

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de
La Recherche Scientifique
Ecole Nationale Polytechnique
Département de Génie Electrique



المدسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

المدسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

THESE
Présentée par

Leïla Mokhnache

Pour l'obtention du grade de

DOCTEUR D'ETAT

**APPLICATION DES RESEAUX DE NEURONES
DANS LE DIAGNOSTIC ET LA PREDICTION DES
ISOLATIONS DE HAUTE TENSION**

Devant le jury :

Prof. M. Bendaoud	Président	USTHB - Alger
Prof. A. Boubakeur	Rapporteur	ENP - Alger
Prof. D. W. Pearson	Examineur	IUT - Roanne - France
Prof. F. Boudjema	Examineur	ENP - Alger
Dr. A. Mekhaldi	Examineur	ENP - Alger
M. A. Ouabdesselam	Membre invité	ENP - Alger
M. M. Ouadjaout	Membre invité	ENP - Alger
M. B. Safta	Membre invité	PDG Transfocentre -Blida

18 janvier 2004

Dédicaces



Au-delà des personnes, des lieux et des temps et avec ma profonde conviction, je dédie ce travail à tous ceux qui ont toujours cru et sans doute à la Science. Cette lumière qui éclaire les esprits et leurs permet de transcender les limites installées par les sociétés et les cultures dans leurs périodes de dégénérescence.

À ceux qui savent concrétiser leurs idées et savent réussir.
À ceux qui ne portent que du bien pour les autres.
À ceux qui ne vivent que dans et par la vérité.

Je dédie ce travail à mes très chers parents
À ma petite famille
À ma grande famille
À Nora
À Monsieur et Madame Boubakeur
À mes collègues et amis à l'université de Batna et à L'ENP

Leïla Mokhnache

Remerciements



Mes premiers remerciements et les plus vifs, je les adresse à mon encadreur le Professeur Ahmed Boubakeur qui m'a honoré d'avoir dirigé ce travail. Je le remercie pour tous les principes qu'il n'a cessé de me transmettre, la rigueur scientifique, l'objectivité et l'orientation vers l'utile.

Je m'incline devant l'homme de science qu'il a toujours été, et je ne considère pas automatiquement comme un Homme scientifique celui qui sait se servir d'instruments et de méthodes jugées scientifiques. Je ne pense qu'à ceux dont l'esprit se révèle vraiment scientifique : cette volonté et cette foi qui les poussent à donner le meilleur de soi.

Mes remerciements vont aussi à Monsieur Mohammed Bendaoud, Professeur à l'université des Sciences et Technologies Houari Boumediene d'Alger de m'avoir fait l'honneur de présider le jury de thèse.

Je tiens à exprimer mes vifs remerciements à Monsieur David W. Pearson, Professeur de l'institut universitaire de technologie de Roanne de l'université Jean Monnet, Saint-Étienne d'avoir accepté aimablement de participer à ce jury d'examen malgré ses multiples responsabilités et d'avoir manifesté sa totale coopération.

Tous mes vifs remerciements à Monsieur Farès Boudjema, Professeur à l'Ecole Nationale Polytechnique d'Alger pour l'honneur qu'il me fait en acceptant de juger ce travail.

Je remercie également Monsieur Abdelouahab. Mekhaldi, Maître de Conférence à l'Ecole Nationale Polytechnique d'Alger de m'avoir fait l'honneur d'examiner ce travail. Je tiens à lui exprimer aussi ma profonde gratitude pour toute l'aide qu'il m'a apportée, pour ses conseils et ses encouragements.

Je tiens à exprimer aussi mes remerciements à Monsieur Mohammed Ouadjaout Chargé de Cours à l'Ecole Nationale Polytechnique d'Alger d'avoir accepté de participer à ce jury d'examen malgré ses responsabilités prenantes.

Je remercie Monsieur Abdelaziz Ouabdesselam pour l'honneur qu'il m'a fait d'examiner ce travail. Je tiens à lui dire ici ma gratitude envers lui pour avoir su porter la symbolique de valeurs qui fondent ce temple de la science qu'est l'École, et de représenter ainsi une référence et un exemple élevé pour nous. C'est pour cela que nous l'aimons tous.

