynamique non linéaire.....	6
I.4.Classification d réseaux de neurones.....	8
I.5.Les types d' apprentissage.....	9
I.6.Réseaux de neurones a apprentissage non supervise.....	10
I.6.1.Mémoire associative.....	10
I.6.1.1.Réseau de Hopfield.....	10
I.6.1.2.Mémoires associatives bidirectionnelles (RAM)	11

I.6.1.3.Memoires associatives line ares optimisées(OLAM)	11
I.6.2.Reseaux compétitifs.....	12
I.6.2.1.Couche de grossberg.....	12
I.6.2.2.Adaptive resonance theory ART.....	13
I.7.Apprentissage supervise.....	14
I.7.1.Regles du perceptron et d' adaline. madaline.....	14
I.7.1.1.Fonction de coût et méthode de gradient.....	16
I.7.1.2.Le gradient stochastique :la règle de widrow-hoff.....	18
I.7.2.Reseaux multicouches statiques.....	19
I.7.3.Reseaux a fonction de base linéaire LBF(linear basis function)	19
I.7.3.1.La retro-probagation du gradient	20
I.7.3.2.L'algorithme de retro-probagation du gradient.....	20
I.7.3.2.1. Probagation des etats.....	21
I.7.3.2.2. Retro-probagation. de l'erreur.....	22
I.7.3.2.3.Backpropagation avec momentané.....	23
I.7.4.Une nouvelle architecture.....	24
I.7.4.1.Les unités.....	24
I.7.4.2.L'organisation des cellules.....	25
I.7.5.Methode d' optimisation aléatoire ROM.....	25
I.7.6.Reseaux a fonction de base radiale RBF.....	26
I.7.6.1.Architecture et fonctionnement des réseaux RBF.....	27
I.7.6.2.Apprentissage des réseaux RBF généralises.....	28
I.7.6.3.Reseau RBF et approximation de fonction.....	29
I.8.Conclusion.....	29

Chapitre .II.Les huile des transformateurs

II.1.Introdction.....	31
II.2.Les huiles minérales isolantes.....	31
II.2.1.Les rôles d'une huile isolantes.....	31
II.2.2.Choix d'un diélectrique liquide.....	31
II.2.3.Caracteristiques d'un huile minérale isolante.....	31
II.2.3.1.Proprites électriques.....	31
II.2.3.2. Proprites physiques	33
II.2.3.3. Proprites chimiques	34
II.2.4.Vieillessement des huile de transformateur.....	34
II.2.4.1. Vieillessement thermique des huile de transformateur.....	35
II.2.4.2.Loïs du Vieillessement des isolants.....	35
II.3.Les mécanismes de claquage des diélectriques liquides.....	36
II.3.1.Mecanisme électronique.....	36
II.3.1.1.Dissociation renforcée par le champ.....	36
II.3.1.2.Production des charges aux interfaces.....	36
II.3.1.3.Consequence du transfert de charge.....	36
II.3.2.Mecanisme de claquage avec phase gazeuse	36
II.3.2.1. Mécanisme thermique.....	37
II.3.2.2. Mécanisme cavitation.....	37
II.3.3.Mecanisme de claquage par pont.....	38
II.3.4.Evolution des décharges dans les diélectriques liquides.....	38
II.4.Les essais sur les diélectriques liquides	39
II.4.1.Essais diélectriques.....	39
II.4.1.1.Mesure du facteur de pertes diélectriques ($\text{tg}\delta$) et de la permittivité relative(ϵ_r).....	39
II.4.1.2.Mesure de la rigidité diélectrique	39
II.4.4.3. Mesure de la résistivité(ρ).....	39
II.4.2.Mesure physico-chimiques.....	39

II.4.2.1.Mesure de la viscosité cinématique (η).....	39
II.4.2.2.La couleur	40
II.4.2.3.L'acidite.....	40
II.4.3.Essais de vieillissement de l'huile.....	40
II.5.Huile de transformateur borac22.....	41
II.5.1.Proprietes de l'huile borac22.....	41
II.5.2.Influence de la température sur les propriétés de l'huile.....	43
II.5.2.1.Facteur de pertes diélectriques.....	43
II.5.3.Influence du vieillissement thermique sur les propriétés de l'huile.....	45
II.5.3.1.Facteur de pertes diélectriques	45
II.5.3.2.Permittivite relative ϵ_r	46
II.6. Conclusion.....	47

Chapitre .III. Réseaux élaborés et résultats

III.1.Introduction.....	48
III.2.Achitecture et fonctionnement des réseaux RBF.....	48
III.3.Codage de l'information.....	49
III.4.Phase d'apprentissage du réseau RBF avec ROM.....	51
III.5.Résultats a paramètre constant.....	53
III.6. Résultats a paramètres variables.....	54
III.6.1.Variations de temps et pas d'apprentissage.....	54
III.6.1.Variations de N_c et taux.....	58
III.7.Comparaison entre les résultats obtenus.....	61
III.8.Intrprtations.....	61
III.9.Conclusion.....	61

Conclusion générale.....	62
--------------------------	----

REMERCIEMENTS

Nous remercions Dieu de nous avoir donné la force physique et morale pour accomplir ce travail.

*Nous remercions chaleureusement nos encadrateurs **M^{rs}** : **G.YOUCEF** pour son suivi, ses précieux conseils et ses nombreuses critiques constructives durant toute l'année.*

Nous remercions également tous les membres du jury, pour avoir bien voulu évaluer notre travail.

Et finalement, nous adressons nos remerciements les plus profonds et les plus distingués à tous ceux qui ont contribué de proche ou de loin à ce modeste travail.

INTRODUCTION GENERALE

Depuis les années 80 les chercheurs dans le domaine de génie électrique ont été obligé de s'orienter vers les réseaux de neurones et s'adapter avec eux, surtout quant il s'agit d'identifier un système dont la dynamique est inconnue, de filtrer de signaux brutes ou de reconnaître les forme et les parole, etc....

Durant les année 90 des travaux on été communiqués sur l'utilisation des réseaux de neurones artificiels [1]

Notre étude fait suite à d'autres travaux concernant l'utilisation des RNA dans le vieillissement thermique de l'huile de transformateur [1]. Ce travail a permis d'élaborer de programme de calcul utilisant différentes techniques d'apprentissage. Le but est de réduire le temps d'expérimentation dans le cas de l'étude du vieillissement d'huile.

L'application des RNA au vieillissement thermique de l'huile de transformateur BORAK22 utilisée par SONELGAZ avec une durée de vieillissement de 5000h. Les résultats expérimentaux que nous avons utilisés, ont été réalisés par le Laboratoire de HT de l'ENP par d'autres chercheurs.

Ce travail est organisé dans trois chapitres :

Dans un premier chapitre, nous présentons les généralités concernant les réseaux de neurones artificiels.

Le deuxième chapitre présentes quelques caractéristiques d'une huile des transformateurs.

Le troisième chapitre, nous détaillons les architectures des réseaux que nous avons étudiés, leurs organigrammes et un exemple d'application.

Enfin, dans ce chapitre nous présentons les différents résultats que nous obtenons ainsi que leurs interprétations.

Finalement, nous terminons ce travail par une conclusion générale.

I.1. INTRODUCTION

L'objectif principal de la recherche sur les réseaux de neurones artificiels était d'accroître les connaissances sur le mécanisme cérébral via l'élaboration de systèmes artificiels capables de produire des calculs complexes (voire intelligents), similaires à ceux qu'effectue le cerveau humain. Ils sont donc à l'origine.

Dans les réseaux de neurones artificiels de nombreux processeurs appelés cellules ou unités, capables de réaliser des calculs élémentaires, sont structurés en couches successives capables d'échanger des informations au moyen de connexions qui les relient. On dit de ces unités qu'elles miment les neurones biologiques.

On commence ce chapitre par définir d'abord le neurone biologique ensuite le neurone artificiel. Puis on exposera les différents modes d'apprentissage des réseaux de neurones.

I.2. LES NEURONES

Les Réseaux de Neurones Artificiels sont fondés sur une représentation schématique d'un neurone biologique. Les notions de mémorisation, d'apprentissage et de généralisation sont communes aux deux. La réponse du réseau de neurone à un signal d'entrée dépend de paramètres appelés poids synaptiques (mémoire). Une loi d'apprentissage modifie progressivement ces paramètres pour réduire un paramètre d'erreur. Si les exemples présentés forment un échantillon représentatif de l'ensemble des entrées, le réseau améliore son comportement pour n'importe quel signal (généralisation). L'élément fonctionnel du réseau de neurones artificiels est le neurone formel.

I.2.1. LE NEURONE BIOLOGIQUE

Les cellules nerveuses, appelées neurones, sont les éléments de base du système nerveux central. Le cerveau se compose d'environ 10^{12} neurones (mille milliards), avec 1000 à 10000 synapses (connexions) par neurone.

Le neurone est une cellule composée d'un corps cellulaire et d'un noyau. le corps cellulaire se ramifie pour former ce que l'on nomme les dendrites. Celles-ci sont parfois si nombreuses que l'on parle alors de chevelure dendritique ou d'arborisation dendritique. C'est par les dendrites que l'information est acheminée de l'extérieur vers le soma, corps du neurone, l'information traitée par le neurone chemine ensuite le long de l'axone (unique) pour être transmise aux autres neurones. La transmission entre deux neurones n'est pas directe. en fait, il existe un espace intercellulaire de quelques dizaines d'angströms (10^{-9} m) entre l'axone du neurone afférent et les dendrites du neurone efférent, la jonction entre deux neurones est appelée la synapse.

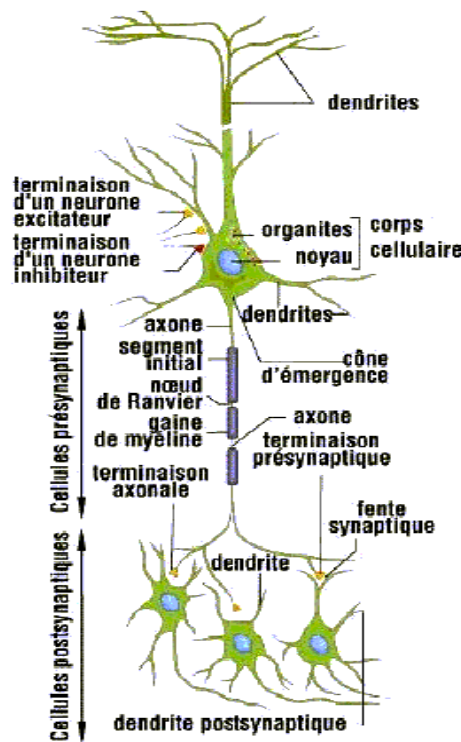


Fig. I.1. Vue d'un neurone biologique

I.2.2. NEURONE ARTIFICIEL

I.2.2.1. DEFINITION

Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique de réseaux est évidemment un réseau.

I.2.2.2. L'ETAT DES NEURONES

Un neurone artificiel est un élément qui possède un état interne. Il reçoit des signaux qui lui permettent éventuellement, de changer d'état. Nous noterons S l'ensemble des états possibles d'un neurone. S pourra être par exemple $\{0, 1\}$ où 0 sera interprété comme l'état inactif et 1 l'état actif. S pourra également prendre un nombre plus grand de valeurs $\{0, 1, \dots, P\}$ pour une image en $P + 1$ niveaux de gris ou même, par extension, un continuum de valeurs $[-1, 1]$ ou $\mathbb{R}$ tout entier. Dans une application en télédétection les valeurs correspondant aux signaux écho-radar peuvent être représentées à l'aide de neurones à valeurs continues. L'état

d'un neurone peut alors être défini dans l'intervalle $S = [-1, 1]$, où -1 représente la valeur minimum du signal, et 1 le maximum.

L'état d'un neurone est fonction des états des neurones auxquels il est relié. Pour calculer l'état d'un neurone il faut donc considérer les connexions entre ce neurone et d'autres neurones. Nous définirons par la suite les connexions entre neurones et leur poids.

I.2.2.3. LES CONNEXIONS ENTRE NEURONES

Architecture est le terme le plus général pour désigner la façon dont sont disposés et connectés les différents neurones qui composent un réseau. On parle également de topologie (terme emprunté de la théorie des graphes). Au niveau des neurones on parle plutôt de voisinage. Ce terme fait allusion à la façon dont un neurone est connecté à d'autres neurones. Il est donc en rapport direct avec l'architecture du réseau. Voyons de plus près la signification du mot voisinage dans une architecture de réseaux de neurones.

I.2.2.3.1. LE VOISINAGE

Le voisinage d'un neurone est l'ensemble des neurones connectés à ce neurone. On parle de voisinage d'ordre n pour un neurone i , s'il y a n neurones connectés à ce neurone. Les connexions entre neurones ont souvent un sens.

I.2.2.3.2. LES CONNEXIONS

Une connexion est un lien établi explicitement entre deux neurones. Les connexions sont aussi appelées synapses, en analogie avec le nom des connecteurs des neurones réels.

Une connexion entre deux neurones a une valeur numérique associée appelé poids de connexion.

I.2.2.3.3. LES POIDS DES CONNEXIONS

Le poids de connexion w_{ij} entre deux neurones j et i peut prendre des valeurs discrètes dans $\mathbb{Z}$ ou bien continues dans $\mathbb{R}$. L'information qui traverse la connexion sera affectée par la valeur du poids correspond. Une connexion avec un poids $w_{ij} = 0$ est équivalente à l'absence de connexion.

On définit une matrice des poids de connexions $\mathbf{W}$ où les lignes et les colonnes correspondent aux neurones et chaque valeur w_{ij} représente le poids de la connexion entre la cellule j et la cellule i du réseau.

I.2.2.4. DESCRIPTION D'UN NEURONE FORMEL

Un neurone formel ou artificiel est un opérateur mathématique très simple. Un neurone possède des entrées qui peuvent être les sorties d'autres neurones, ou des entrées de signaux extérieures, et une sortie. La valeur de la sortie résulte du calcul de la somme des entrées,

pondérées par des coefficients (dits poids de connexions ou poids synaptiques) et du calcul d'une fonction non linéaire (dite fonction d'activation) de cette somme pondérée. L'état du neurone, appelé aussi activité, est défini comme la somme pondérée de ses entrées. Son schéma de fonctionnement est donné en Figure I-2. L'information est ainsi transmise de manière unidirectionnelle. Un neurone se caractérise par trois concepts : son état, ses connexions avec d'autres neurones et sa fonction d'activation.

Nous utiliserons par la suite les notations suivantes.

- S_i : l'état à la sortie du neurone i .
- e_j ($j=1,2,\dots,p$): les entrée de neurone
- f_i : la fonction d'activation associée au neurone i .
- w_{ij} : le poids de la connexion entre les neurones j et i .
- w_{i0} : le poids de la connexion entre le neurone biais (+1) et les neurones i .

Ainsi, le neurone i recevant les informations de p neurones effectue l'opération suivante :

$$S_i = f \left(\sum_{j=1}^p w_{ij} e_j - w_{i0} \right) \quad (\text{I.1})$$

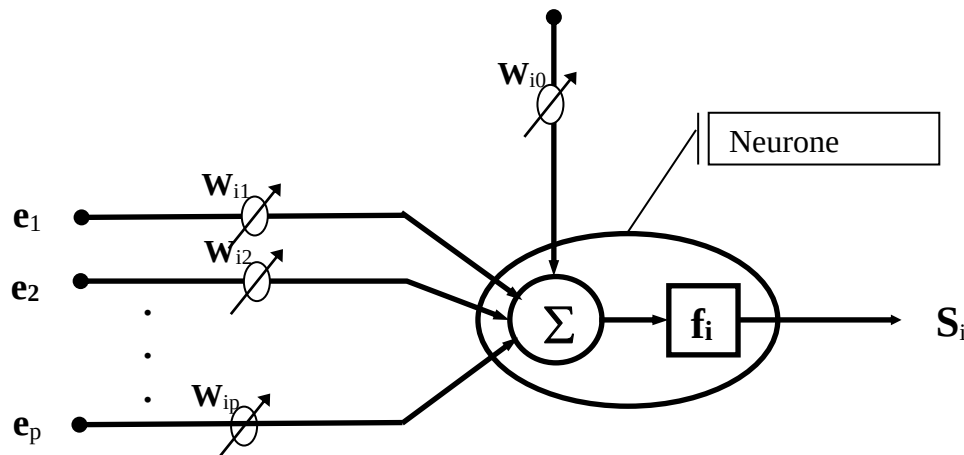


Fig. I.2. Schéma de fonctionnement d'un neurone *formel*.

I.3. MODELE MATHEMATIQUE GENERAL DU NEURONE

I.3.1. FONCTION DISCRIMINANT (FONCTION DE BASE)

Elle définit l'activité du neurone. Dans le cas de Perceptron, c'est la fonction linéaire qui est utilisée. Cette fonction est depuis longtemps la plus usitée. C'est le cas pour les réseaux multicouches [3] et de manière générale les réseaux LBF (Linear Basis Function).

D'autres formes de fonctions discriminants non linéaires ont été élaborées ces dernières années afin d'améliorer les performances des réseaux.

Nous citons ci dessous les fonctions de base les plus utilisées.

I.3.1.1. FONCTION DE BASE LINEAIRE LBF

Cette fonction fait la somme pondérée des entrées vers le neurone. Sa forme est en générale définie par :

$$\Phi(z, w) = z^T \cdot w \quad (I.2)$$

Où W représente la matrice des poids et z les entrées qui viennent de l'extérieur ou de la part des autres neurones vers le neurone en question. En plus de son efficacité, cette fonction est plus simple à implémenter.

I.3.1.2. FONCTION DE BASE RADIALE RBF (RADIAL BASIS FUNCTION)

La forme de cette fonction est définie par :

$$\Phi(x, w) = -\frac{\|x - w\|^2}{2} \quad (I.3)$$

Où les poids W sont représentés comme étant les centroïdes de chaque classe de l'espace d'entrée. Le discriminant calcule le rayon entre les entrées x et le centre de chaque classe.

I.3.1.3. FONCTION DE BASE ELLIPTIQUE EBF (ELLIPTIC BASIS FUNCTION)

La forme de ce discriminant est une généralisation de la fonction de base radiale

$$\Phi(x, w) = \sum_{k=1}^n \alpha_k (x - w)^2 + \theta \quad (I.4)$$

Celle ci a une forme modulée par des paramètres α_k et dotée d'un Biais θ . Ces paramètres font qu'elle ne soit pas symétrique

I.3.2. FONCTION DYNAMIQUE LINEAIRE

Ce bloc détermine la dynamique du réseau. Cette fonction est régie par l'équation différentielle de premier ordre suivante

$$\alpha_0 \dot{u}_i(t) + \alpha_1 u_i(t) = v_i \quad (I.5)$$

Où u_i représente l'activité du $i^{\text{ème}}$ neurone et v_i l'entrée du système dynamique décrit par cette équation.

I.3.3. FONCTION NON DYNAMIQUE NON LINEAIRE

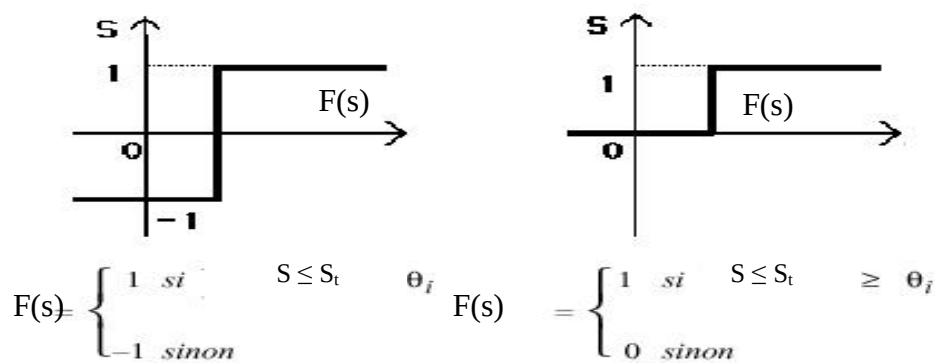
Elle a pour objectif de rendre l'activité du neurone bornée, la fonction d'activation peut être une fonction de seuil (Simple Threshold Function), sigmoïde (Sigmoidal), tangente hyperbolique (Hyperbolic Tangent) ou fonction de base Radiale [4].

nom de fonction	La forme mathématique
Linéaire	$f(s) = s$
Seuil	$f(s) = \begin{cases} +1, & \text{si } s > s_t \\ -1, & \text{si non} \end{cases}$
Sigmoïde	$f(s) = a \frac{e^{ks} - 1}{e^{ks} + 1}$
Tangente hyperbolique	$f(s) = \frac{1 - e^{-2s}}{1 + e^{-2s}}$
Fonction de base radiale	$f(s) = e^{-s^2 / \beta^2}$

Tableau 1- les fonctions d'activation.

La fonction d'activation doit être continue, dérivable et monotone.

Il existe une autre fonction d'activation qui n'est pas monotone, s'est la fonction Gaussienne; elle se distingue des autres, par sa linéarité significative et son comportement impulsionnelle. La caractéristique de comportement local rend cette fonction très puissante en approximation de fonction et en classification.

**Fig. I. 3.** Fonction à seuil

En générale, toutes les fonctions citées ont un fondement biologique [5] par exemple les neurone du système visuel sont connus ayant un comportement Gaussien. [6]. La recherche de L'efficacité a permis à d'autres fonctions de voir le jour ces dernières années comme les fonctions logarithmique, cubiques ou exponentielles logarithmiques.

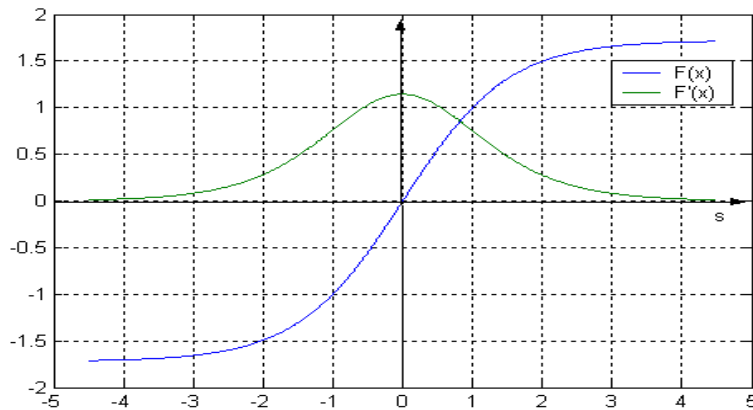


FIG. I.4. Fonction sigmoïde f et dérivée première f' . f est une fonction Quasi-linéaire, pour $a=1.72$ et $k=1.33$ (voir tableau 1)

I.4. CLASSIFICATION D RESEAUX DE NEURONES

La figure suivante (**Fig.I.5.**) résume les trois différentes possibilités suivant lesquelles on peut classifier les réseaux de neurones. Dans ce qui suit, nous utilisant ces différentes classifications pour faire l'étude de plusieurs points liés aux réseaux de neurones.

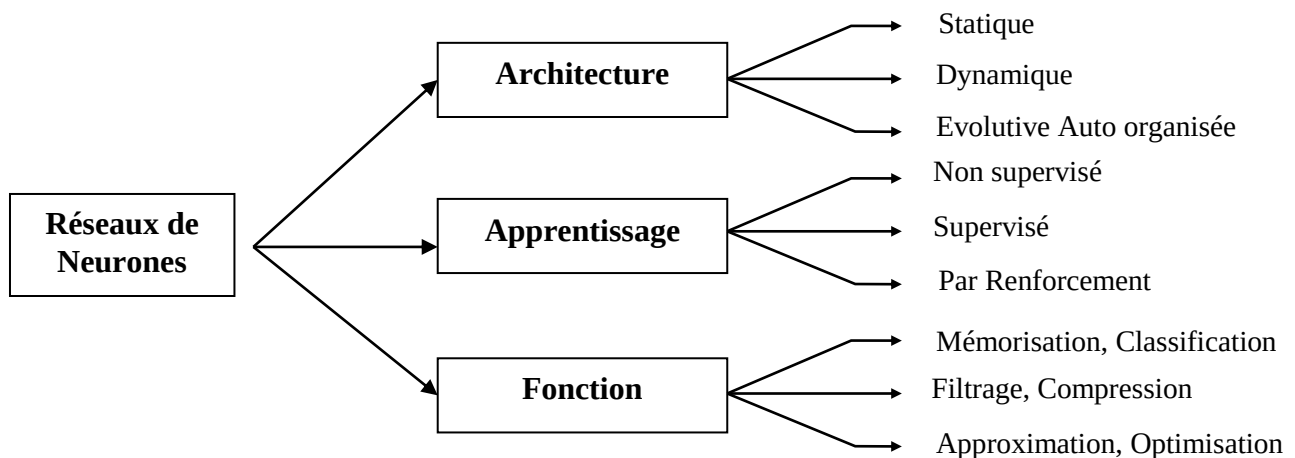


Fig.I.5. Les différentes possibilités de classification des réseaux de neurones

1.5 les TYPES D'APPRENTISSAGE

Les techniques d'apprentissage des réseaux de neurones se repartissent en deux grandes familles:

- L'apprentissage est dit supervisé lorsque les exemples sont constitués de couples de valeurs du type : (valeur d'entrée, valeur de sortie désirée). Tout le problème de l'apprentissage supervisé consiste, étant donné un ensemble d'apprentissage E de N couples (entrée-sortie désirée) $(x_i, y_i)_{i=1,2,\dots,n}$, à déterminer le vecteur des poids W d'un réseau F_w capable de mettre ces informations en correspondance, c'est à dire un réseau tel que

$$F_w(x_i) = y_i, \quad i = 1, 2, \dots, n. \quad (I.6)$$

L'apprentissage est qualifié de non-supervisé lorsque seules les valeurs d'entrée sont disponibles. Dans ce cas, les exemples présentés à l'entrée provoquent une autoadaptation du réseau afin de produire des valeurs de sortie qui soient proches en réponse à des valeurs d'entrée similaires [8].

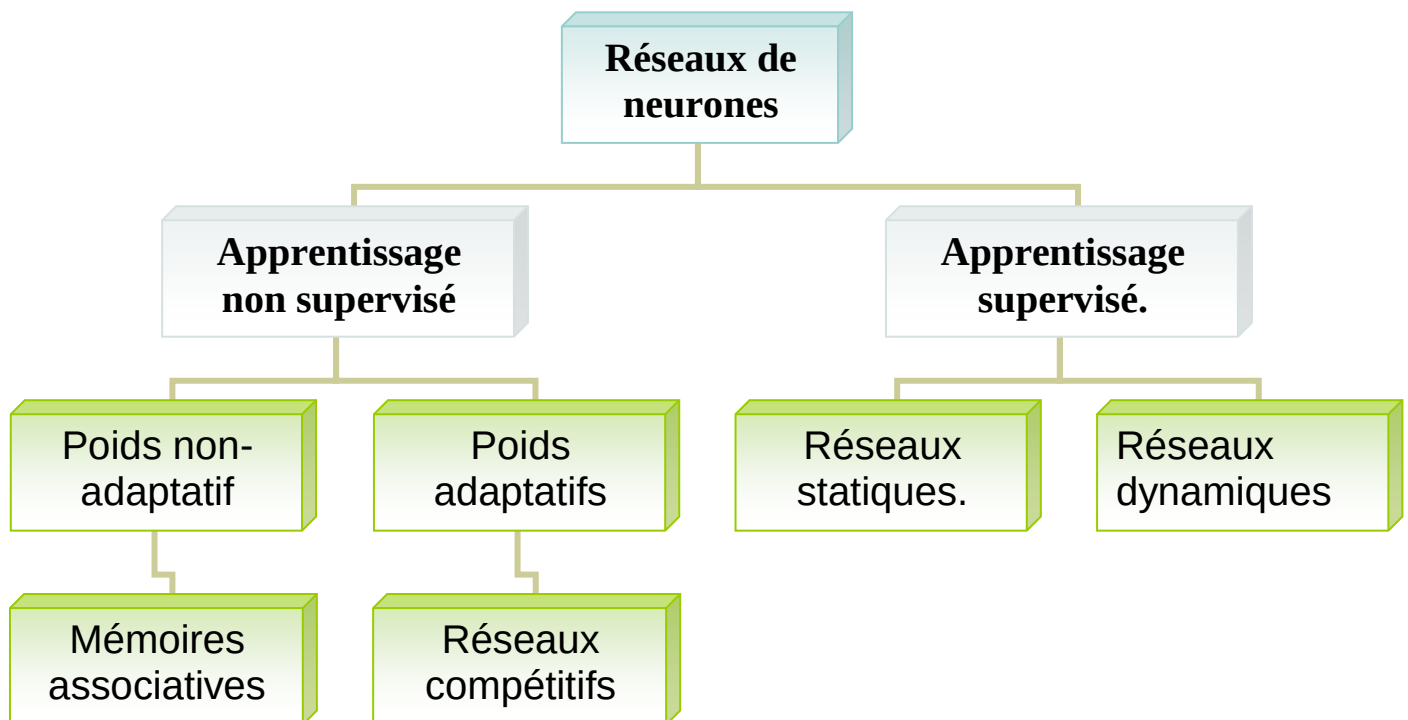


Fig. I.6. Classification des réseaux de neurones suivant leur apprentissage

I.6. RESEAUX DE NEURONES A APPRENTISSAGE NON-SUPERVISE

Ces réseaux traitent des échantillons en entrée sans avoir d'informations sur les sorties que ceux ci doivent générer. L'objectif de ses réseaux n'est pas de faire une approximation. Le rôle de ces réseaux est basé sur la mémorisation et la classification de leurs entrées.

I.6.1. MEMOIRE ASSOCIATIVE

La mémoire associative est une des premières applications issues des réseaux de neurones adaptatifs. Il s'agit de réaliser un dispositif capable de mémoriser des informations que l'on peut ensuite retrouver, non par leur adresse comme dans une mémoire classique, mais en fournissant des données même incomplètes ou bruitées relatives aux informations stockées [7].

Les premiers modèles de mémoires associatives sont ceux de Widrow et de Rosenblatt, suivis par les modèles de Kohonen et Anderson. [1]

La mise en mémoire des informations à l'intérieur de ces réseaux se fait grâce à un apprentissage non-supervisé. Pendant cette étape, appelée **Encoding Step**, le réseau calcule les poids synaptiques W , qui représente un codage de l'information à stocker. Grâce à la présentation interne déduite de l'apprentissage, les neurones calculent leurs activations. A la sortie, il doit se rappeler de se qu'il appris. Cette étape est appelée **Recal Step**. [8]

On distingue deux types de mémoires associatives suivant le mode de fonctionnement :

a) Mémoires auto-associatives :

Elles sont destinées à stocker des vecteurs prototypes

$$P^k = p_1^k, p_2^k, \dots, p_n^k. \quad (I.7)$$

Tel que $k = 1, 2, \dots, M$ représente les différentes classes.

b) Mémoires hétéro-associatives :

Elles stockent les couples Prototypes (X^k, Y^k) , où :

$$Y^k = y_1^k, y_2^k, \dots, y_n^k \quad \text{avec } k = 1, 2, \dots, M \quad (I.8)$$

A chaque vecteur X^k en entrée, cette mémoire associe le vecteur Y^k en sortie et vice versa.

I.6.1.1. RESEAU DE HOPFIELD

Ce modèle a été développé en 1982 selon une analogie formelle avec des réseaux de particules orientées en interaction. Ce réseau est une mémoire auto-associative constituée par une couche de neurones entièrement interconnectés (**Fig.I. 7.**)

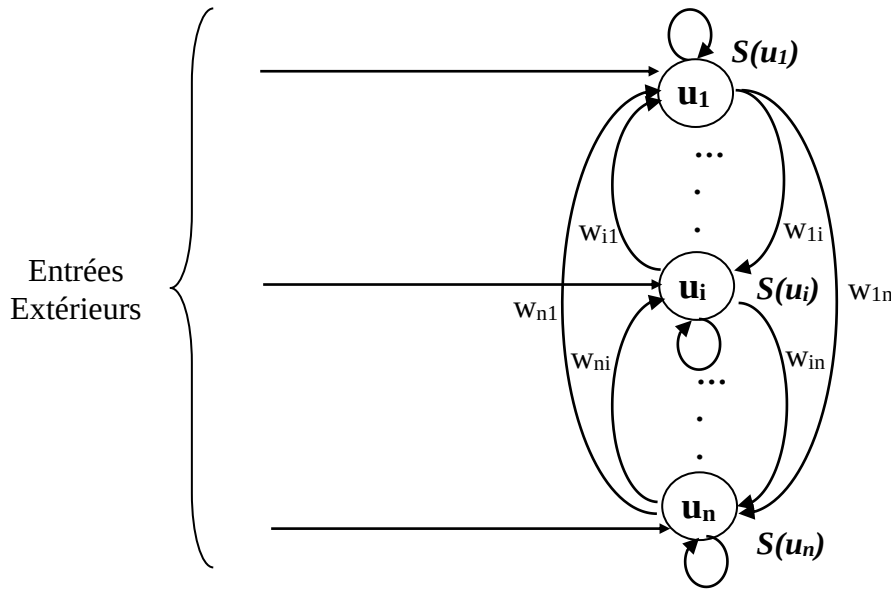


Fig.I.7. Architecture du réseau de Hopfield

L'originalité du réseau de Hopfield réside notamment dans son modèle électrique très connu, ce modèle traduit le comportement dynamique intégrateur des neurones, celui-ci est défini par le système d'équation suivant :

$$\dot{u}_i = \sum_{j=1}^n w_{ij} S_j(u_j) - b_i u_i + I_i \quad i = 1, 2, \dots, n \quad (\text{I.9})$$

Où I_i représentent les entrées extérieures, w_{ij} des paramètres définissant les forces des liens synaptiques représentées par les conductances de ces connections, b_i représente la conductance équivalente à l'entrée du neurone, et $S_j(u_j)$ représente la sortie du neurone [7].

La sortie du neurone est binaire comme son entrée, donc la fonction de sortie est une fonction seuil définie par l'équation (I. 9).

I.6.1.2. MEMOIRES ASSOCIATIVES BIDIRECTIONNELLES (RAM)

Elles représentent une extension du réseau de Hopfield vers une mémoire hétéro - associative, ces réseaux sont composés de deux couches de neurones, l'information à l'intérieur de ce réseau peut circuler dans les deux sens (de la première couche vers la deuxième couche et vice versa).

I.6.1.3. MEMOIRES ASSOCIATIVES LINEAIRES OPTIMISEES (OLAM)

Le réseau de Hopfield ne garantit un bon fonctionnement que si les exemples sont mutuellement orthogonaux, ce qui n'est pas toujours vérifié en pratique.

Les mémoires associatives optimisées constituent une solution pour les problèmes d'orthogonalité, ces réseaux utilisent la règle de projection pour leur apprentissage, mais leur architecture est la même que celle de Hopfield [7].

En injectant une entrée égale à un des prototypes stockés ou proche de lui, on doit retrouver ce prototype en sortie. Ceci revient à résoudre l'équation suivante :

$$W X^k = X^k \quad \text{avec } k = 1, 2, \dots, M \quad (\text{I. 10})$$

Une solution de cette équation est :

$$W = X \cdot X^1 \quad (\text{I.11})$$

Où X est une matrice regroupant tout les vecteurs prototypes à mémoriser et X^1 représente le pseudo-inverse de cette matrice.

I.6.2. RESEAUX COMPETITIFS

Ces réseaux s'adaptent au fur et à mesure de leur utilisation, contrairement aux réseaux que nous avons étudié jusqu'à maintenant, qui sont à poids fixes. Ces réseaux contiennent, en plus de la couche d'entrée, une couche compétitive, où les neurones entrent en compétition entre eux, afin de déterminer lequel est le plus représentatif de la classe de l'espace d'entrée.

Nous présentons maintenant les modèles des principaux réseaux compétitifs.

I.6.2.1. COUCHE DE GROSSBERG

Elle est constituée de neurones entièrement interconnectés, sa dynamique est régie par l'équation différentielle suivante :

$$\dot{u}_i = -\mu \cdot u_i + \delta \sum_{j=1}^n S(u_j) w_{ij} + I_i \quad i = 1, 2, \dots, n \quad (\text{I.12})$$

Où :

u_i : représente l'activité du $i^{\text{ème}}$ neurone.

I_i : l'entrée extérieure vers ce neurone.

$S(u_i)$: représente la sortie du neurone.

w_{ij} : les poids synaptiques.

δ et μ : constantes positives.

Lors de la présentation d'une entrée X^k au réseau, la dynamique de l'équation (I.12) fait que le neurone de la classe à laquelle cet exemple est le plus proche est activé au maximum. A l'équilibre c'est le seul qui reste activé et le seul à se voir réadapter ses poids. L'apprentissage se fait de la manière suivante :

$$\dot{w}_{ij} = S(u_i)[- \alpha w_{ij} + \beta S(u_i)] \quad (I.13)$$

Où :

α : un paramètre contrôlant le régime transitoire.

β : le taux d'apprentissage.

Le défaut de ce modèle est qu'il traite de la même manière que le signal. En effet l'équation (I.13), qui est linéaire, ne permet aucun filtrage du bruit. Ceci conduit le modèle rapidement vers la saturation.

I.6.2.2. ADAPTIVE RESONANCE THEORY ART

Le fonctionnement de ce réseau constitue une analogie avec la résonance physique. A partir d'un faible signal, l'information fait des 'aller-retour' entre la couche d'entrée et celle de sortie, constituant des vibrations, 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 [7].

Ce réseau est constitué de deux couches ; une couche d'entrée ordinaire et une couche de sortie appelée Maxnet, dont les neurones sont interconnectés (**Fig.1.8**)

Chaque neurone de la couche de sortie représente une classe. Lorsque une entrée x est présentée au réseau, celle-ci est transmise vers la couche supérieure en passant par des liens synaptiques allant dans le sens direct appelés Bottom-Up « bas-haut ». Chaque neurone du Maxnet calcule son activation. Le neurone ayant ses poids les plus corrélés avec l'exemple en entrée, présente le plus grand résultat en sortie, celui-ci sera candidat pour être désigné comme étant la classe à laquelle cet exemple appartiendrait. Avant de prendre cette décision, l'exemple x est envoyé vers la couche d'entrée, en passant par les liens synaptiques qui vont dans le sens inverse appelés Top-Down « haut-bas », là, un test dit de vigilance, est appliqué afin d'estimer à quel point cet exemple ressemble à ceux appartenant à la classe gagnant. Si celui-ci réussit le test, l'exemple est accepté. Sinon le neurone gagnant est disqualifié et une autre classe est recherchée [7].

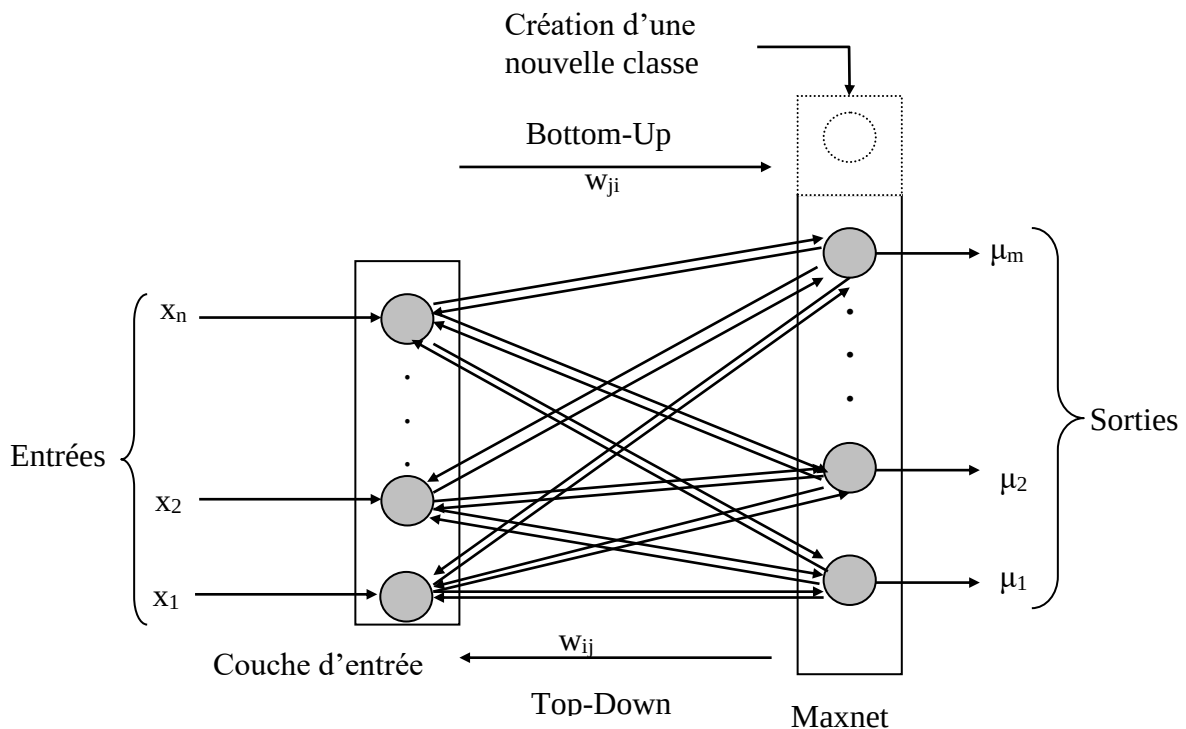


Fig.I.8. Architecture du ART

I.7. APPRENTISSAGE SUPERVISE

Nous considérerons par la suite un réseau comme un associer défini par son architecture et un ensemble de paramètres à ajuster. Pour simplifier la présentation, nous considérerons que les seuls paramètres à ajuster sont les poids des connexions.