Je remercie Monsieur Bachir Safta Président Directeur Général de Transfocentre-Filiale Sonelgaz à Blida d'avoir accepté de participer à ce jury d'examen et pour sa totale coopération quant à la création d'un groupe de recherche sur l'utilisation des techniques intelligentes dans le diagnostic de l'huile des transformateurs utilisant la méthode des gaz dissous

Je ne saurai oublier de manifester ma profonde reconnaissance à Monsieur Chabane Hamouda, chef de l'unité de recherche de l'université de Batna, de m'avoir accueilli à l'unité de recherche et d'avoir mis à ma disposition ce qui est nécessaire pour la réalisation de ce travail.

C'est avec tout le plaisir que j'exprime ici mes remerciements et mes sincères sentiments de respects à Monsieur Naceredine Nait-Said, chargé de Cours à L'université de Batna pour les discussions fructueuses et l'aide qu'il a apporté à l'aboutissement de ce travail.

Je remercie également Monsieur Hocine Moulai, Maître de conférence à l'université des Sciences et Technologies Houari Boumediene d'Alger pour ses remarques enrichissantes et son aide précieuse qu'il a apportée à l'aboutissement de ce travail.

INTRODUCTION GÉNÉRALE

I-III

Chapitre I INTRODUCTION AUX RÉSEAUX DE NEURONES

1.1 Historique	1
1.2 Fondement biologique	3
1.3 Fonctionnement du neurone	5
1.3.1 Physiologie du neurone	5
1.3.2 Création d'un potentiel d'action	5
1.3.3 Le corps cellulaire comme sommateur à seuil	6
1.4 Les neurones artificiels	7
1.4.1 Le neurone formel	7
1.4.1.1 Fonction discriminant	7
a) Fonction de base linéaire (LBF)	7
b) Fonction de base radiale RBF (radial basis function)	8
c) Fonction de base elliptique EBF (elliptic basis function)	8
1.4.1.2 Fonction dynamique linéaire	8
1.4.1.3 Fonction non dynamique non linéaire	8
1.4.2 Architecture des réseaux de neurones	9
1.4.3 Apprentissage des réseaux de neurones	9
1.5 Réseaux de neurones à apprentissage supervisé	10
1.5.1 Réseaux de neurones basés sur la décision	10
1.5.1.1 Le Perceptron	10
1.5.1.2 Adaline et Madaline	11
1.5.1.3 Méthode des moindres carrés (Least Mean Square Error)	12
1.5.1.4 Méthode de descente de gradient	12
1.5.2 Réseaux multicouches	13
1.5.2.1 Back-propagation	13
a) Data adaptive learning (FFN pattern)	14
b) Block adaptive learning (batch learning)	14
1.5.3 Variantes de la Back-propagation	15
1.5.3.1 Back-propagation avec Momentum	15
1.5.3.2 Back-propagation à taux d'apprentissage adaptatif	15
1.5.3.3 Méthode d'optimisation aléatoire (ROM)	16
1.5.4 Dilemme Précision-Généralisation	17
1.5.5 Les réseaux bayesiens	18
1.5.5.1 Théorème de Bayes	18
1.5.5.2 La Régularisation bayésienne	19
a) Optimisation des paramètres de régularisation	20
b) Approximation de Gauss-Newton à la hessienne	21
1.5.5.3 Avantages et inconvénients de l'approche bayésienne	22
1.5.6 Les réseaux à fonction de base radiale RBF	23
1.5.6.1 Principe de base	23
1.5.6.2 Architecture et fonctionnement des réseaux RBF	23
1.5.6.3 Apprentissage des réseaux RBF	25
a) Méthode de Centrage Adaptative (Adaptive Centering Method)	
b) Méthode du regroupement (clustering algorithm)	26
1.5.6.4 Réseaux RBF et Approximation de fonctions	27

1.6 Réseaux de neurones a apprentissage non supervise	28
1.6.1 Fondement de l'apprentissage non supervise	28
1.6.1.1 Principes généraux	29
a) L'architecture	29
b) Le neurone gagnant	29
1.6.2 L'apprentissage	29
1.6.2.1 Réseaux du type 'winner takes all'	30
a) L'algorithme LBG	30
b) Les algorithmes adaptatifs	30
1.6.2.2 Réseaux du type 'winner takes most'	32
1.6.3 Réseaux à activation compétitive	32
1.6.3.1. Principe de fonctionnement d'une couche compétitive	32
1.6.3.2 Règle d'apprentissage dans une couche compétitive	33
1.6.4 Carte auto-organisatrice de Kohonen	34
1.6.4.1 Neurone formel de Kohonen	34
1.6.4.2 Compétition dans une couche de Kohonen	35
1.6.4.3. Apprentissage et voisinage dans une couche de Kohonen	36
1.6.4.4 Caractéristiques du modèle physiologique des SOM	37
1.7 Monographie des travaux publiés dans l'application des RNA en technique de haute tension	37
1.8 Conclusion	41
1.9 Bibliographie	42

Chapitre II MÉCANISMES DIÉLECTRIQUES DANS LES MATÉRIAUX ISOLANTS

2.1 Introduction	48
2.2 Caractéristiques de vieillissement du PVC	49
a) Propriétés mécaniques des polymères	49
b) Propriétés électriques des polymères	49
2.2.1. Vieillissement thermique des polymères	49
2.2.1.1 Processus de dégradation des isolants électriques solides	49
2.2.1.2 Vieillissement des isolants électriques	50
a) Vieillissement thermique	50
b) Vieillissement électrique	51
c) Couplage de contraintes (thermique et électrique)	52
2.2.1.3 Endurance thermique du PVC	52
2.2.2. Vieillissement thermique du PVC	53
2.2.2.1 Présentation du PVC	53
2.2.2.2 Propriétés du PVC	55
a) Propriétés physico-chimiques	55
b) Propriétés électriques	55
c) Propriétés mécaniques	56
2.2.2.3 Résultats expérimentaux du vieillissement thermique du PVC	56
a) Variation de l'allongement à la rupture en fonction du temps de vieillissement	56
b) Résistance à la traction	56
c) Endurance thermique	57
d) Perte de masse	57

2.3 Propriétés et caractéristiques de vieillissement de l'huile de transformateur	
2.3.1 Pro propriétés électriques	58
2.3.2 Propriétés physiques	58
2.3.3 Propriétés chimiques	60
2.3.4 Caractéristiques de service	60
2.3.4.1 Aspect, couleur et caractéristiques optiques	60
2.3.4.2 Stabilité chimique	61
2.3.4.3 Stabilité électrique aux décharges partielles	61
2.3.5 Sécurité d'emploi	61
2.3.6 Propriétés de la BORAK22	62
2.3.6.1 Rigidité diélectrique	62
2.3.6.2 Acidité	62
2.3.6.3 Teneur en eau	62
2.3.6.4 Couleur	63
2.3.6.5 Viscosité cinématique	63
2.4 Vieillesse des huiles de transformateurs	63
2.4.1 Influence de la température sur les propriétés de l'huile	63
2.4.1.1 Pertes diélectriques	64
2.4.1.2 Permittivité relative (ϵ_R)	64
2.4.1.3 Résistivité transversale	64
2.4.1.4 Viscosité	65
2.4.2 Influence du vieillissement thermique sur les propriétés de l'huile	65
2.4.2.1 Facteur de pertes diélectriques	66
2.4.2.2 Permittivité relative (ϵ_R)	67
2.5 Rupture des intervalles d'air pointe-barrière-plan	68
2.5.1 Influence de la barrière isolante sur la tension de décharge disruptive	69
2.5.1.1 Influence de la position de la barrière dans l'intervalle d'air	69
2.5.1.2 Influence de la largeur de la barrière	70
2.5.1.3 Influence de la forme de tension	70
2.5.1.4 Influence de la conductivité de la barrière	70
2.5.1.5 Influence de la longueur de l'intervalle	71
2.6 Conclusion	72
2.7 Bibliographie	72

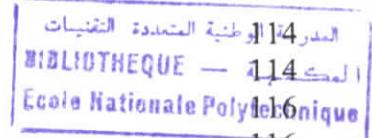
Chapitre III SURVEILLANCE ET DIAGNOSTIC D'HUILE DE TRANSFORMATEURS ET APPLICATIONS DES RNA

3.1 Introduction	75
3.2 Diagnostic	75
3.2.1 Méthodes internes de diagnostic	75
3.2.2 Méthodes externes de diagnostic	76
3.2.3 Réseaux de neurones artificiels et diagnostic	76
3.2.4 Diagnostic d'une isolation	76
3.3 Mesures et surveillance comme outils de diagnostic de l'isolation d'un transformateur	77
3.3.1 Défaillances et problèmes de transformateurs de puissance	78
3.3.2 Gestion de la vie d'un transformateur	78