Il existe deux approches qui peuvent être faites pour ce type d'apprentissage. La première est basée seulement sur la validité de la décision du réseau. Ce type de réseau est appelé (DBNN) réseaux basés sur les décision, la deuxième catégorie est basée sur l'optimisation d'un critère.

I.7.1. REGLES DU PERCEPTRON ET D'ADALINE, MADALINE

Les règles d'apprentissages utilisées dans le Perceptron de Roenblatt et ADALINE (ADAPtive LInear NEuron) de Widrow et Hoff reposent toutes deux sur la minimisation d'un critère d'erreur quadratique instantanée.

Les architectures des réseaux proposés dans les deux cas son très voisines (**Fig.I.9**).

A savoir une couche d'entrée connectée à une où plusieurs sorties, dans le cas de plusieurs sorties, il est alors appelé MADALINE (Multiple-Adaline).

La caractéristique du neurone dans le Perceptron est une fonction Echelon :

$F(p) = \text{Ech}(p - \theta)$, où θ est un seuil.

L'apprentissage se fait avec superviseur, c'est-à-dire que l'on présente un vecteur à l'entrée du réseau et en même temps on fournit la sortie désirée : dans le cas d'Adaline, il s'agit du potentiel désiré $P_{\text{désiré}}$, dans le cas du Perceptron, il s'agit de la classe désirée $y_{\text{désirée}}$.

Pour un seul « neurone », à chaque pas d'apprentissage, l'erreur quadratique E est minimale est donc :

$E = (P - P_{\text{désiré}})^2$, erreur analogique, dans le cas d'Adaline,

$E = (y - y_{\text{désirée}})^2$, erreur binaire, dans le cas du Perceptron.

Le principe de minimisation d'erreur n'est autre qu'un cas particulier de l'application de la méthode de descente de gradient.

$$\frac{\partial E}{\partial w_k} = 2(P - P_{\text{désiré}}) \frac{\partial P}{\partial w_k} \quad (\text{I.14})$$

Or $P = \sum_j w_j x_j$,

Donc $\frac{\partial P}{\partial w_k} = x_k$,

Et finalement

$$\frac{\partial E}{\partial w_k} = 2(P - P_{\text{désiré}}) x_k.$$

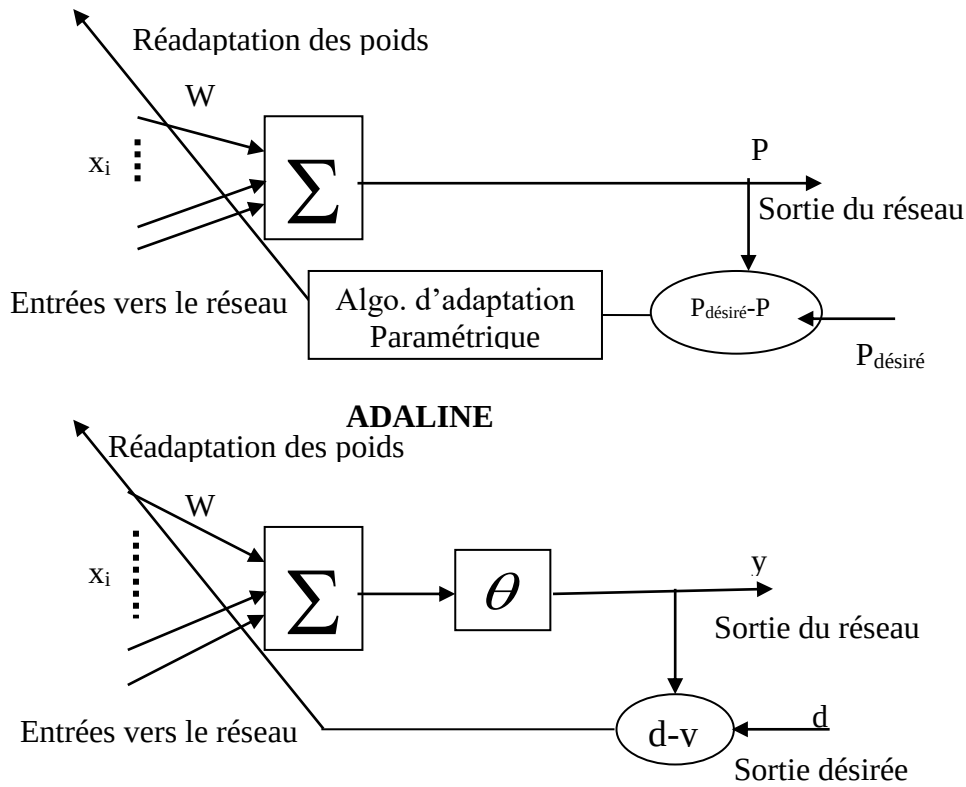


Fig. I.9. Neurones D'Adaline et Perceptron

D'où :

$$\Delta w_k = -\alpha(P - P_{\text{désiré}})x_k, \quad (\text{I.15})$$

Où α est un gain d'adaptation positif.

Le même type d'équation que (I.15) est utilisé dans l'algorithme du Perceptron, les poids sont modifiés selon la règle suivante :

$$\Delta w_k = -\alpha(y - y_{\text{désirée}})x_k \quad (\text{I.16})$$

Cet algorithme est très simple, mais il présente l'inconvénient des algorithmes de type gradient stochastique en ce qui concerne la convergence. Par ailleurs, il est évident que minimiser l'erreur quadratique instantanée n'est pas identique à minimiser l'erreur quadratique moyenne calculée sur toute la base d'apprentissage, et que ce dernier critère est le plus performant [7].

I.7.1.1. FONCTION DE COUT ET METHODE DE GRADIENT

Les techniques de gradient permettent l'estimation des paramètres W par l'optimisation d'une fonction de coût qui caractérise les performances du réseau. Cette fonction de coût sera calculée entre les vecteurs désirés et les vecteurs calculés par le réseau.

Soit l'ensemble des vecteurs d'entrée $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$, définis dans R_n , les vecteurs réponses $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$, définis dans R_p et $\mathbf{W}$ la matrice de poids de connexions définie dans R_{pn} . Alors, on a :

$$\mathbf{x}_k = [x_1^k, x_2^k, \dots, x_n^k]^T, \mathbf{y}_k = [y_1^k, y_2^k, \dots, y_p^k]^T \text{ et}$$

$$\mathbf{W} = [w_1, w_2, \dots, w_p]^T = [\mathbf{w}_{ij}]_{p \times n}.$$

Où $\mathbf{w}_i = [w_{1i}, w_{2i}, \dots, w_{pi}]^T$ est le vecteur colonne composé par les poids des connexions entre les n cellules d'entrée et l'automate i en sortie. Définissons la fonction de coût suivante :

$$\mathbf{C}(\mathbf{W}) = \frac{1}{2} \sum_{k=1}^m \|\mathbf{w}^T \cdot \mathbf{x}_k - \mathbf{y}_k\|^2 \quad (\text{I.17})$$

Où $\|\cdot\|^2$ est la norme quadratique. Le problème est alors de trouver un $\mathbf{W}^*$ Pour $\mathbf{W}$ tel qu'il minimise $\mathbf{C}(\mathbf{W})$. Une condition nécessaire est la suivante :

$$\nabla \mathbf{C}(\mathbf{w})_{\mathbf{w}=\mathbf{w}^*} = \left\{ \frac{\partial \mathbf{C}(\mathbf{w})^T}{\partial \mathbf{w}_{ij}} \right\}_{p \times n} \Big|_{\mathbf{w}=\mathbf{w}^*} = 0 \quad (\text{I.18})$$

Les techniques de gradient sont des méthodes d'optimisation itératives qui permettent de trouver des solutions vérifiant la condition (3.8). La version la plus simple de ces techniques est celle dite du *gradient de la plus grande pente* :

Définir la configuration initiale $\mathbf{W} = \mathbf{W}_0$

$\forall t_k, k = 1, 2, \dots$

$$\mathbf{W}_{k+1}^T = \mathbf{W}_k^T - \varepsilon \nabla \mathbf{C}(\mathbf{W}_k) \quad (\text{I.19})$$

Où ε est un scalaire positif, souvent dépendant de k –auquel cas on écrit ε^k –et qui est appelé le pas d'itération. Ce coefficient est incorporé dans le terme $\Delta \mathbf{w}_{ij}^k$.

Dans le cas quadratique, en remplaçant (I.17) dans (I.19) on obtient l'expression :

$$\mathbf{W}_{k+1}^T = \mathbf{W}_k^T - \varepsilon^k \sum_{k=1}^m (\mathbf{W}_k^T \mathbf{x}_k - \mathbf{y}_k) \cdot \mathbf{x}_k^T \quad (\text{I.20})$$

Dans (I.20) les changements de chaque w_{ij} sont fonction des toutes les données $(\mathbf{x}_k, \mathbf{y}_k)$, $k = 1, 2, \dots, m$, m étant le nombre d'exemples.

I.7.1.2. LE GRADIENT STOCHASTIQUE : LA REGLE DE WIDROW-HOFF

Ces techniques de gradient ne sont pas adaptées au cas des réseaux connexionnistes où l'on désire mettre en oeuvre des systèmes adaptatifs capables de modifier leurs paramètres en continu en fonction d'un environnement qui peut changer. Widrow et Hoff ont proposé en 1960 la règle adaptative qui porte leur nom et qui est également connue sous le nom de règle LMSE [7]. Il s'agit tout simplement d'une règle de gradient suivant la plus grande pente où les paramètres w sont modifiés non pas en tenant compte à chaque itération de l'erreur globale réalisée sur un ensemble de formes, mais après chaque présentation de forme on modifie le vecteur de paramètres. La formulation mathématique de la règle est la suivante :

$$\mathbf{W}_{k+1}^T = \mathbf{W}_k^T - \varepsilon^k (\mathbf{W}_k^T \cdot \mathbf{x}_k - y_k) \cdot \mathbf{x}_k^T \quad (\text{I. 21})$$

Où $(\mathbf{x}_k, y_k)$ est le couple présenté à l'instant de temps t_k . Il a été démontré que, dans le cas d'une erreur quadratique, (I.21) converge avec une probabilité 1 vers $\mathbf{W}^*$, la solution de $\mathbf{W}$ que minimise la fonction de coût $C(\mathbf{W})$.

L'algorithme de l'Adaline est très général puisqu'il permet d'apprendre des associations quelconques. Quelque soit la nature du problème étudié, cet algorithme fournira une solution qui est la meilleure que l'on puisse avoir avec cette architecture au sens de l'erreur aux moindres carrés. Toutefois, l'architecture utilisée étant extrêmement simple, il est évident que la qualité de l'approximation trouvée ne sera pas toujours excellente.

L'algorithme de l'Adaline est extrêmement employé, c'est notamment l'algorithme de base en traitement adaptatif du signal. Toutefois, à cause des limitations évoquées plus haut, on a cherché des architectures et des algorithmes permettant de réaliser des tâches plus complexes ou avec une meilleure approximation que celle fournie par une simple Adaline. Ces recherches ont donné lieu à de très nombreuses publications dans les années 60-70. Toutefois, aucun algorithme général n'a été formulé permettant d'entraîner des réseaux plus complexes et d'usage général. Widrow par exemple a lui même proposé la notion de Madaline [2] où plusieurs Adalines sont utilisées simultanément ou séquentiellement pour résoudre des problèmes non solubles de façon exacte par une seule Adaline. Les Madalines, comme tous les algorithmes présentés à la même époque, nécessitent une importante connaissance a priori du problème qui permette de définir l'architecture à employer. Ce type de connaissance n'est bien sûr quasiment jamais disponible quand on s'attaque à un problème réel. Il ne s'agit donc pas d'une machine d'apprentissage général.

Il faut attendre le milieu des années 80 pour voir apparaître des machines beaucoup plus puissantes que celle des années 60 avec des procédures d'apprentissage générales et efficaces. Le modèle le plus célèbre qui est également celui que nous avons employé pour nos applications est celui des Perceptrons multicouches entraînés par l'algorithme de rétro-propagation *du gradient*. Toutefois, cet algorithme n'est qu'une généralisation à un type d'architecture plus complexe de celui de l'Adaline.

I.7.2. RESEAUX MULTICOUCHES STATIQUES

L'entrée dans ces réseaux est suivie par une succession de couches dites couches cachées, pour finalement aboutir à la sortie (**Fig.I.10**) ; ce passage par ces couches contribue à la richesse du traitement de l'information à l'intérieur du réseau afin d'obtenir le résultat désiré en sortie [7].

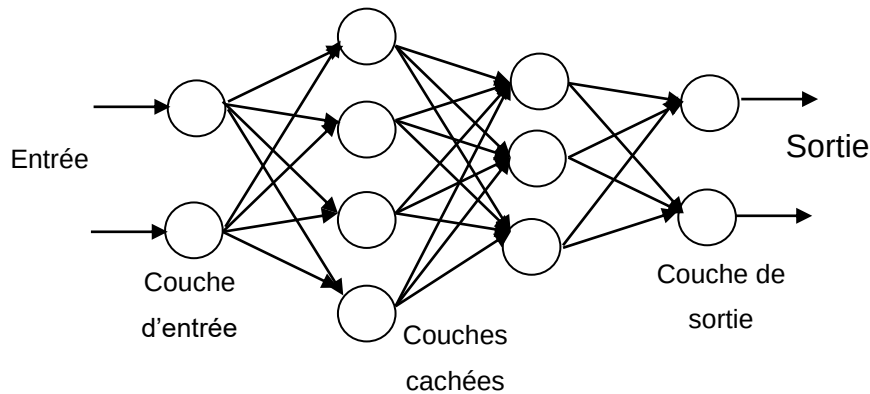


Fig.I.10. Réseau statique multicouche

I.7.3. RESEAUX A FONCTION DE BASE LINEAIRE LBF (LINEAR BASIS FUNCTION)

Le réseau LBF a un discriminant linéaire, il est régi par l'équation dynamique suivante :

$$u_i^l(t) = \sum_{j=1}^{N_{l-1}} w_{ij}^l(t) y_j^{l-1}(t) + \theta_i^l(t) \quad (I.22)$$

$$y_i^l(t) = f(u_i^l(t)) \quad (I.23)$$

Tel que :

l : le nombre de couches.

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

$y_i^l(t)$: la sortie du $i^{\text{ème}}$ neurone dans la $l^{\text{ème}}$ couche, en provenance du $j^{\text{ème}}$ neurone de la couche précédente.

$\theta_i^l(t)$: est un paramètre appelé Bias, il a pour objectif de donner de plus de souplesse à la fonction de sortie du neurone en approximation de fonctions.

La fonction d'activation f du neurone est monotone non décroissante. Elle a pour rôle de limiter l'activité du neurone. Pour les neurones de la couche d'entrée, cette fonction n'est autre que l'identité. Ainsi dans la couche d'entrée, chaque neurone fait passer une composante du vecteur entrée vers sa sortie sans aucun traitement, c'est à dire $y_i^1(t) = x_i(t)$ où $x_i(t)$ représente la $i^{\text{ème}}$ entrée du réseau à l'instant t .

I.7.3.1. LA RETRO-PROPAGATION DU GRADIENT

Les limitations des modèles neuronaux des années 60 comme l'Adaline ou le Perceptron [1], ont conduit les chercheurs à abandonner progressivement cette ligne de méthodes d'apprentissage au profit des approches symboliques de ce que l'on appelle aujourd'hui l'Intelligence Artificielle classique.

Dans les années 80, grâce aux travaux de Teuvo Kohonen et de J. Hopfield [7], cette voie a été remise au goût du jour et a suscité un intérêt croissant de la part de nombreux chercheurs issus de différentes disciplines.

C'est ainsi que fut proposé simultanément par plusieurs équipes travaillant indépendamment l'algorithme de la rétro-propagation du gradient. David E. Rumelhart, Geoffrey E. Hinton et Ronald J. Williams en 1986 [4], ont présenté la méthode, qu'ils ont appelé la Règle Delta Généralisée [4]. Presque simultanément, Yan le Cun [7] et David B. Parker [7] présentaient chacun séparément des algorithmes d'apprentissage similaires.

Des dérivations formelles de la méthode de rétro-propagation du gradient peuvent être trouvées dans [1].

I.7.3.2. L'ALGORITHME DE RETRO-PROPAGATION DU GRADIENT

Notons x_i l'état d'un neurone i , $x_i \in \mathbf{S}$ qui est l'ensemble des états possibles. l'état x_i du neurone i est calculé par :

$$x_i = f(A_i - \theta_i) \quad (\text{I.24})$$

Où, pour les MLP, f est une fonction sigmoïde, A_i l'activité du neurone i et θ_i le seuil associé au neurone i . L'activité A_i est donnée par :

$$A_i = \sum_j W_{ij} x_j \quad (\text{I.25})$$

Où j est indice des neurones “en amont” du neurone i , et w_{ij} est le poids de la connexion du neurone j au neurone i .

I.7.3.2.1. PROPAGATION DES ETATS

On considère encore une fois l'ensemble des vecteurs d'entrée $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$, définis dans $\mathbb{R}^n$ et les vecteurs réponses $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m$, définis dans $\mathbb{R}^p$ et $\mathbf{W}$ la matrice de poids de connexions cette fois-ci définie dans $\mathbb{R}^{nc \times nc}$, nc étant le nombre total de cellules. Alors,

$$\mathbf{x}_k = [x_1^k, x_2^k, \dots, x_n^k]^T, \mathbf{y}_k = [y_1^k, y_2^k, \dots, y_p^k]^T \text{ et}$$

$$\mathbf{W} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ w_1 & 0 & \dots & 0 & 0 \\ 0 & w_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & w_p & 0 \end{bmatrix} = \{ \mathbf{W}_{ij} \}_{nc \times nc}, \text{ où } w_i = \{ w_{kl} \}_{nc(i-1) \times nc(i)} \quad (\text{I.26})$$

Chaque sous-matrice $\mathbf{W}_i$ contient les poids de connexions entre la couche $i-1$ –avec $nc_{(i-1)}$ cellules– et la couche i –avec $nc_{(i)}$ cellules–.

A chaque instant k , un vecteur $\mathbf{X}_k$ est présenté dans la couche d'entrée du réseau. Les états des n cellules en entrée prennent les valeurs $(x_1^k, x_2^k, \dots, x_n^k)$

Ces états sont propagés selon les équations (I.24) et (I.25), vers des unités se trouvant dans les couches en aval de la couche d'entrée jusqu'à arriver à la couche de sortie.

Appelons $\mathbf{S}_k = (s_1^k, s_2^k, \dots, s_p^k)$ la réponse du réseau pour une entrée $\mathbf{X}_k$ donnée. $\mathbf{S}_k$ est le vecteur composé par les états de p cellules de la couche de sortie. Leurs états ont été calculés par les équations (I.24) et (I.25) en fonction des cellules j connectées à la couche de sortie

Couche :

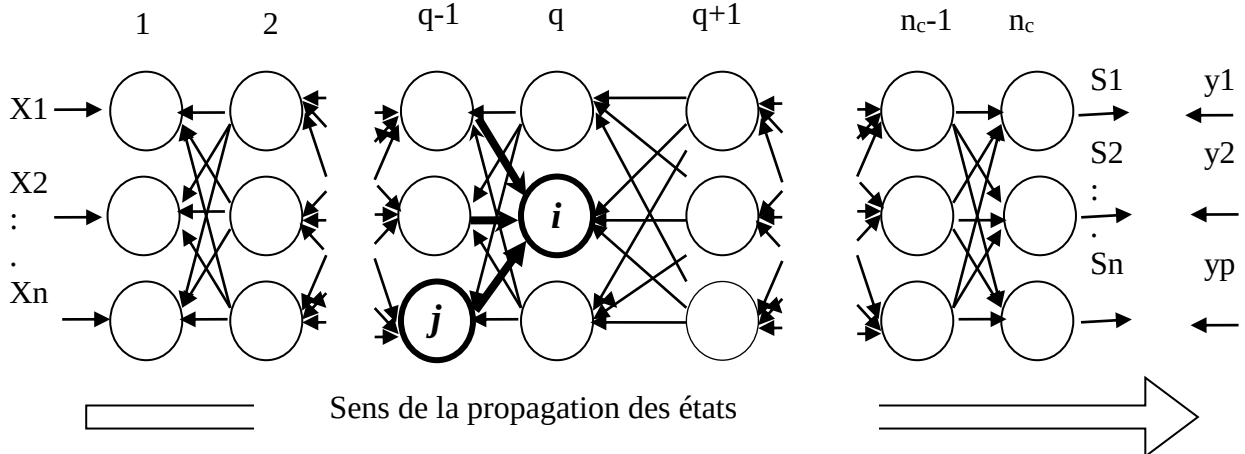


Fig.I.11. Propagation des états dans un réseau multicouche.

La propagation des états se fait “en aval” dès la couche d’entrée vers la couche de sortie. L’état de la cellule i dans la couche q est fonction des états des cellules j des couches précédentes $q-1$, etc.

I.7.3.2.2. RETRO-PROPAGATION DE L’ERREUR

Le but étant d’obtenir pour une entrée $\mathbf{X}_k$ une réponse $\mathbf{S}_k$ la plus proche possible du vecteur $\mathbf{y}^k$ désiré correspondant, on est amené à optimiser notre réseau afin de réduire la différence entre sortie désirée et calculée comme pour l’Adaline.

Définissons la fonction de coût suivante :

$$C^k(W) = \sum_{i=1}^p (s_i^k - y_i^k)^2 \quad (\text{I.27})$$

qui est l’erreur ou distance quadratique entre la sortie calculée s^k et la sortie désirée y^k . La totalité des poids de connexions du réseau doit être ajustée en fonction de cette erreur. Cette fonction de coût dépendra bien sûr de l’état du système et de l’exemple qui lui est présenté. Pour minimiser la fonction de coût $C^k(W)$ on doit calculer son gradient par rapport à W .

$$\frac{\partial C^k(W)}{\partial w_{ij}} = \frac{\partial C^k(W)}{\partial A_i} \frac{\partial A_i}{\partial w_{ij}} = \frac{\partial C^k(W)}{\partial A_i} x_j = \delta_i^k x_j \quad (\text{I.28})$$

Ceci en posant
$$\delta_i^k = \frac{\partial C^k(W)}{\partial A_i}$$

- Pour les cellules i de la couche de sortie on peut réécrire δ_i^k de la manière suivante :

$$\delta_i^k = \frac{\partial C^k(W)}{\partial A_i} = \frac{\partial C^k(W)}{\partial s_i} f'(A_i) = 2(s_i^k - y_i^k) f'(A_i) \quad (\text{I.29})$$

- Pour les cellules i ne faisant pas partie des cellules de la couche de sortie on a :

$$\delta_i^k = \frac{\partial C^k(W)}{\partial A_i} = \frac{\partial C^k(W)}{\partial x_i} f'(A_i)$$

Mais

$$\frac{\partial C^k(W)}{\partial x_i} = \sum_l \frac{\partial C^k(W)}{\partial A_l} \frac{\partial A_l}{\partial x_i} = \sum_l \delta_l^k w_{li} \quad (\text{I.30})$$

Où les cellules l sont “en aval” de la cellule i . On obtient alors :

$$\delta_i^k = f'(A_i) \sum_l \delta_l^k w_{li} \quad (\text{I.31})$$

Pour les neurones i dans les couches intermédiaires du réseau.

Enfin, selon l'équation générale d'adaptation des coefficients w_{ij} (I. 32) :

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

Dont les Δw_{ij} sont données par l'expression

$$\Delta w_{ij} = \varepsilon^k \frac{\partial C^k(W)}{\partial w_{ij}} = \varepsilon^k \delta_i^k x_j \quad (\text{I. 33})$$

La valeur du pas d'apprentissage ε $\square$ k est typiquement assez petite et décroît dans le temps.

Les gradients δ_i^k sont donnés par les équations (I. 29) et (I. 31).

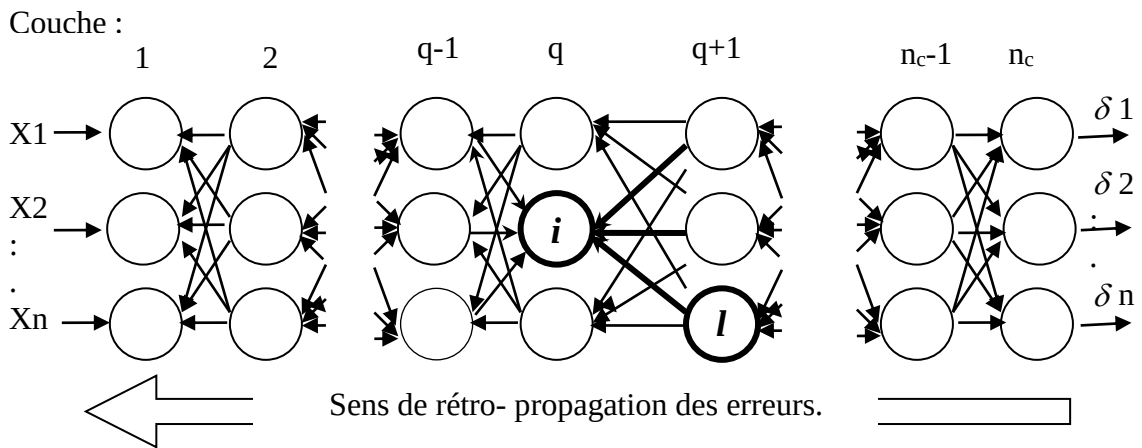


Fig.I.12. *Rétro- propagation des gradients.*

La rétro- propagation des gradients se fait “en amont” dès la couche de sortie vers la couche d’entrée. Le gradient de la cellule i dans la couche q est fonction des gradients des cellules l des couches supérieures $q+1$, etc.

I.7.3.2.3. BACKPROPAGATION AVEC MOMENTIUM

D.E.Rumelhart [4] a proposé une solution très efficace pour accélérer la convergence, qui consiste à utiliser les changements précédents des poids pour la réadaptation des poids actuels.

L'équation d'adaptation devient :

$$w_{ij}^1(t+1) = w_{ij}^1 + \eta \Delta w_{ij}^1(t) + \alpha \Delta w_{ij}^1(t-1) \quad (\text{I.34})$$

Le terme ajouté est appelé “momentum” (élan), en analogie avec le système mécanique classique, où un objet en mouvement garde l'élan acquis grâce à la quantité de mouvement qui lui a été communiquée précédemment pour accélérer son mouvement (**Fig.I.13**).

Le paramètre α est utilisé pour pondérer l'effet de ce terme. Sa valeur est généralement prise entre 0.8 et 0.9.

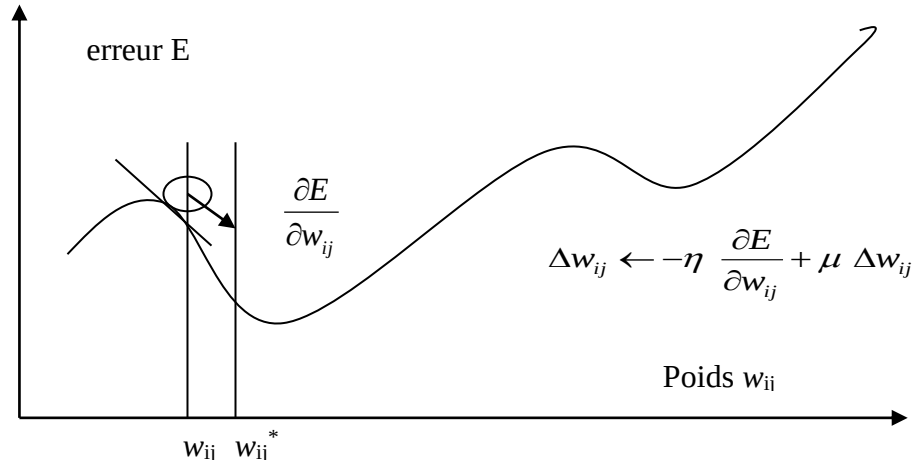


Fig.I.13. Effet du momentum pour échapper d'un minimum

I.7.4. UNE NOUVELLE ARCHITECTURE :

L'apport fondamental de ce nouveau type de réseau a été de proposer une méthode d'apprentissage efficace pour entraîner des réseaux composés d'unités non linéaires assemblés dans des architectures complexes.

I.7.4.1. LES UNITES :

Dans les Perceptrons multicouches, les unités peuvent éventuellement être linéaires comme dans l'Adaline; elles peuvent être aussi non linéaires et c'est l'intérêt de ce type d'unités. Pour des raisons à la fois historiques et pratiques, les unités employées dans la plupart des modèles neuronaux sont du type quasi-linéaire au sens où on l'a défini au paravent.

Ainsi, l'état d'une cellule i est calculé par :

$$x_i = f\left(\sum_{j=1}^n w_{ij} x_j - \theta_0\right) \quad (\text{I.35})$$

Où f est une fonction différentiable, la plus communément employée étant celle originellement proposée dans le modèles des MLP, il s'agit de la *fonction sigmoïde*; x_j représente les états des cellules connectées à la cellule i ; w_{ij} le poids de connexion entre les cellules j et i ; et θ_0 le seuil associé à la cellule i .

I.7.4.2. L'ORGANISATION DES CELLULES

L'Adaline possède des unités d'entrée et des unités de sortie. Les deux types de cellules étant accessibles par le monde extérieur et sont dites unités externes. Elles constituent respectivement l'interface d'entrée et celle de sortie du système.

Les réseaux MLP possèdent le même type d'interface, mais ont également des cellules qui ne sont ni des unités d'entrée ni des unités de sortie et qui sont appelées *cellules cachées*. L'état des cellules cachées n'étant pas accessible de l'extérieur, il est modifié en fonction des états des autres cellules du réseau. Une donnée présentée en entrée du réseau est propagée à travers

Les connexions jusqu'aux cellules cachées. De même, les états de cellules cachées sont propagés à travers, éventuellement d'autres *couches cachées*, jusqu'à la couche de sortie du réseau. La rétro-propagation du gradient permet de travailler avec un réseau à plusieurs couches de neurones, ou réseau multicouches; on parle souvent de Perceptron Multicouches. L'utilisation des automates quasi-linéaires dans l'algorithme de rétro-propagation du gradient conjuguée aux architectures multicouches va permettre à cette méthode de résoudre certain type de problèmes qu'il n'était pas possible de résoudre efficacement par les réseaux plus simples comme l'Adaline et le Perceptron.

I.7.5. METHODE D'OPTIMISATION ALEATOIRE ROM

En 1965, J. Mathias a présenté cette méthode, elle a été développée par F.J.Sollis et J.B.Wells en 1981 [4], qui ont démontré la convergence de l'algorithme vers un minimum global [7].

Nous présentons ci-dessous l'algorithme de cette méthode:

1. initialiser aléatoirement tous les poids W du réseau. Initialiser la variance v et la moyenne de la séquence aléatoire $b(o) = 0$.
2. Générer une nouvelle séquence aléatoire $\xi(k)$ de moyenne $b(k)$ et variance $v(k)$.
3. Calculer les erreurs $E = \sum_{p=1}^m E_p$ à la sortie du réseau pour chaque cas:
 - $E(w(k)) \rightarrow E_1$
 - $E(w(k) + \xi(k)) \rightarrow E_2$
 - $E(w(k) - \xi(k)) \rightarrow E_3$
4. si $E_2 < E_1$ alors
$$\begin{cases} w(k+1) = w(k) + \xi(k) \\ b(k+1) = 0.2b(k) + 0.4\xi(k) \end{cases}$$

$$\text{si } E2 > E1 \text{ et } E3 < E1 \text{ alors } \begin{cases} w(k+1) = w(k) - \xi(k) \\ b(k+1) = 0.4b(k) - 0.4\xi(k) \end{cases}$$

$$\text{Sinon } \begin{cases} w(k+1) = w(k) \\ b(k+1) = 0.5b(k) \end{cases}$$

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

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

I.7.6. RESEAUX A FONCTION DE BASE RADIALE RBF

Les réseaux de neurones à fonction de base radiale, outre les couches d'entrée et de sortie, ils comprennent une couche cachée de n unités qui ne réagissent significativement qu'à une partie restreinte dans l'espace d'entrée suivant une fonction d'activation de type gaussien[7]. A travers une combinaison linéaire de 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. Un estimateur à noyau considère des fonctions de R^+ vers R et sont de la forme :

$$\phi_i(x) = \phi(\|x - \xi_i\|) \quad (I.36)$$

L'idée principale a été introduite par M.J.D.Powell [20], toute fonction $f(x)$ d'une variable x de R^n peut être approchée par une interpolation composée par la somme de P fonctions noyaux de forme fixée $\phi(\|x - \xi_i\|)$:

$$f(x) = \sum_{i=1}^P \lambda_j \phi(\|x - \xi_i\|) \quad (I.37)$$

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

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

$$f(x) = \gamma_j \quad i = 1, n \quad (I.38)$$

$\phi(\|x - \xi_i\|)$ est une fonction assurant la continuité aux nœuds et la dérivabilité d'ordre supérieur en ces points [7].

I.7.6.1. ARCHITECTURE ET FONCTIONNEMENT DES RESEAUX RBF

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

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

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

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

Où r est un réel quelconque, remplaçant le discriminant des neurones. Mais la fonction gaussienne est la plus utilisée.

L'architecture d'un réseau RBF est représentée par la figure I. 14.

Il est composé de trois couches. La première reçoit les entrées injectées au réseau, la couche cachée est constituée de neurones, dont le discriminant est à fonction de base radiale.

Chacun de ces neurones est doté d'un vecteur C_i appelé centre, les valeurs de ces vecteurs représente chacune le barycentre de la classe qui lui correspond.

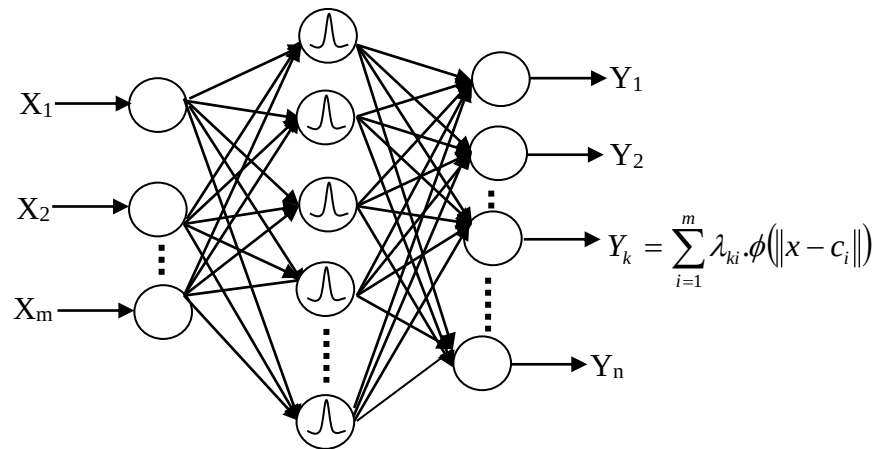


Fig.I.14. Architecture d'un réseau RBF.

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

$$d_i = \|X - C_i\| \quad (\text{I. 39})$$

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

$$s_i = \exp\left(\frac{(X - C_i)^T (X - C_i)}{2\sigma_i}\right) \quad (\text{I. 40})$$

Où σ_i est un paramètre qui représente une mesure de la dispersion des données associées à chaque nœud. La couche de sortie est constituée de neurones linéaires, effectue l'approximation définie par la somme pondérée suivante :

$$y_k(X) = \sum_{i=1}^m \lambda_{ki} \phi(\|X - \xi_i\|) \quad 1 \leq k \leq n \quad (\text{I. 50})$$

Avec n le nombre de neurones de sortie.

I.7.6.2. APPRENTISSAGE DES RESEAUX RBF GENERALISES

Ces réseaux permettent non seulement l'apprentissage des poids des connexion w , mais aussi celui des centre et des rayons de ces reseaux à base radiale généralisés. L'ajustement des centres et des rayons de ces réseaux peut être aisément effectué au moyen d'un algorithme de type descente de gradient .

Nous présentons ci-dessous les etapes de cette technique d'apprentissage supervisé :

1. initialiser les centres $c_i(0)$, les paramètre de la sortie de chaque neurone de la couche cachée $\sigma_i(0)$, $i=1, \dots, m$, et les poids synaptique $\lambda_{ij}(0)$, $j=1, \dots, n$.
2. choisir un taux d'entraînement $\alpha < 1$ variable ou fixe .
3. présentation successive de P exemples d'entraînement.

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

$$E = \sum_{p=1}^M \sum_{j=1}^n (d_j^p - y_j^p)^2 \quad (\text{I. 51})$$

avec :

M représentant le nombre d'exemples d'entraînement

y_j^p d_j^p représente respectivement la $j^{\text{ème}}$ sortie du réseau et celle désirée pour le $p^{\text{ème}}$ vecteur d'entrées.

4. Réajuster les centres et les paramètres des fonctions noyau :

$$c_i(t+1) = c_i(t) + \alpha \Delta c_i(t) \quad (\text{I. 52})$$

$$\sigma_i(t+1) = \sigma_i(t) + \alpha \Delta \sigma_i(t)$$

Avec :

$$\Delta c_i(t) = - \frac{\partial E}{\partial \lambda_{ij}(t)} \quad 1 \leq i \leq m \quad (\text{I. 53})$$

$$\Delta \sigma_i(t) = - \frac{\partial E}{\partial \sigma_i(t)}$$

6. réajuster les poids synaptiques :

$$\lambda_{ij}(t+1) = \lambda_{ij}(t) + \alpha \Delta \lambda_{ij}(t) \quad (\alpha \text{ est le taux d'apprentissage}) \quad (\text{I. 54})$$

Avec :

$$\Delta \lambda_{ij}(t) = - \frac{\partial E}{\partial \lambda_{ij}(t)}, \quad 1 \leq j \leq n; \quad 1 \leq i \leq m \quad (\text{I. 55})$$

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

I.7.6.3. RESEAU RBF ET APPROXIMATION DE FONCTION

Les réseaux RBF constituent des modèles de réseaux très efficaces notamment pour l'approximation des 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 ce réseau, qui consiste en 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 utilisés avec une seule couche cachée, ceci libère l'utilisateur du choix du nombre de couches [8].

Néanmoins, ces réseaux présentent quelques inconvénients. En effet, lorsque le nombre d'exemples d'apprentissage est important, la couche cachée risque d'être surchargée en neurones, compliquant ainsi les calculs et ralentissant l'apprentissage.

I.8. CONCLUSION

Les réseaux multicouches entraînés par l'algorithme de rétro propagation du gradient sont les modèles connexionnistes les plus étudiés et utilisés à ce jour.

Les champs d'application de cette méthode sont très vastes tels que la prédiction, classification, l'identification de processus de série temporelles, la commande de processus et de robot, le traitement d'image et de parole.

Il est clair que le réseau apprend les exemples de base d'apprentissage, dont les réponses désirées sont élaborées par le modèle que constitue le superviseur. Si les exemples de la base de données concernent un fonctionnement dans un domaine réduit, le réseau ne saura pas répondre correctement en dehors de ce domaine. Si le superviseur donne des informations erronées le réseau apprendra un modèle erroné.

Si le processus que l'on cherche à modéliser est non stationnaire ou change brutalement, le réseau doit répondre au bon modèle.

D'un point de vue théorique, on sait qu'il existe toujours un réseau à trois couches capable d'approcher toute fonction continue.

Cependant, pour un réseau donné, on ne connaît pas le nombre optimum de neurones par couches. De plus, on ne sait pas quel est le nombre de couches du réseau de complexité minimale, en nombre de paramètres (connections).

Un nombre de neurones trop petit induit une modélisation insuffisante, mais un trop grand nombre de neurones entraîne une sur paramétrisation du modèle, qui nuira aux performances en généralisation. La taille du réseau et celle de la base d'apprentissage sont liées.

D'un point de vue pratique, la principale difficulté est l'optimisation de la phase d'apprentissage. Le choix de l'architecture adéquat ou la détermination du « pas d'apprentissage » se fait par essais successifs. L'utilisation d'une base indépendante de celle d'apprentissage, appelée base de test permet de déterminer le réseau optimal. On détermine les poids du réseau à partir de la base d'apprentissage et on calcule les performances sur la base de test. Le réseau « optimal » est celui qui minimise l'erreur commise sur la base de test.

Une autre difficulté est liée aux caractéristiques de la base d'apprentissage, aussi bien en terme de taille et de représentativité que de répartition des exemples. Le nombre d'exemples doit être suffisamment grand devant les paramètres (les poids) à déterminer. Le domaine de validité de l'algorithme neuronal est directement lié à la représentativité des exemples de la base d'apprentissage. L'irrégularité des exemples dans certaines régions de l'espace des données peut conduire à une mauvaise optimisation du réseau. Dans ces conditions, un soin particulier doit être appliqué lors de la création des bases de données; une méthode d'échantillonnage apparaît primordiale.