3.4 Surveillance et méthodes de diagnostic d'une isolation de transformateur	79
3.4.1 Méthodes de diagnostic traditionnelles	79
3.4.1.1 Diagnostic avec analyse de gaz dissous	79
3.4.1.2 Qualité d'une huile de transformateur (ou diagnostic physico-chimique)	81
a) Critères de fin de vie basés sur les tests physico-chimiques	82
b) Efficacité du diagnostic physico-chimique	82
3.4.2 Méthodes non traditionnelles pour le diagnostic des transformateurs	83
3.4.2.1 Essai en service des décharges partielles (DP)	83
3.4.2.2 Mesure de tension de rétablissement	83
3.4.2.3 Logiciels et systèmes intelligents et experts pour le diagnostic	84
3.5 Surveillance-maintenance de l'huile de transformateur	84
3.5.1 Surveillance du transformateur en service	84
3.5.1.1 Échauffement	84
3.5.1.2 Protection du diélectrique contre l'humidité	84
3.5.2 Surveillance de l'huile en service	85
3.5.3 Remplacement et régénération des huiles de transformateurs	87
3.6 Application des RNA au diagnostic de la BORAK22	90
3.6.1 La Back-propagation (BP)	91
3.6.1.1 Back-propagation avec momentum	91
a) Réseau à une couche cachée	91
b) Réseau à deux couches cachées	92
3.6.1.2 Back-propagation avec momentum et adaptation du taux d'apprentissage (MADPTA)	93
3.6.1.3 Méthode de Levenberg-Marquardt (L.M)	94
3.6.2 Régularisation bayésienne (RB)	96
3.6.3 Paramètres influant l'apprentissage des réseaux du diagnostic	100
3.6.3.1 Méthode du momentum	100
3.6.3.2 Méthode MADPTA	101
3.6.3.3 Méthode L.M	101
3.6.3.4 Méthode de régularisation bayésienne (RB)	102
3.6.4 Influence de la base de données sur l'apprentissage	102
3.6.4.1 Phase d'apprentissage	102
3.6.4.2 Phase de généralisation	103
3.7 Architectures et apprentissages des RNA utilisés dans la classification de la BORAK22	105
3.7.1 Les Réseaux Compétitifs	105
3.7.2 Les cartes auto-organisées de Kohonen (SOM)	106
3.7.3 Les réseaux bayésiens	107
3.8 Discussions	108
3.9 Conclusion	109
3.10 Bibliographie	110

Chapitre IV APPLICATIONS DES RNA DANS LA PRÉDICTION

4.1. Introduction	
4.2. Application des réseaux supervisés dans la prédiction	
4.2.1 Entraînement du réseau RBFG	
4.2.1.1 Phase d'apprentissage	116
4.2.1.2 Phase de prédiction	116
4.2.2 Réseau RBFG entraîné par la BP	117
4.2.3 Réseaux RBFG entraînés par la ROM	118
4.2.3.1 Prédiction du vieillissement thermique de la BORAK22 et du PVC	119
a) En utilisant la technique du FFN pattern	119
b) En utilisant la technique du Batch learning	122
4.2.3.2 Prédiction de la tenue diélectrique des intervalles d'air pointe-barrière-plan	125
4.2.4 RBFG entraîné par la ROM couplée avec les K-Means clustering	126
4.2.4.1 Taux de recouvrement	127
4.2.4.2 Méthode du classement K-means (K-means clustering)	127
4.2.4.3 Application du RBFG entraîné par la ROM couplée à la méthode des K-means clustering pour la prédiction du vieillissement thermique	128
4.3 Application des réseaux non-supervisés dans la prédiction	130
4.3.1 Réseau de neurones SOM et règle d'apprentissage	130
4.3.2 Phase d'apprentissage du réseau de prédiction de Kohonen	131
4.3.3 Prédiction du vieillissement thermique du PVC et de la BORAK22	133
4.3.4 Prédiction de la tenue diélectrique des intervalles d'air pointe-barrière-plan	135
4.4 Discussions	137
4.5 Conclusion	138
4.6 Bibliographie	138
CONCLUSION GENERALE	142
ANNEXE1	145
ANNEXE2	146