L'algorithme d'optimisation aléatoire ROM, constitue une alternative très intéressante. En effet, cet algorithme ne dépend pas du système pour lequel le réseau est utilisé. De plus, sa technique de recherche permet d'atteindre un minimum global.

Les réseaux LBF souffrent de l'indétermination du nombre optimal d'éléments dans chaque couche. Mais concernant la dimension du réseau, il est clair qu'un réseau à deux couches constitue un approximateur universel.

Les réseaux RBF se distinguent par leur caractéristique de meilleurs approximateurs.

En effet, ces réseaux sont dotés de la caractéristique de calcul et d'adaptation paramétrique locaux. Ceci fait que ces réseaux soient à l'abri de plusieurs problèmes que connaissent les LBF, comme le sur-apprentissage, et la difficulté d'introduction de nouveaux exemples sans toucher à la précision d'apprentissage des exemples précédemment appris.

II.1. INTRODUCTION

Les huiles isolantes sont largement utilisées comme diélectrique dans l'industrie d'électrotechnique à haute tension, et plus précisément les huiles minérales qui sont utilisées dans l'élaboration de l'isolation et le refroidissement du matériel électrique en service comme les transformateurs, les cabales, les disjoncteurs et les condensateurs.

Durant l'exploitation, les huiles se détériorent progressivement à cause des différentes contraintes. Cette dégradation au cours du temps est connue sous le nom de vieillissement.

Dans cette partie, nous présentons quelques notions et définitions qui concernent les diélectriques liquides. Ainsi que quelques caractéristiques des huiles de transformateur utilisée par la SONEGAS et dénommée BORAK22.

II.2. LES HUILES MINÉRALES ISOLANTES

Depuis 1938, les bruts naphténiques paraissent bien adaptés à la production des huiles minérales isolantes. Elles représentent la fraction lourde de pétrole brut, et sont constituées donc de mélanges très complexes de composés appelés hydrocarbures [1].

II.2.1. LES ROLES D'UNE HUILE ISOLANTES

Les principaux rôles d'une huile isolante sont le refroidissement (éviter la surchauffe et évacuer les calories vers les parois) et l'isolement électrique (prévenir les arcs, éviter la reprise d'humidité et éviter les fines bulles d'air et de gaz piégés dans les enroulements). L'huile de transformateur a aussi une très bonne aptitude à résister au vieillissement.

II.2.2. CHOIX D'UN DIELECTRIQUE LIQUIDE

Le choix d'une huile isolante est déterminé en fonction de ses caractéristiques physiques et chimiques que l'on désire lui conférer. Comme la viscosité, la densité, la stabilité ..., ainsi qu'en fonction de la nature chimique recherchée (paraffinique, naphténique ou mixte).

II.2.3. CARACTERISTIQUES D'UNE HUILE MINÉRALE ISOLANTE

Comme on a dit, le choix d'une huile isolante est lié à certaines propriétés physiques, chimiques et électriques.

II.2.3.1. PROPRIETES ELECTRIQUES

- La **permittivité** définit la possibilité sous l'action du champ électrique, de libérer des charges. Elle résulte de diverses contributions de déplacement de charges [9]. la permittivité relative « ϵ_r » est définit comme le rapport « ϵ / ϵ_0 » de la permittivité absolue « ϵ » du liquide à

celle de vide « ϵ_0 » ($\epsilon_0 = (1/36\pi)10^{-9}$ F/m). la permittivité relative des huiles est comprise entre 2.1 et 2.5 à 90°C [9].

- **La conductivité** σ d'un milieu qui renferme n porteurs de mobilité μ_i et de densité q_i exprime le rapport constant de la densité de courant $\vec{j}$ au champ $\vec{E}$, selon l'équation suivante [10] :

$$\vec{J} = \left(\sum_{i=1}^n \mu_i q_i \right) \vec{E}$$

- **La résistivité** ρ est définie comme l'inverse de la conductivité σ . Dans le cas d'une huile isolante la résistivité est comprise entre 20 et 2000GΩm à 90°C [9].

En pratique, pour mieux caractériser un diélectrique du point de vue pertes, il est usuel d'utiliser la tangente de l'angle de pertes δ (**facteur de pertes diélectrique**). C'est un facteur qui augmente avec l'augmentation de la température et dépend de la tension appliquée [10bija]. Le milieu du diélectrique sous tension alternative sinusoïdale de pulsation ω , il obéit à la loi d'Ohm et il peut être représenté par un schéma équivalent comportant une résistance R et une capacité C en parallèle [10].

Donc :

$$\text{tg}(\delta) = 1 / RC\omega = \sigma / \epsilon\omega = 1 / \rho\epsilon\omega.$$

- **La rigidité diélectrique** d'une huile est la valeur maximale du champ électrique que l'on peut lui appliquer sans décharge.

On peut représenter quelques facteurs qui peuvent influencer sur la rigidité diélectrique :

- **L'humidité** : la présence d'un peu d'humidité dans le liquide diminue considérablement la rigidité diélectrique de ce dernier.
- **La pression hydrostatique** : la tension de claquage croît de façon remarquable quand la pression augmente sans indication de saturation en polarité négative, cette croissance est bien moins rapide et tend à disparaître en polarité positive [11]. Cette forte dépendance de pression en polarité négative indique que le mécanisme des bulles est dominant, sans pour autant exclure les mécanismes gazeux [12].
- **Distance entre électrodes** : en augmentant la distance entre électrodes la tension de claquage augmente, mais la rigidité diélectrique du liquide diminue.
- **Température** : l'influence de la température est double. En augmentant la température, la concentration de gaz et de l'humidité diminue dans le liquide. Cependant les conditions pour avoir un claquage purement thermoélectrique s'améliorent [13].

- **Durée d'application de la tension** : plus la durée d'application augmente plus l'effet des différentes impuretés se fait ressentir et la tension de claquage diminuent [9].

II.2.3.2. PROPRIÉTÉS PHYSIQUES

- **La viscosité dynamique** ou **La viscosité absolue** η est le coefficient de proportionnalité entre les réactions du milieu (qui sont provoquées par les contraintes de cisaillement ou de glissement dans un liquide) et le gradient de vitesse [14].

La viscosité diminue avec la température et augmente avec la pression. L'expérience montre que dans le cas d'une huile minérale, la viscosité double quand la pression passe de la pression atmosphérique à 350 bar.

- **Le point d'écoulement** est la température à laquelle les liquides se figent, il nous permet d'évaluer les performances d'une huile à basse température [14]. En particulier, dans les transformateurs avec circulation forcée. Ainsi, la température à laquelle la viscosité permet le fonctionnement des pompes est de l'ordre de -25°C à 35°C pour des points d'écoulement de -45°C [16][13].

- **La conductivité thermique** est mesurée en Watt par mètre Kelvin, elle exprime le flux thermique s'écoulant, en régime permanent sous l'effet d'un gradient thermique entre deux isothermes du liquide. Elle décroît lorsque la température et la masse volumique augmentent [16].

- **Le point d'éclair** indique la température de l'huile pour laquelle les vapeurs émises au contact de l'air et approchées d'une flamme prennent feu sans atteindre la combustion totale. Sa détermination renseigne quant à la présence d'hydrocarbures gazeux dissous dans l'huile par la suite de défauts électrique ou thermique [1].

- **Le point de feu** est le point d'inflammation de l'huile. Il indique la température minimale pour laquelle la combustion des vapeurs d'un liquide sera entretenue [16].

- **La couleur** est aussi un paramètre souvent considéré. En générale, les liquides isolants sont incolores ; certains, comme les huiles minérales ont une couleur ambrée ou une fluorescence caractéristique. La variation de la couleur dans des livraisons d'huile neuve peut indiquer un changement dans l'origine de l'huile ou dans la fabrication, sans que cela soit absolument certain. Dans le cas d'une huile usagée, une coloration foncée indique sa détérioration [1].

II.2.3.3. PROPRIETES CHIMIQUES

- **L'acidité** d'une huile en service résulte des produits d'oxydation. Pour une huile neuve l'acidité est de l'ordre de 0,5 mgKOH/g [13].

- Les propriétés électriques des liquides sont affectées par **leur teneur en eau** ou, mieux, par l'association eau-particules. La quantité d'eau dissoute dans un liquide à saturation dépend de sa nature et s'exprime en parties par million en masse [p.p.m., mg/kg ou 10^{-6}]. L'eau dissoute dans un liquide, avant saturation, est à l'état de vapeur. Sa teneur dépend de la température et de la pression partielle d'eau dans l'atmosphère au-dessus du liquide, c'est-à-dire de l'humidité relative extérieure. Si cette dernière est de cent pour cent, le liquide se sature en eau et, à une température donnée, cette saturation correspond aux premières gouttes d'eau condensée précipitables, généralement sous l'aspect d'un trouble [14].

- **La stabilité chimique** concerne la stabilité à l'oxydation, la stabilité thermique et, dans certains cas, la stabilité à l'hydrolyse, ensemble de processus réactionnels désignés sous le terme de vieillissement et responsables de la durée de vie des liquides en service, qui devrait correspondre à celle des matériels.

La stabilité à l'oxydation des liquides repose sur l'expérience acquise avec les huiles minérales. L'oxydation de ces huiles se traduit par un accroissement de l'indice d'acidité totale, en abrégé TAN (de l'anglais Total Acidity Number). Cet indice est exprimé en [mg KOH/g d'huile] et il est la somme de l'acidité soluble ou indice de neutralisation et de l'acidité volatile, tout deux exprimés également en [mg KOH/g d'huile]. L'acidité soluble a une influence sur la dégradation thermique de l'isolation cellulosique. Les huiles, de couleur claire à jaune, deviennent de plus en plus foncées et, à partir d'une certaine valeur de TAN forment, des dépôts insolubles ou boues d'oxydation qui se déposent sur l'isolant solide, réduisant l'évacuation des pertes ; l'huile est à régénérer. Les liquides de synthèse ne forment généralement pas de dépôt [14].

II.2.4. VIEILLESSEMENT DES HUILES DE TRANSFORMATEUR

Pendant leurs utilisations, les huiles minérales isolantes sont soumises à plusieurs contraintes électriques, thermique et chimiques. Sous l'action combinée ou séparée de ces contraintes, les huiles isolantes se dégradent au cours de temps, ce qui est connue sous le nom de vieillissement. Donc, on appelle vieillissement tout phénomène se traduisant par une évolution lente et irréversible des propriétés du matériau.

II.2.4.1. VIEILLISSEMENT THERMIQUE DES HUILES DE TRANSFORMATEUR

On convient d'appeler vieillissement thermique tous phénomène traduisant une évolution qui ne fait intervenir que [17] la température, l'atmosphère et le matériaux. Il y a deux buts principaux pour l'étude du phénomène de vieillissement thermique d'un isolant liquide qui sont :

- Evaluer de durée de vie de cet isolant sous l'effet des contraintes thermiques.
- Trouver une éventuelle corrélation entre le processus de vieillissement et les contraintes qui le provoquent.

II.2.4.2. LOIS DU VIEILLISSEMENT DES ISOLANTS

Les premiers travaux ont porté sur le vieillissement des matériaux isolants solides. En 1930 V.M.Montsinger, après avoir étudié la diminution des caractéristiques mécaniques des papiers isolant imprégnés d'huiles minérales sous l'influence de la température, propose une loi générale pour traduire le vieillissement qui s'exprime par la relation suivante [18] :

$$L(h) = C \cdot \exp(-b \cdot \theta)$$

Avec :

C et b sont des constantes.

L(h) est la durée de vie.

$\theta(^{\circ}\text{C})$ est la température.

Des études expérimentales sont venues confirmer que, dans un très grand nombre de réaction chimiques et tout au moins en première approximation, les constantes de vitesse k varient avec la température conformément à l'équation d'Arrhenius :

$$k = k_0 \exp\left(-\frac{W}{RT}\right)$$

Avec :

k_0 est un constante.

R est la constante de gaz parfaits.

T est la température.

W est l'énergie d'activation.

En s'appuyant su cette équation, W.Dakin(1948) à montré que la durée de vie L d'un matériaux isolant ou d'un système d'isolation en fonction de la température T peut se mettre sous la forme simple [2] :

$$\log(L) = A + (B / T) .$$

A et B sont des constantes.

II.3. LES MECANISMES DE CLAQUAGE DES ELECTRIQUES LIQUIDES

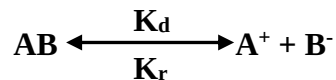
Les processus fondamentaux conduisant au claquage des isolants liquides sont encore mal connus. Ceci est lié en grande partie au fait que la connaissance des propriétés physique de l'état liquide est encore moins bien développée que dans les gaz et les solides. Ainsi, de nombreux chercheurs ont tenté d'appliquer aux liquides les théories proposées pour les gaz et les solides.

II.3.1. MECANISME ELECTRONIQUE

La conduction dans les diélectriques liquides, même sous champ intense ($E \geq 1 \text{ MV/cm}$) est presque toujours assurée par des ions de mobilité faible, qui préexistent dans le milieu ou qui se créent aux interfaces [15].

II.3.1.1. DISSOCIATION RENFORCEE PAR LE CHAMP

Il existe une réaction d'équilibre entre les molécules neutres AB d'un liquide et les ions A^+ et B^- , qui s'écrit :



Avec :

K_d : constante de vitesse de dissociation.

K_r : constante de vitesse de recombinaison.

L'application d'un champ électrique au liquide, abaisse l'énergie de liaison de molécules neutres AB. En conséquence de quoi, le nombre des ions dissociés devient supérieur aux nombres des ions recombinaison.

II.3.1.2. PRODUCTION DES CHARGES AUX INTERFACES

La création des ions aux électrodes a pour origine des réactions électrochimiques résultant de la mise en contact de l'électrode et du diélectrique. Des échanges électroniques ont lieu lorsque l'ion pénètre dans une zone de 10^{-7} cm d'épaisseur, où règne un champ très intense ($\sim 10^7 \text{ V/cm}$). Cette région est formée par des ions non déchargés fixés à la surface du métal en permanence [19].

II.3.1.3. CONSEQUENCE DU TRANSFERT DE CHARGE

Le transfert de charge s'accompagne d'une augmentation ou d'une réduction de la résistivité du liquide selon que l'élément qui réagit est un ion ou une espèce neutre [15].

II.3.2. MECANISME DE CLAQUAGE AVEC PHASE GAZEUSE

Dans les liquides ultra purs, et en géométrie pointe- plan, sous tension continue, il a été observé, qu'à partir d'un certain seuil de tension, des bulles gazeuses se forment au voisinage de la pointe, puis sont violemment chassées vers le plan. Cette phase gazeuse résulterait soit de la vaporisation et de la nucléation des bulles, soit de la cavitation [15][9].

II.3.2.1. MECANISME THERMIQUE

D'après Watson et Sharbaugh, une bulle de vapeur est générée dans le liquide par une injection de courant à partir d'aspérités sur la cathode. Ce courant serait limité par la charge d'espace et à la forme suivante :

$$I = A \cdot V^n$$

Avec :

A : est une constante.

V : est la tension appliquée, $1.5 \leq n \leq 2$.

Un modèle plus élaboré et tenant compte du régime transitoire a été proposé par KAO. Si l'intensité de courant est suffisante pour ramener la température du liquide à son point d'ébullition, il y a claquage [1].

II.3.2.2. MECANISME DE CAVITATION

Le phénomène de cavitation dans un liquide se manifeste lorsque la pression de ce liquide à température constante tombe au dessous de sa vapeur saturante [1].

D'après Krasucki, les bulles se forment en des points où le champ est élevé (particules solides de très petites dimensions dans le liquide ou attachées aux électrodes) pour qu'il y ait formation d'une cavité, le critère de pression nulle est :

$$P_{em} = P_h + P_s.$$

Avec :

P_{em} : Pression électromécanique.

P_h : Pression hydrostatique.

P_s : Pression due à la tension superficielle du liquide.

Un critère analogue a été proposé par Thomas, en supposant que la pression d'électrostriction s'oppose au décollement du liquide.

II.3.3. MECANISME DE CLAQUAGE PAR PONT

Les impuretés se trouvent dans les isolants liquides sous différentes natures (bulles de gaz, gouttelettes d'eau, particules solides isolantes ou conductrices).

La présence de ces impuretés renforce le champ localement. La déformation du champ dépend de plusieurs paramètres comme :

- La forme et la dimension des impuretés.
- L'intervalle et la forme des électrodes.
- La permittivité et la conductivité des impuretés.
- La concentration des impuretés entre les électrodes.
- La valeur des charges libres existant à la surface des particules.

L'élévation du champ peut conduire à des claquages partiels dans le cas des impuretés solide conductrices et des impuretés liquides.

Dans le cas des impuretés solides isolantes, celle-ci se polarisent sous l'effet du champ électrique et tendent à se déplacer vers les zones où le champ est le plus intense. Pour qu'un pont de particules se forme entre les électrodes, il faut que le temps d'application soit relativement lent. Quand ce pont relie les électrodes, il s'établit un fort courant à travers le pont qui est plus conducteur que le liquide. On aura ainsi, un échauffement local important qui peut engendrer le claquage [9].

Dans le cas de l'eau condensé, les gouttes d'eau sont initialement sphériques et tendent à se déformer sous l'influence du champ électrique. Elles prennent la forme d'ellipsoïdes qui s'allongent dans la direction du champ. Lorsque le volume de la goutte est suffisant, il peut ainsi se former un canal d'eau reliant les électrodes. Ce dernier engendre un fort courant qui provoquerait la vaporisation de l'eau et par suite le claquage dans le canal « vapeur - eau » en cours de formation [9].

II.3.4. EVOLUTION DES DECHARGES DANS LES DIELECTRIQUES LIQUIDES

De nombreux travaux ont porté sur la propagation de streamers, dans des systèmes d'électrodes à champ divergent. Il est généralement admis que le claquage est précédé d'une phase de pré-claquage pouvant elle-même se séparer en deux phases [1] :

- Une phase de génération caractérisée par un temps t_g dit de génération, pendant lequel apparaît au voisinage d'une électrode une perturbation révélant la forme d'arborescence.

- Une phase de propagation durant laquelle se développe la perturbation précédemment créée. Cette phase est caractérisée par un temps t_p .

Les streamers sont caractérisés par leur indice de réfraction qui est différent de ce lui du liquide. Ils produisent des courant, et émettent de la lumière.

II.4. LES ESSAIS SUR LES DIELECTRIQUES LIQUIDES

II.4.1. ESSAIS DIELECTRIQUES

II.4.1.1. MESURE DU FACTEUR DE PERTES DIELECTRIQUES ($\tan\delta$) ET DE LA PERMITTIVITE RELATIVE (ϵ_r)

Le facteur de pertes diélectriques ($\tan\delta$) et la permittivité relative (ϵ_r), sont mesurés à l'aide d'un pont de Schering, selon la norme CEI250. La cellule d'essai contenant l'échantillon d'huile est constituée de deux cylindres coaxiaux de distance entre électrodes de 5mm [20].

Les essais sont effectués sous une tension de 2kV, 50Hz, et sous une température variant de 30 °C à 120 °C, la sensibilité peut aller jusqu'à 10^{-6} .

II.4.1.2. MESURE DE LA RIGIDITE DIELECTRIQUE

La mesure de la tension de claquage est effectuée conformément à la norme CEI 156.

L'appareil utilisé est un Spintermètre. C'est une cellule d'essais ayant des électrodes sphériques de diamètre $d=12.5\text{mm}$ et un écartement $e=2.5\text{mm}$ [20].

II.4.1.3. MESURE DE LA RESISTIVITE (ρ)

La résistivité (ρ) est mesurée à l'aide d'un Meghommètre selon la norme CEI 93.

Les mesures sont effectuées sous une tension continue de 500V, appliquée durant 10mn, et sous une température variant de 30°C à 120°C [20].

II.4.2. MESURES PHYSICO-CHIMIQUES

II.4.2.1. MESURE DE LA VISCOSITE CINEMATIQUE (η)

La viscosité cinématique (η) est mesurée selon la norme NFT60-100. On procède à la mesure du temps d'écoulement d'un volume du liquide dans un viscosimètre à capillaire en verre.

La viscosité cinématique est le produit de ce temps par la constante de temps d'étalonnage de l'appareil [20].

II.4.2.2. LA COULEUR

La couleur d'une huile isolante est déterminée à l'aide d'un colorimètre en lumière transmise. Elle est exprimée par un indice obtenu par comparaison à une série de couleurs de verres étalons. L'étalon de couleur correspondant, ou celui dont la couleur est juste supérieure, est pris comme valeur de couleur [20].

II.4.2.3. L'ACIDITE

La mesure de l'acidité est réalisée par un procédé chimique simple : la neutralisation.

L'acidité est exprimée en mg de KOH par gramme d'huile. Elle correspond au nombre de mg d'une solution de Potasse alcoolique de normalité 0.1, nécessaire pour neutraliser une gamme d'huile à essayer [21].

Elle se calcule par l'équation suivante :

$$IA = 0.1 M_1 \cdot V / M$$

Où :

0.1 : normalité de la solution KOH (g/l).

M_1 : masse moléculaire de la solution KOH (g/moles).

M : masse d'huile (g).

V : volume en ml de KOH nécessaire pour avoir la neutralisation.

II.4.3. ESSAIS DE VIEILLISSEMENT DE L'HUILE

Ces essais ont pour objectif de définir les variations des propriétés de l'huile sous l'effet de contraintes appliquées durant une longue période, de préférence jusqu'à la dégradation définissant la limite de durée de vie.

Les principaux types de vieillissement de l'huile de transformateur sont le vieillissement thermique, le vieillissement électrochimique et le vieillissement électrique.

Le vieillissement thermique correspond à l'échauffement cyclique répété ou continu de longue durée, à des températures relativement élevées.

Le vieillissement électrochimique est caractérisé surtout par les conséquences, sous l'effet du champ électrique, de l'action de longue durée de certains agents chimiques accidentellement introduits dans l'huile où provenant de la propre dégradation de cette dernière.

Le vieillissement électrique est surtout caractérisé par l'action de longue durée des décharges partielles ou tout phénomène d'ionisation dû au champ électrique [20].

II.5. HUILE DE TRANSFORMATEUR BORAC22

Dans ce suit, nous présentons Les résultats d'essais effectués par des chercheurs du laboratoire de H.T. dans le cadre d'un projet de recherche en collaboration avec SONAELGAZ

II.5.1. PROPRIETES DE L'HUILE BORAC22

- **rigidité diélectrique** : La valeur de la tension de claquage de l'huile BORAK 22 neuve est de 51kV, elle correspond a une valeur de la rigidité diélectrique de 235 kV /cm [21].

Le tableau (II.1) représente la tension de claquage de l'huile avant et après vieillissement

Tableau II.1.

Etat de l'huile	Tension de claquage (kV)	Rigidité diélectrique (kV/cm)
Neuve	51	235
Vieillie à 80°C	51	235
Vieillie à 120°C	49	225

- **L'acidité** : L'indice d'acidité de l'huile neuve utilisée est très faible, à savoir 0.026mg KOH/g d'huile. Cette acidité reflète l'existence de composés organiques acides inhérents à la constitution chimique de l'huile [21].

Le tableau (II.2) représente l'indice d'acidité de l'huile avant et après vieillissement

Tableau II.2.

Etat de l'huile	Indice d'acidité (mg KOH/g)
Neuve	0.026
Vieillie à 80°C	0.046
Vieillie à 120°C	0.108

- **La teneur en eau** La mesure de la teneur en eau de l'huile neuve étudiée a donné une quantité d'eau de 30 p.p.m, (valeur acceptable) [21]

Le tableau (II.3) représente la teneur en eau de l'huile avant et après vieillissement.

Tableau II.3.

Etat de l'huile	Teneur en eau (p.p.m)
Neuve	30
Vieillie à 80°C	31
Vieillie à 120°C	35

• **La couleur** : L'indice de couleur de l'huile neuve est 1, cette huile contient une fraction minime de la fonction aromatique [21].

Le tableau (II.4) représente l'indice de couleur de l'huile avant et après vieillissement.

Tableau II.4.

États de l'huile	Indice de couleur
Neuve	1
Vieillie à 80°C	1
Vieillie à 120°C	3

1 : blanc pur

3 : citron pâle

• **La viscosité cinématique** : La viscosité cinématique, correspondant à une température fixe, est donnée par :

$$V = C \cdot t$$

Avec :

C : constante de temps d'étalonnage de l'appareille.

t : temps d'écoulement en secondes.

Pour La BORAK22 : C = 0.04787, t = 262 s.

D'où :

$$V = C \cdot t = 12.54 \text{ cSt}$$

Le tableau (II.5) représente la viscosité cinématique de l'huile avant et après vieillissement [21].

Tableau III.5.

États de l'huile	Viscosité cinématique (CST)
Neuve	12.54
Vieillie à 80°C	12.54
Vieillie à 120°C	12.49

II.5.2. INFLUENCE DE LA TEMPERATURE SUR LES PROPRIETES DE L'HUILE

II.5.2.1. FACTEUR DE PERTES DIELECTRIQUES

Le facteur de pertes diélectrique augmente en exponentiel en fonction de la température (**Fig.II.1**). Cette évolution est due en même temps à la réduction de la viscosité causant une augmentation de la mobilité des porteurs de charges et à l'augmentation du nombre des porteurs à cause d'une plus grande dissociation thermique.

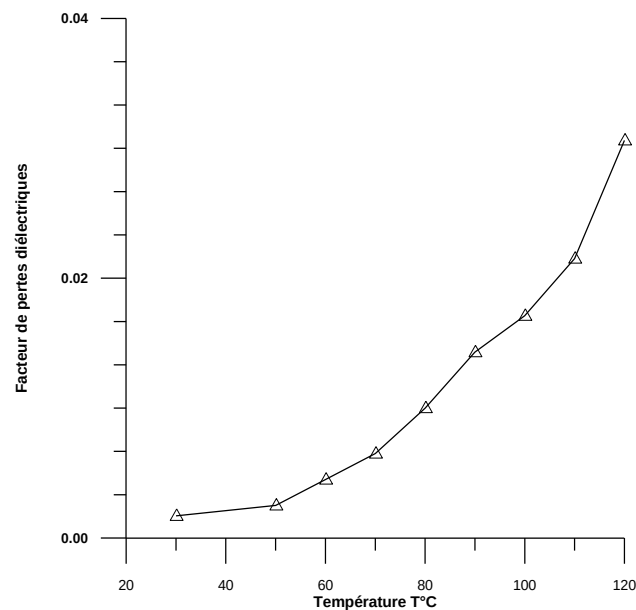


Fig.II.1. Variation du facteur de pertes diélectriques en fonction de la température [22]

PERMITTIVITE RELATIVE ϵ_R

La permittivité relative décroît presque linéairement avec la température (**Fig.II.2.**).

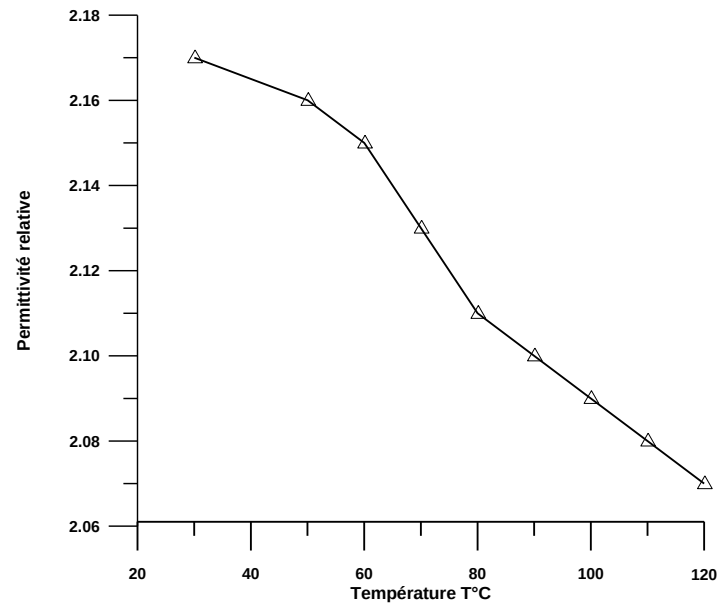


Fig.II.2. Variation de la permittivité relative en fonction de la température [22]

RESISTIVITE

La résistivité décroît en fonction de la température (**Fig.II.3.**).

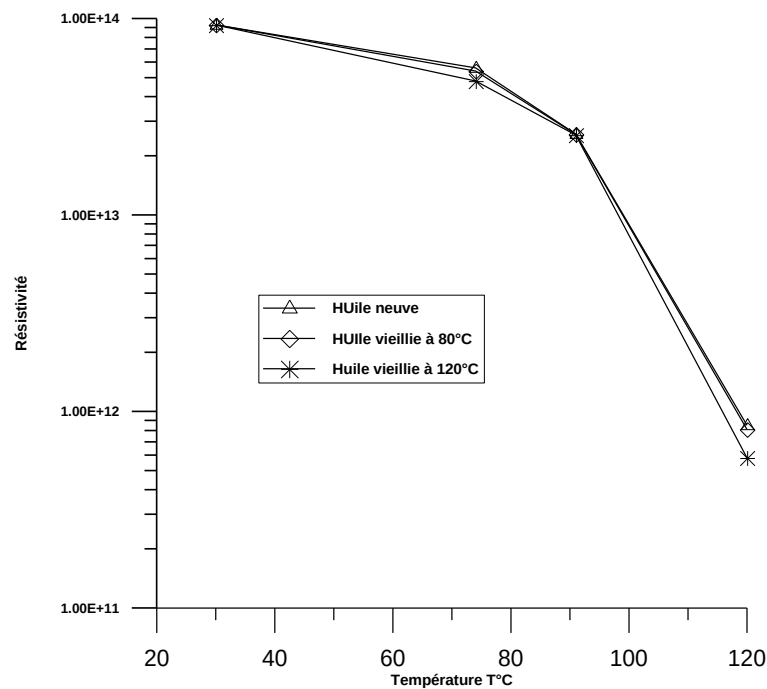


Fig.II.3. résistivité en fonction de la température avant et après vieillissement [22].

II.5.3. INFLUENCE DU VIEILLISSEMENT THERMIQUE SUR LES PROPRIETES DE L'HUILE

II.5.3.1. FACTEUR DE PERTES DIELECTRIQUES

Le facteur de pertes diélectrique $\text{tg}(\delta)$ pour des différentes valeurs de température augmente au cours de vieillissement thermique, et plus que la température augmente plus il augmente rapidement (**Fig.II.4.** et **Fig.II.5.**). On peut remarquer que pour $T=50^{\circ}\text{C}$ le facteur de pertes diélectrique augmente linéairement.

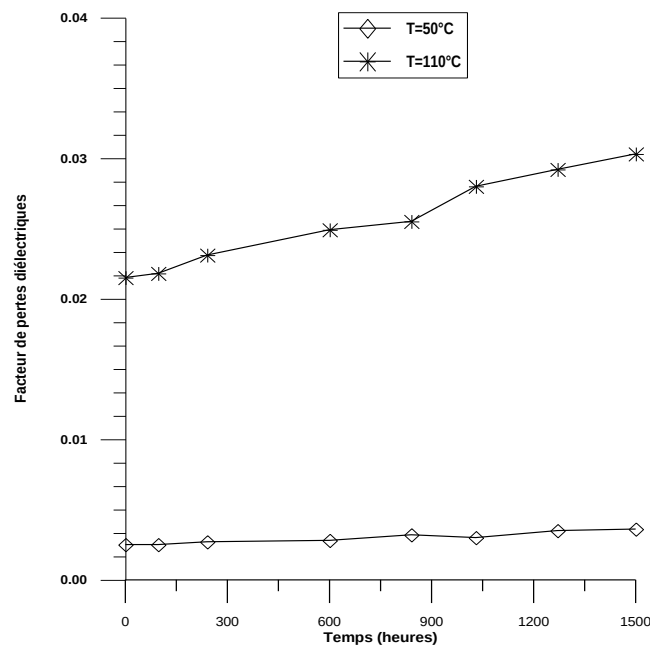


Fig.II.4.facteur de pertes diélectriques durant le vieillissement à $T=80^{\circ}\text{C}$ [22].

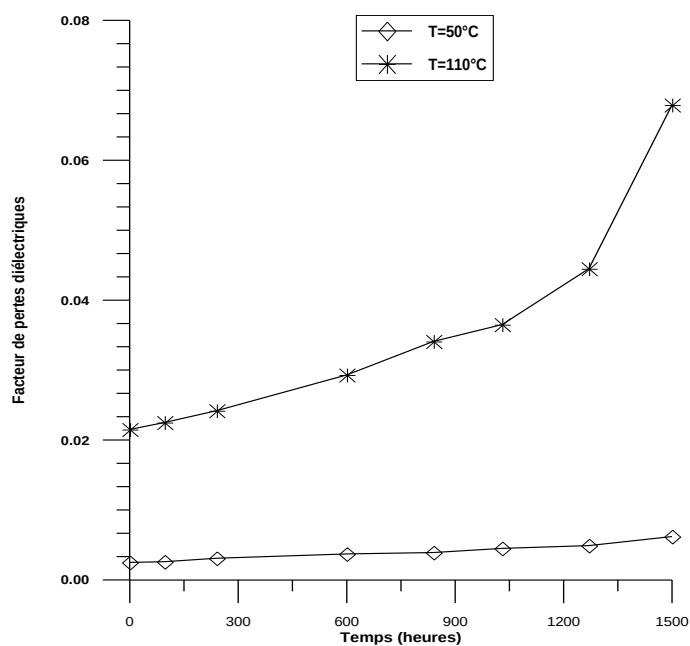


Fig.II.5.facteur de pertes diélectriques durant le vieillissement à T=120°C [22].

II.5.3.2. PERMITTIVITE RELATIVE ϵ_R :

Les figures suivantes (**Fig.II.6.**et **Fig.II.7.**) représentent la variation de la permittivité relative en fonction de temps de vieillissement thermique à T=80°C et à T=120°C. On remarque que la permittivité décroît avec une très faibles variation durant le vieillissement.

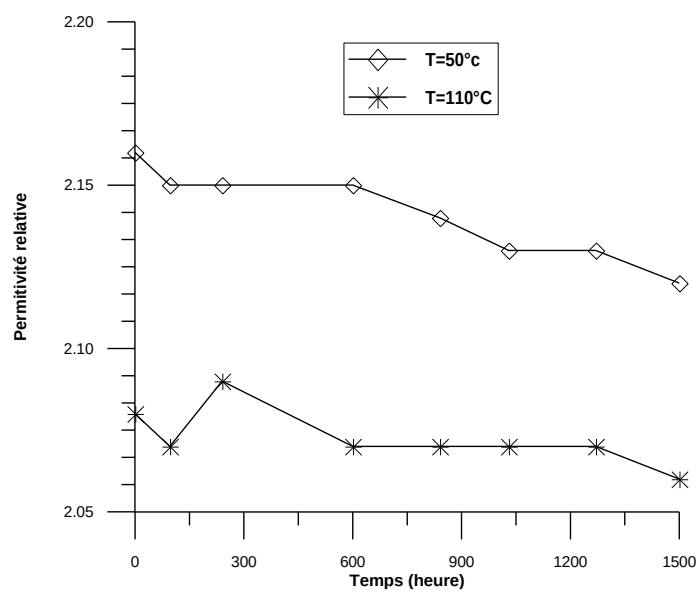


Fig.II.5.permittivité relative ϵ_r durant le vieillissement à T=80°C [22].

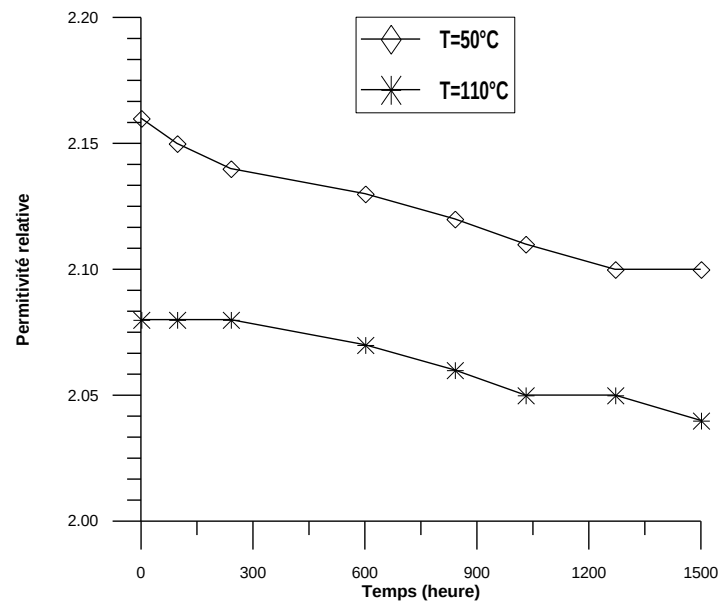


Fig.II.6.permittivité relative ϵ_r durant le vieillissement à $T=120^\circ\text{C}$ [22].

II.7. CONCLUSION

- Le facteur de pertes diélectrique augmente avec la température.
- La permittivité relative et la viscosité diminuent avec l'augmentation de la température, mais la permittivité subit à une très faible variation.
- pour le vieillissement thermique de l'huile, les propriétés électriques, les propriétés physiques et les propriétés chimiques varient en fonction de la durer de vieillissement, et surtout pour celle vieillie à température élevée.

III.1. INTRODUCTION

L'objectif de ce chapitre est de réaliser un programme de calcul et leurs résultats qui nous permettant de prédire certaines propriétés électriques d'une huile isolante (BORAC 22) en fonction du temps de vieillissement sous une température constante.

Pour une bonne prédiction, l'élaboration de la meilleure architecture possible du réseau est très importante. Ceci fait l'objet de la première partie de ce chapitre.

Si on arrive à prédire la caractéristique du programme donnée avec une erreur acceptable par rapport à la caractéristique expérimentale, avec une durée d'apprentissage relativement courte par rapport au temps global de vieillissement et d'un pas d'apprentissage le plus grand possible. Donc on dira que la prédiction de notre réseau est bonne.

On ce qui concerne les résultats de prédiction, nous présentons quelques résultats
Des trois programmes que nous avons élaborés et on jouant sur les paramètres suivants :

- L'intervalle d'apprentissage.
- Le pas d'apprentissage.
- Le nombre de neurones dans la couche cachée.
- Le taux de recouvrement

Notre intérêt est porté sur la recherche des paramètres communs qui nous donnent une bonne prédiction pour toutes les courbes à utiliser, en s'approchant de ces courbes pratiques c'est-à-dire choisir un pas en heures où le manipulateur peut prendre des valeurs, et d'essayer d'optimiser le temps d'apprentissage de telle sorte qu'il soit le plus court possible, cela pour voir le comportement du réseau dans un intervalle élevé.

Les résultats prédits vont être comparés aux valeurs de la courbe expérimentale. S'ils coïncident avec une certaine marge d'erreur, on juge la prédiction comme bonne, et que notre réseau a une certaine fiabilité de prédire des points à l'extérieur de la courbe expérimentale.

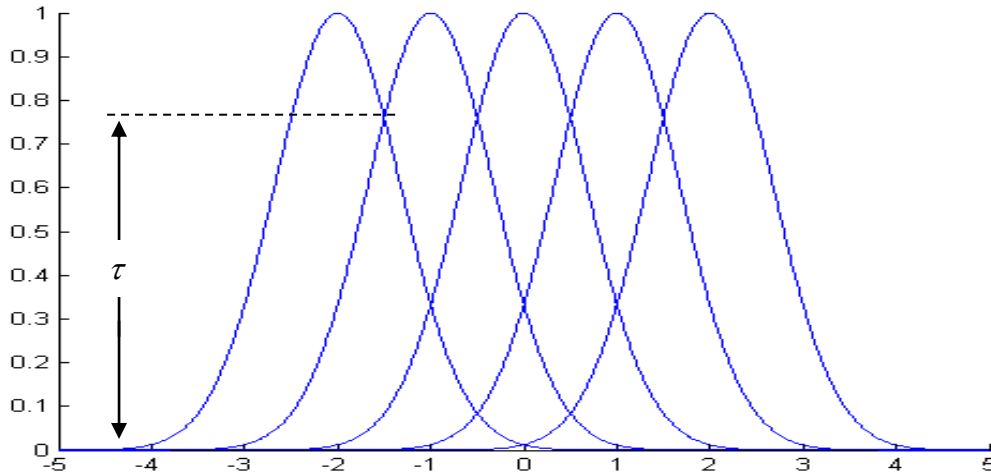
Les résultats de prédiction présentés dans ce chapitre concernent aux facteurs des pertes diélectriques à différentes températures d'essai.

III.2.ARCHITECTURE ET FONCTIONNEMENT DES RESEAUX RBF

Parmi toutes les architectures que nous avons citées, nous avons opté pour l'architecture d'un réseau RBF composé de trois couches, une seule couche de sortie Y, et une couche cachée de nombre de neurones N_c variable de fonction d'activation gaussienne (l'utilisation de cette fonction permet de bénéficier de sa caractéristique locale pour faciliter

l'apprentissage et améliorer la généralisation). Chacun de ces neurones est doté d'un vecteur C_i appelé centre, les valeurs de ces vecteurs représentent chacune le barycentre de la classe qui lui correspond.