LISTE DES SYMBOLES



Symbole	Signification
$b(k)$	moyenne de la séquence aléatoire
c_i	Centre de la $i^{\text{ème}}$ classe de l'espace des entrées
d^p, D_s	Sortie désirée pour la $p^{\text{ème}}$ entrée, vecteur de sorties désirées
d_i	Distance euclidienne d'un exemple par rapport au centre de la classe c_i qu'il représente
$d_i(i, w)$	Distance euclidienne entre le gagnant et le $i^{\text{ème}}$ neurone
E_p	Erreur quadratique à la sortie pour la $p^{\text{ème}}$ entrée
e_i^p	Erreur du $i^{\text{ème}}$ neurone lié à la $p^{\text{ème}}$ entrée
e, E	Erreur instantanée à la sortie du réseau, erreur quadratique
$E_{K\text{-means}}$	Distances euclidiennes quadratiques totales entre les points d'apprentissage assignés à chaque classe et le centre de cette classe
E_w	Somme quadratique des poids
e, e_i	Vecteur entrée et $i^{\text{ème}}$ élément du vecteur e des réseaux à apprentissage non supervisé
F_i^l	Vecteurs des dérivées de l'erreur à la sortie du réseau par rapport à ces vecteurs poids.
F et F^T	Vecteur rassemblant les vecteurs F_i^l et sa transposée.
$G(i, e)$	Fonction gaussienne définie autour du $i^{\text{ème}}$ neurone gagnant du Vecteur entrée e
I	Matrice identité
J	Matrice Jacobienne de l'ensemble des erreurs d'apprentissage
H, H^{MP}	Matrice hessienne, matrice hessienne au point minimum
L	Nombre de couches
L_t	Durée de vie sous contrainte thermique
L_0	Durée de vie sous température ambiante.
n_l	Nombre de neurones dans la $l^{\text{ème}}$ couche
n_{int}	Nombre de nœuds d'interpolation
n	Nombre d'entrées
net	Sortie du neurone après la sommation
$P(C1 x)$	Probabilité à posteriori d'appartenir à la classe C1
$P(C1)$	Probabilité à priori de la classe C1
$p(x C1)$	Densité de probabilité du vecteur x dans la classe C1
$P_m(t)$	Valeur moyenne d'une caractéristique pour le temps t
$P(t)$	Valeur réelle d'une caractéristique pour le temps t
$P_r, P_{rd}(t)$	Profil et profil prédit
$P_d(t)$	Caractéristique réelle prédite
$s(net)$	Sortie du réseau, (s): fonction seuil
T	Température (°C)
$tg\delta$	Facteur de pertes diélectriques
$u_i^l(t)$	Activité du $i^{\text{ème}}$ neurone dans la $l^{\text{ème}}$ couche.
$v(k)$	Variance de la séquence aléatoire
$v(t)$	Variance des données d'apprentissage
v_s, η	Viscosité cinématique et absolue
w^{MP}	Poids minimum de la densité postérieure

w_r	Vecteur des poids du neurone gagnant r .
w_{ij}^l	Poids synaptiques de la $l^{\text{ème}}$ couche reliant les $i^{\text{ème}}$ et $j^{\text{ème}}$ neurones
$w(t+1)$, $\Delta w(t)$ et δ_{wi}	Nouveau poids synaptique, taux d'adaptation du poids $w(t)$, taux d'adaptation partielle du poids $w(t)$.
W^*	Vecteur de poids optimal
W_e , W_s	Matrices de poids à l'entrée et à la sortie du réseau, respectivement
x_i	Entrées du réseau
y_i^l	Sortie du $i^{\text{ème}}$ neurone dans la $l^{\text{ème}}$ couche
θ_i^l	Biais
α	Taux d'apprentissage
α_m	Terme du momentum
α_p , β_p	Paramètres de la fonction objective du réseau bayésien à optimiser
γ	Nombre efficace de paramètres du réseau bayésien
$\delta(k)$	Séquence aléatoire
δ_d	Distance entre deux centres voisins
σ	Écart type (rayon) de chaque gaussienne
$\phi(x)$	Fonction noyau
ξ_i	Nœuds d'interpolation
ϵ_r , ϵ	Permittivité relative et absolue
ϵ_t	Erreur permise à la sortie du réseau
$\rho_j(x_i)$	État (sortie) du neurone caché j lié à la $i^{\text{ème}}$ entrée
ρ_m	Masse volumique de l'huile de transformateur
ρ_R	Résistivité transversale de l'huile de transformateur
τ	Taux de recouvrement des données
η_w , η_c et η_σ	Taux d'apprentissage pour ajuster respectivement les poids, les centres et les rayons gaussiens