Les centres des fonctions gaussiennes sont choisis par plusieurs techniques afin de les mettre comme points représentatifs dans les zones de grandes concentrations de données.



FigIII.1. Schéma d'un treillis régulier unidimensionnel

Mais il existe une méthode plus simple qui consiste à disposer les centres en treillis réguliers pour couvrir uniformément la partie utile de l'espace d'entrée. 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. La figure III.1. définit le taux de recouvrement τ d'un réseau à base radial à treillis régulier unidimensionnel, en supposant les gaussiennes d'écart type constant σ et une distance entre centres de δ , le taux de recouvrement est donné par :

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

Pour un réseau à treillis et uniforme sur le domaine d'approximation, le nombre de centres (L) et le taux de recouvrement τ définissent totalement la topologie du réseau.

Mais quelquefois, la disposition des centres à priori s'avère impraticable, car, le nombre de centres nécessaires est parfois trop élevé (dans le cas où l'on décide d'adopter une disposition en treillis régulier).

Mais en ce qui concerne le nombre d'entrées, nous avons utilisé un modèle, avec une seule entrée

Donc l'architecture de notre réseau est comme suit :

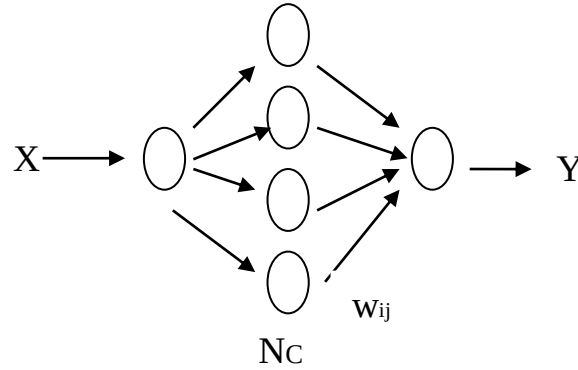


Fig.III.2.Architecture de réseau RBF

$$y_i(X) = \sum_j^{Nc} w_{ji} \rho_j(\|X - C_j\|) \quad 1 \leq i \leq n + f \quad (\text{III.2.})$$

Où :

$$\rho_{ij}(\|X_i - C_j\|) = \exp(-1/2 \sum_{j=1}^i \frac{(X_i - C_j)^2}{\sigma_j^2}) \quad 1 \leq j \leq Nc \quad (\text{III.3.})$$

Tel que :

Nc : est le nombre de neurones dans la couche cachée.

n : est le dimension de vecteur Y_1 au début de l'apprentissage.

f : est le nombre de valeurs à prédire.

C_j : est le centre du $j^{\text{ème}}$ neurone de la couche cachée.

ρ_{ij} : est l'élément de la matrice de réponse de fonction d'activation du $j^{\text{ème}}$ neurone.

III.3. CODAGE DE L'INFORMATION

Pour que notre réseau reste souple durant le test des différents signaux nous avons préféré de laisser quelques paramètres variables afin de choisir le réseau convenable pour une telle caractéristique. Ces paramètres sont :

- La durée globale du signal (pour notre application, c'est le temps global de vieillissement).
- La durée relative d'apprentissage par rapport à la durée globale (en pour cent).

- Le pas d'apprentissage relatif par rapport à la durée d'apprentissage (en pour cent).
- Le nombre de neurones dans la couche cachée
- Le taux de recouvrement

III.4. PHASE D'APPRENTISSAGE DU RESEAU RBF AVEC ROM

Nous avons essayé de tester les performances des réseaux RBF (pour une seule entrée et) avec la méthode d'optimisation aléatoire dont nous présentons ci-dessous l'algorithme.

1. Initialiser aléatoirement tous les poids W_j du réseau. Initialiser la variance v et la moyenne de la séquence aléatoire $b_j(o) = 0$.
2. Générer une nouvelle séquence aléatoire $\xi_j(k)$ de moyenne $b_j(k)$ et variance $v(k)$.
3. Calculer les erreurs $E = \sum_{p=1}^m E_p$ à la sortie du réseau pour chaque cas :
 - $E(w(k)) \rightarrow E_1$
 - $E(w(k) + \xi(k)) \rightarrow E_2$
 - $E(w(k) - \xi(k)) \rightarrow E_3$

$$\text{a. Si } E_2 < E_1 \text{ alors } \begin{cases} w_j(k+1) = w_j(k) + \xi_j(k) \\ b_j(k+1) = 0.2b_j(k) + 0.4\xi_j(k) \end{cases} \quad (\text{III.3.})$$

$$\text{b. Si } E_2 > E_1 \text{ et } E_3 < E_1 \text{ alors } \begin{cases} w_j(k+1) = w_j(k) - \xi_j(k) \\ b_j(k+1) = 0b_j(k) - 0.4\xi_j(k) \end{cases} \quad (\text{III.4.})$$

$$\text{Si non } \begin{cases} w_j(k+1) = w_j(k) \\ b_j(k+1) = 0.5b_j(k) \end{cases} \quad (\text{III.5})$$

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

Organigrammes :

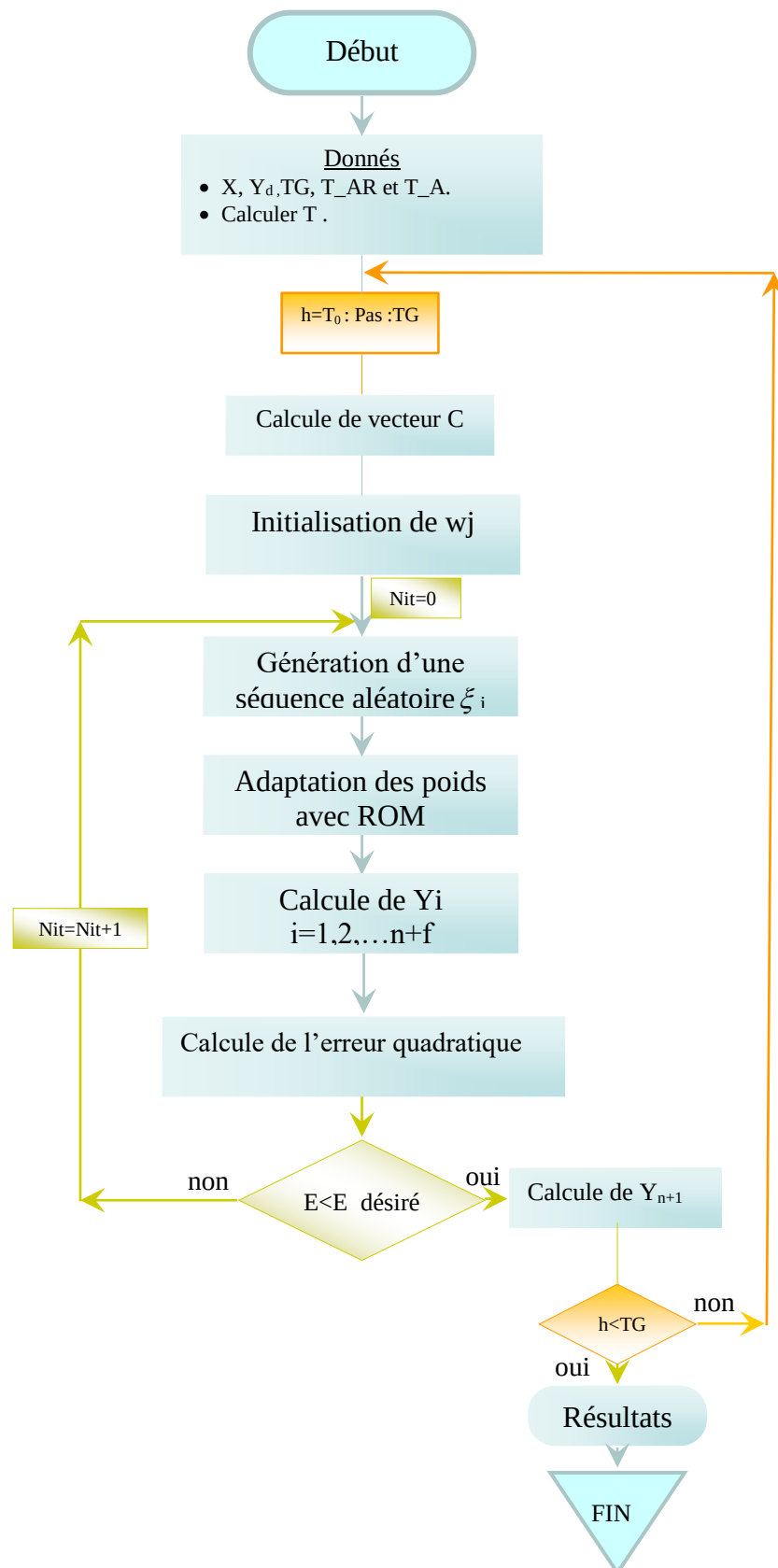


Fig.III.3.Organigramme du réseau RBF

III.5. RESULTATS A PARAMETRE CONSTANTE

Après la recherche de la bonne configuration de notre réseau (taux de recouvrement et nombre de neurones dans la couche cachée pour un temps d'apprentissage 2500 heures et un pas de 500 heures), nous avons obtenus des paramètres qui donnent une bonne prédiction. Ces paramètres sont fixés pour tous les réseaux, c'est-à-dire un taux de recouvrement $\tau = 98\%$ et un nombre de neurones de la couche cachée $N_c = 8$.

Voici les résultats obtenus par le réseau RBF avec un pas constant de 500h et un temps d'apprentissage de 2500h, pour la propriété qui concerne le facteur des pertes diélectriques

D'après les résultats représentés ci-dessous, nous pouvons dire que ce réseau à bien prédit sur la courbe (a) pour le facteur des pertes diélectrique avec un erreur moyenne inférieur à 2% (Tableau III.1). Malgré que la marge d'erreur soit relativement importante pour le reste des courbes de facteur des pertes diélectriques (Fig.III.4. b, c et d) mais la sortie du réseau suit l'allure de ces courbes expérimentales.

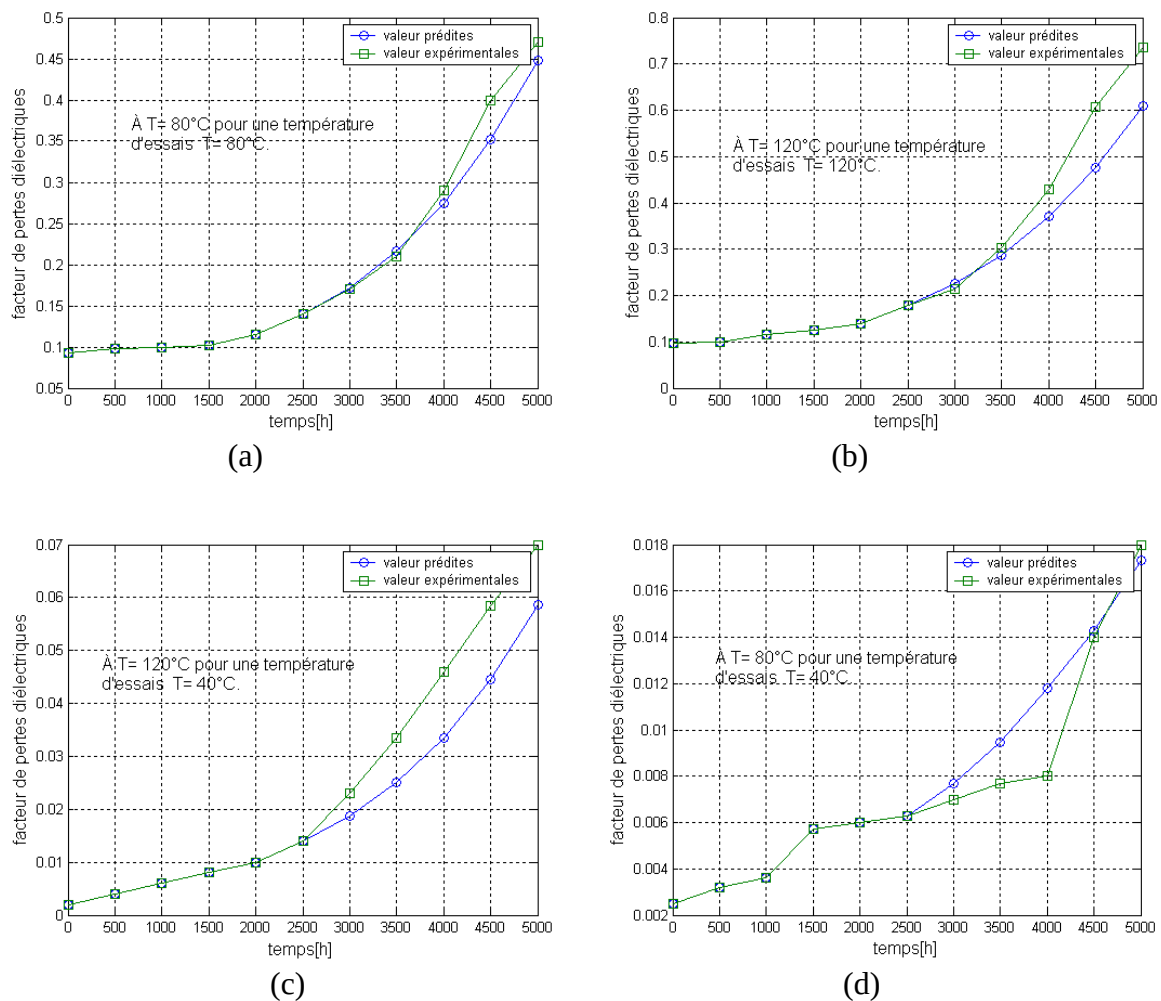


Fig.III.4. Facteur des pertes diélectriques en fonction du temps de vieillissement

	(a)	(b)	(c)	(d)
L'erreur relative minimale	2.0075	4.6291	19.3175	0.4029
L'erreur relative maximale	11.3760	26.3545	37.3085	34.0165
L'erreur relative moyenne	1.1173	3.3517	13.1692	6.4086

. **Tableau III.1.** L'erreur relative concernant du facteur des pertes diélectriques.

III.6. RESULTATS A PARAMETRES VARIABLES

Pour tester notre réseau, nous avons pris trois pas pour chaque a trois temps d'apprentissage et aussi trois valeurs de taux pour chaque trois de neurone dans la couche cachée, pour une propriété que nous avons choisie (la propriété de facteur des pertes diélectriques)

III.6.1 VARIATIONS DE TEMPS ET PAS D'APPRENTISSAGE

Dans ce cas nous avons pris trois temps d'apprentissages en heures différents (1500h, 2500h et 3500h) et Pour chaque temps nous avons pris trois pas d'apprentissages (250h, 500h et 750h).

- Temps d'aprentissage de 1500h (30%)

Pour propriété (**Fig.III.5**), il n'y a pas un pas d'apprentissage permettant d'avoir une convergence entre les résultats prédits et les résultats expérimentaux ; Donc le temps d'apprentissage 1500h n'a pas donné de résultats encourageants.

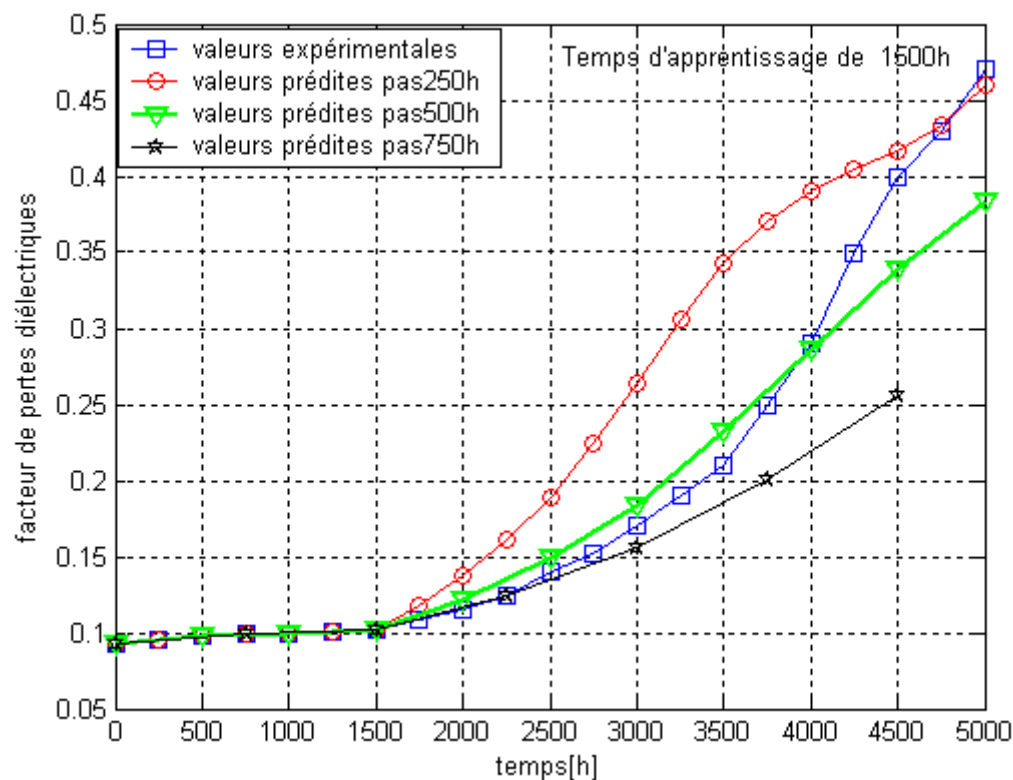


Fig.III.5.Facteur de pertes diélectriques en fonction du temps de vieillissement

pas d'apprentissages	250h	500h	750h
L'erreur relative moyenne	14.2252	6.4843	11.6334
L'erreur relative minimale	1.0846	0.3022	1.9605
L'erreur relative maximale	38.7320	21.4210	51.9954

Tableau III.2. L'erreur relative concernant le facteur des pertes diélectriques.

- Temps d'apprentissage de 2500h (50%)

D'après le figure, on constat que la courbe du valeur de pas d'apprentissage 500h suivre l'allure de la courbe expérimentale, Par contre les deux courbes de valeur (250h et 750h) seront divergente par rapport de l'allure de la courbe expérimentale.

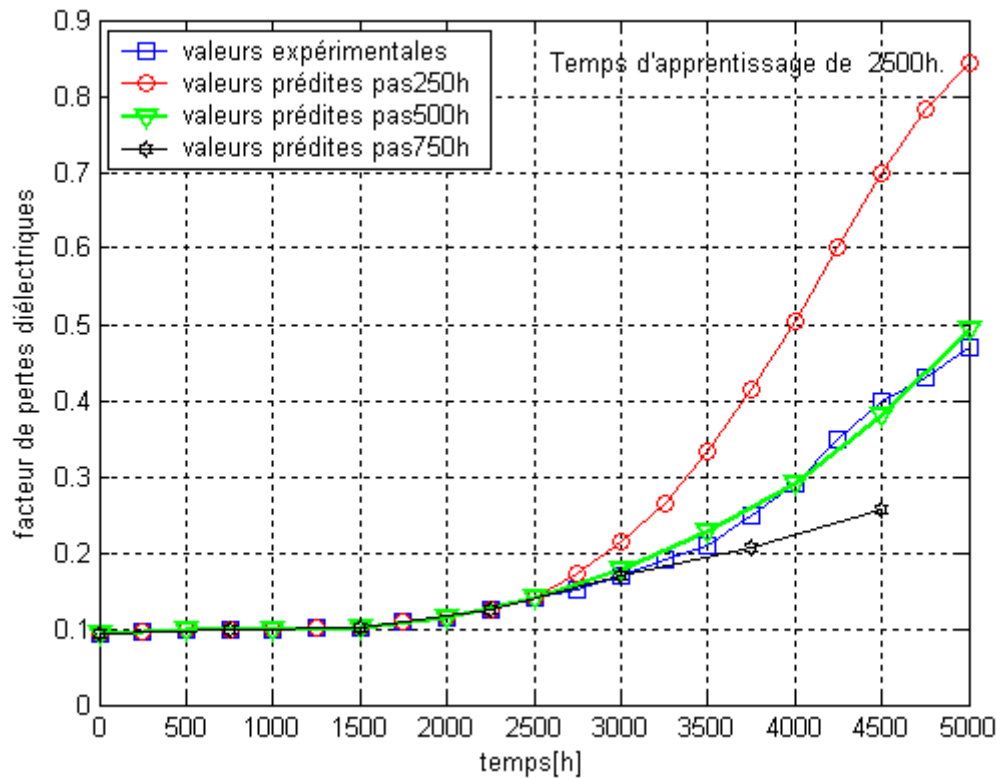


Fig.III.6.Facteur de pertes diélectriques en fonction du temps de vieillissement

pas d'apprentissages	250h	500h	750h
L'erreur relative moyenne	19.1374	3.0424	10.4867
L'erreur relative minimale	9.5933	1.3126	3.4003
L'erreur relative maximale	46.6536	15.5996	54.7628

Tableau III.3. L'erreur relative concernant du facteur des pertes diélectriques

- **Temps d'apprentissage de 3500h (70%)**

On remarque que les courbes sont totalement divergentes par rapport à la courbe expérimentale.

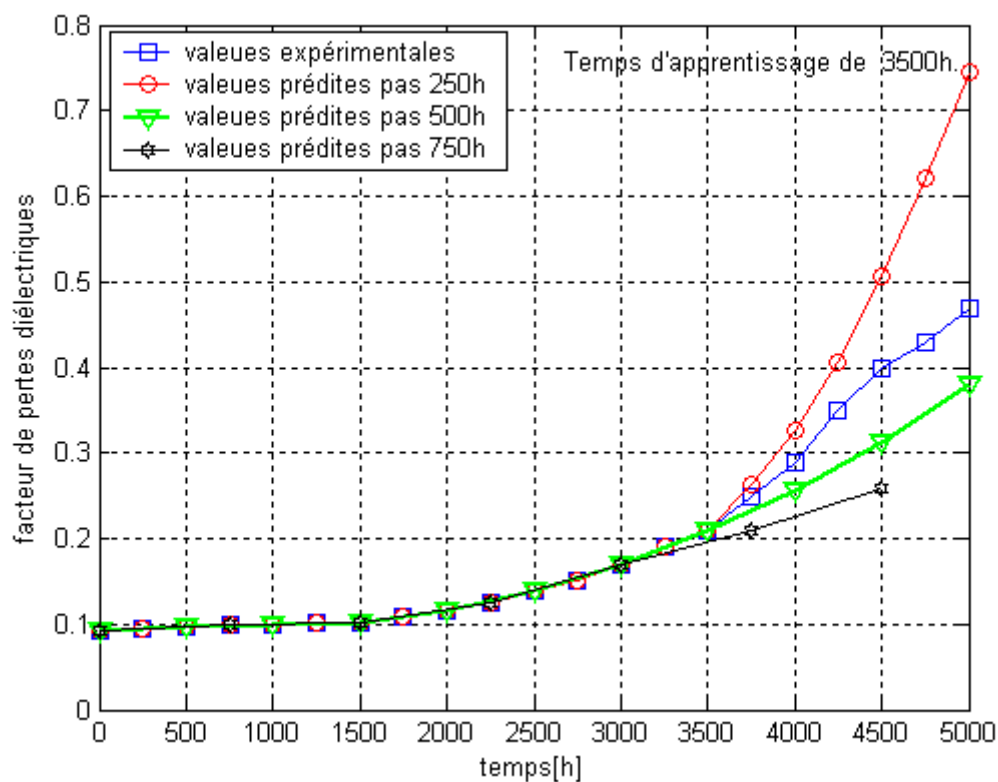


Fig.III.7. Facteur des pertes diélectriques en fonction du temps de vieillissement

pas d'apprentissages	250h	500h	750h
L'erreur relative moyenne	5.0716	5.8756	10.8282
L'erreur relative minimale	3.3950	13.3371	19.8425
L'erreur relative maximale	34.6333	28.3556	55.9549

Tableau III.4. L'erreur relative concernant du facteur des pertes diélectriques

III.6.2 VARIATIONS DE (N_c ET TAUX)

Dans ce cas, nous avons pris trois nombres différents des neurones dans la couche cachée en différents (4, 8 12 neurones). Pour chaque couche nous avons pris trois taux (0.3, 0.5 et 0.8).

- **Nombre des neurones dans couches cachée $N_c=4$**

On observant que tous les courbes sont convergentes par rapport de la courbe expérimentale, en plus de ça la courbe de valeur du taux 0.5 est coïncide avec celle de la courbe expérimentale.

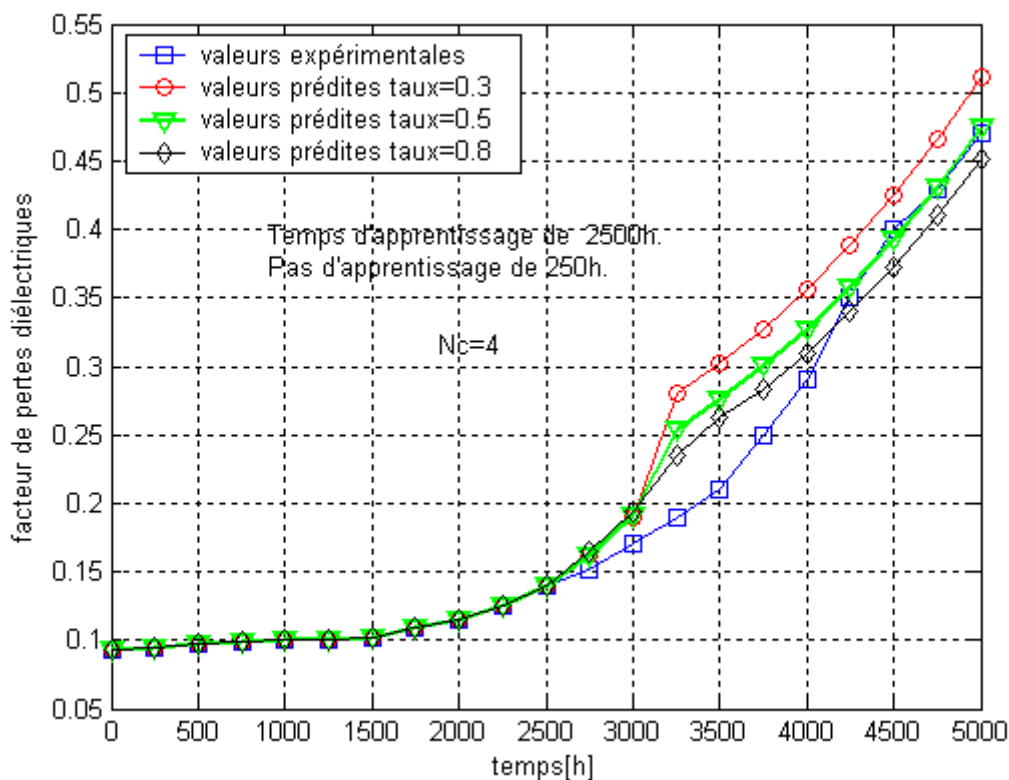


Fig.III.8. Facteur de pertes diélectriques en fonction du temps de vieillissement

Taux	0.3	0.5	0.8
L'erreur relative moyenne	7.2848	4.7610	4.6517
L'erreur relative minimale	5.8036	0.2084	2.7684
L'erreur relative maximale	32.2682	25.2310	20.5318

Tableau III.5. L'erreur relative concernant du facteur des pertes diélectriques

- **Nombre des neurones dans couches cachée NC=8**

On remarquant que les deux courbes de taux (0.3 et 0.8) suivent la courbe expérimentale mais pour la courbe de taux 0.5 s'éloigne par un intervalle remarquable de la courbe expérimentale

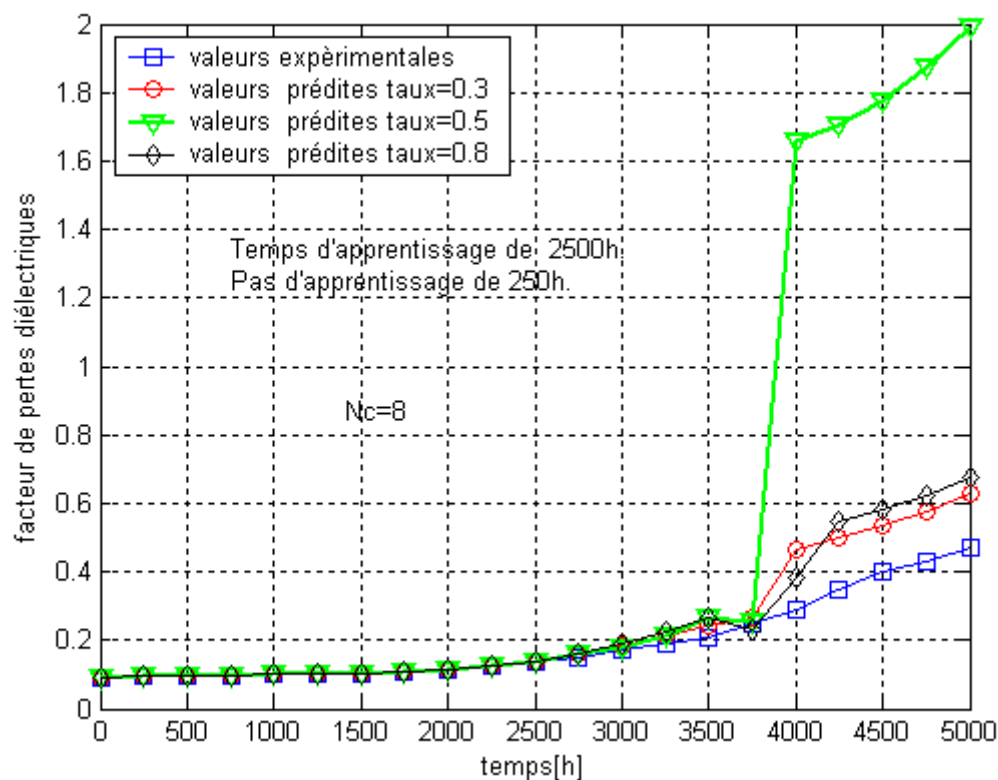


Fig.III.9.Facteur de pertes diélectriques en fonction du temps de vieillissement

Taux	0.3	0.5	0.8
L'erreur relative moyenne	7.4275	21.0821	8.3067
L'erreur relative minimale	2.9387	3.8625	7.0286
L'erreur relative maximale	34.8757	82.3246	28.5207

Tableau III.6. L'erreur relative concernant du facteur des pertes diélectriques

- **Nombre des neurones dans couches cachée NC=12**

On observant que la courbe de taux 0.3 suit la courbe expérimentale avec un minimum d'erreur, par contre les autres deux courbes de taux (0.5 et 0.8) s'éloignent d'une marge importante de la courbe expérimentale

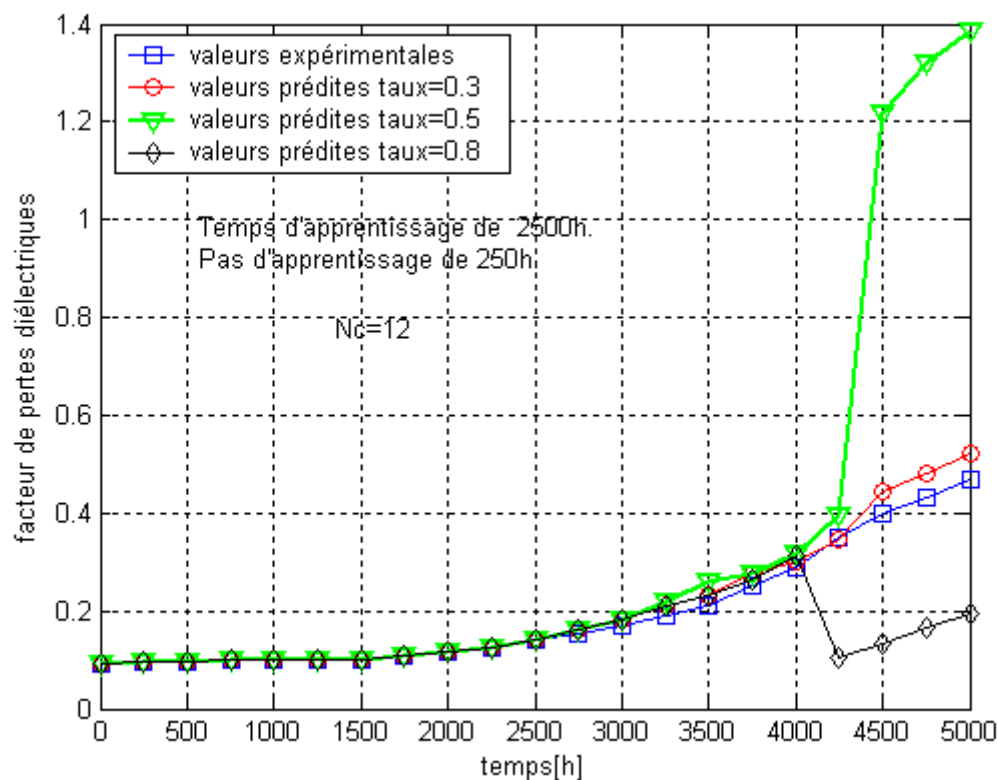


Fig.III.10. Facteur de pertes diélectriques en fonction du temps de vieillissement

Taux	0.3	0.5	0.8
L'erreur relative moyenne	3.6384	13.9169	30.0787
L'erreur relative minimale	1.3944	6.4526	7.8334
L'erreur relative maximale	13.2438	71.4363	166.1227

Tableau III.7. L'erreur relative concernant du facteur des pertes diélectriques

III.7. COMPARAISON ENTRE LES RESULTATS OBTENUS

D'après les résultats obtenus nous pouvons dire que le programme avec un temps d'apprentissage 2500h et un pas d'apprentissage 500h est le meilleur par apport aux autres.

Il faut donc choisir un pas et un temps d'apprentissage relatif convenable pour l'adaptation des poids du réseau.

Mais pour le résultat qui concernant les nombres des neurones dans la couche cachée (NC=4 et le taux 0.5) est le meilleur par apport aux autres résultats (NC=8 et 12).

III.8. INTERPRETATIONS

A l'exécution de ce programme, le nouveau point à prédire dépend sur les valeurs prédites précédemment, si nous avons une erreur au niveau de ces valeurs, celle-ci s'accumule au cours de la prédiction.

L'augmentation du nombre des neurones NC de la couche caché augmente la durée d'exécution et la prédiction est mauvaise.

Le choix correcte d'un pas et d'un temps d'apprentissage donne une bonne prédiction.

Ce réseau ne donne pas un bon résultat dans le cas où nous diminuons le temps et le pas d'apprentissage.

Si le nombre des neurones dans la couche cachée est nombreux, il faut que un taux petit pour une bonne prédiction.

III.9. CONCLUSION

Les réseaux RBF se distinguent par leurs caractéristiques de meilleure approximation, et sont à l'abri de plusieurs problèmes que connaissent les autres réseaux.

Après la discussion de tous les résultats obtenus, nous pouvons sélectionner le meilleur résultat comme approximation de propriété. Avec l'injection du temps à l'entrée du réseau et l'adaptation de la sortie de ce dernier aux valeurs expérimentales, on pourra bien avoir une approche de la caractéristique expérimentale on effectuant sur l'apprentissage du temps et le pas avec un intervalle relativement court.

Le choix du nombre des neurones dans la couche cachée et le taux de recouvrement peuvent influencer sur le temps d'exécution et les valeurs à prédire.

Les résultats obtenus par les réseaux RBF sont acceptables, mais ils nécessitent un temps d'apprentissage relativement moyen.

CONCLUSION GENERALE

Le travail que nous avons effectué consiste à utiliser les réseaux des neurones artificiels dans la prédiction de l'évolution de facteur diélectrique d'une huile isolante, soumise à un vieillissement thermique. Ce travail est l'amélioration de la prédiction d'un vieillissement thermique d'une huile de transformateur, qui prolonge la durée de prédiction de 1500h à 5000h.

Nous avons élaboré un programme de calcul du réseau RBF qui utilise le temps comme entrée et la méthode ROM pour la phase d'apprentissage.

Les résultats obtenus permettent d'aboutir aux conclusions suivantes :

- Les réseaux à base radiale normalisés se distinguent par leur caractéristique de meilleurs approximateurs des fonctions.
- Durant les tests, nous avons constaté que si on augmentant le nombre des neurones N_c de la couche cachée la durée d'exécution augmente mais la prédiction sera mauvaise.
- Avec un pas d'apprentissage petit ; la prédiction de réseau RBF est très lente quand on essaye de diminuer le temps d'apprentissage ; ceci nous oblige d'entraîner le réseau sur des intervalles de temps élevés pour que le réseau puisse trouver une dépendance entre les valeurs qu'ils ont apprises.
- Les résultats obtenus par les réseaux RBF sont acceptables, mais ils nécessitent un temps d'apprentissage relativement moyen.
- Dans notre étude, le nombre de neurones dans la couche cachée et le taux de recouvrement sont fixés, nous proposons d'optimiser ces paramètres avec une méthode convenable pour avoir des bons résultats.
- Nous mentionnons qu'à travers ce travail, nous avons constaté que le réseau des neurones est un outil très diversifié. Ce modèle a ces propres caractéristiques peut être destiné à des opérations spécifiques. Nous pensons qu'il serait intéressant de construire un réseau qui peut bien prédire.

En fin dans notre étude ; nous proposons d'étudier des autres propriétés avec un réseau de deux entrées

BIBLIOGRAPHIE

- [01] Bedja. A/Rahim et O.N.M^{ed} Elboukhary. « *Utilisation des Réseaux de Neurones Artificiels dans les Vieillessement Thermique de l'Huile de Transformateur* », thèse de PFE, ENP, Juin 1999.
- [02] H.Ailam. « *Utilisation des Réseaux de Neurones dans le Domaine de la Pollution des Isolateurs*», thèse de Magister, ENP, décembre 2001.
- [03] Rumelhart D.E, G.E.Hinton and R.J.Wiliams, « *Learning Internal Representation by Error Propagation*», 1986
- [04] Liu, « *Neural Networks for Identification, Prediction and Control* », Springer, 1994.
- [05] Kosko.B, « *Neural Networks and Fuzzy Systems. a Dynamical Machine Approach to Machines Systems* », Printece Hall International, 1992
- [06] . Freeman,J.D. and M.Skapura. « *Neural Networks, Algorithms, Applications and Programming Technique* », Adison Wesley Company, 1992
- [07] M.Yazid M'hamed Yeddou, « *Etude de Synthèse sur les Réseaux de Neurones et leur Applications* », thèse de Magister, ENP, juin 1998
- [08] J.Hérault, C.Jutten,«*Réseaux Neuronaux et Traitement du Signale*», Traité des Nouvelles Technologies, HERMES, 1994
- [9] A.Boubakeur, «*Claquage des Diélectriques Liquides* », cours de technique de HT, quatrième année électrotechnique.
- [10] R.Tobazéon, « *Conduction Electrique dans les Liquides* », techniques de l'ingénieur, D2 430, 1986.
- [11] R.Tobazéon, « *Liquides Diélectriques, Claquage et Préclaquage* », techniques de l'ingénieur, D226, 1986.
- [12] H.Moulai, « *Contribution a L'étude des Phénomènes de Streamers dans les Diélectriques Liquides en Géométrie Pointe-Plan sous Tension de Choc de Foudre* »,thèse de magister, ENP, 1990.
- [13] D.Ouakli, H.Brouri, « *Propriétés Diélectriques de L'huile de Transformateur en Présence de Gaz Dissous* », thèse de PFE, ENP, Juillet 1997.
- [14] R.Tobazéon, « *Huiles et Liquides Isolants* », techniques de l'ingénieur, D230, 1986.

- [15] S.Abdi, S.Zelmat, « *Influence du Vieillissement Thermiques sur les Propriétés de l'Huile de Transformateur* », thèse de PFE, ENP, Juin 1998.
- [16] B.Chalabi, S.Ouabdesselam, « *Analyse des Gaz Dissous dans les Huiles Isolantes* », thèse de PFE, ENP, Juillet 1997.
- [17] M.Mameri, « *Contribution a la Caractéristiques Physico-Chimique et Electrique d'une Huile Isolante de Transformateur Neuve et Usagée* », thèse de PFE, ENP, Juillet 1997.
- [18] SM.sekhri, K.Nouasri, « *Influence de la Température sur Certaines Propriétés de l'Huile de Transformateur Utilisée par La SONELGAZ* », deuxième conférence nationale sur la la haute tension, Tizi-ouzou 1997.
- [19] R.Fournié, « *Les Isolants en Electrotechnique, Essais, Mécanismes de Dégradation, Application Industrielles* », Ed Eyrolles, Paris 1990.
- [20] H.A.Nadhra, « *Contribution a l'Etude des Phénomènes de Préclaquage dans une Huile de Transformateur sous Champs Alternatif Divergent* », thèse de magister,ENP, Octobre 1997.
- [21] A.Boubakeur, « *Les Essais sur les Diélectriques Liquides* », cours de technique de HT, quatrième année électrotechnique.
- [22] A.Boubakeur, et autres, « *Influence du Vieillissement Thermique sur les Propriétés de BORAC22* », rapport interne, ENP.