INTRODUCTION GENERALE



Avec les nouvelles politiques de régulation des réseaux électriques, la prévention des pannes accidentelles est de plus en plus importante. Les pertes majeures d'une mise hors service accidentelle d'un transformateur de puissance par exemple, se comptent en général par millions de dollars, sans compter les coûts associés à la réparation ou au remplacement de l'équipement [Mye-98]. Cela exige des tests préventifs pour prédire les défauts et un programme de maintenance et de remplacement.

La dégradation des isolants peut diminuer la vie utile des équipements de puissance et provoquer des arrêts de systèmes coûteux. Pour prévenir de telles situations et pour aider les industries et les réseaux indépendants à assurer à coût concurrentiel la maintenance préventive de leurs équipements en charge, des systèmes de surveillance-diagnostic et maintenance doivent être installés.

La maintenance préventive et la maintenance prévisionnelle incluent des tâches telles que la surveillance, le diagnostic, le contrôle, la prédiction. Pour la maintenance préventive du matériel électrique, la détection et le diagnostic de défauts sont réalisés à l'aide de l'interprétation de signatures complexes et la prédiction de propriétés est réalisée à l'aide d'une assez large base de données de mesures.

Le diagnostic et la classification de l'huile de transformateur de puissance, basés sur la connaissance, deviennent populaires; mais l'application des normes exige beaucoup d'expérience. Après la phase de mesures, l'expertise rend la phase de prise de décision critique car il y a presque toujours des conflits entre les données, ainsi que des cas 'd'indécisions'. Par conséquent, le diagnostic de l'huile de transformateur basé sur la connaissance est traité en général comme un art au lieu d'une science et aussi utilisé pour être la référence d'experts hautement habiles [Wan-00].

Les recommandations de la maintenance sont aussi importantes après le diagnostic. En général, la connaissance du type de défaut ou de l'état présent de l'isolation n'est pas suffisante; l'ingénieur de terrain doit avoir en plus, une idée de la durée de vie restante de l'équipement. Si l'estimation peut être faite avec une précision acceptable, il peut commencer à planifier la maintenance de l'isolation et décider des solutions à apporter à l'avance.

Toujours dans le domaine de la maintenance des équipements de puissance, la prédiction des propriétés d'un isolant peut aider considérablement à améliorer la qualité de sa maintenance. Le comportement diélectrique des isolants peut être surveillé par plusieurs mesures de propriétés. Pour prédire l'état de l'isolation, nous avons besoin d'une longue période d'essais pour avoir une base de données assez complète afin de donner une bonne qualité de prédiction.

Pour s'approcher de la qualité de diagnostic des experts, nous avons utilisé plusieurs techniques de réseaux de neurones artificiels (RNA), comparé leurs performances et discuté les problèmes de leurs applications.

Les RNA sont des outils particulièrement adaptés pour aider les spécialistes de maintenance dans les activités de classification, de diagnostic et de prises de décisions, de prédiction, etc. Pour ces raisons, de nombreuses études pilotes sont entreprises dans le secteur de la maintenance pour évaluer les apports des réseaux de neurones avant leur mise en œuvre de façon opérationnelle. Dans notre thèse, nous avons élaboré plusieurs types de RNA appliqués au diagnostic, la classification et la prédiction du vieillissement thermique des isolants de haute tension. L'application des RNA permet de réduire considérablement le temps d'essai de vieillissement au laboratoire en apprenant au

réseau de prédire les propriétés pour des durées plus longues que celles des essais et constituant ainsi, un outil rendant plus économique (du point de vue temps et argent) les essais de haute tension en général.

Ce travail est réparti en quatre chapitres. Dans le premier, une description des méthodes des réseaux de neurones utilisées et une synthèse monographique des travaux publiés dans l'application des RNA en technique de haute tension ont été présentées. La partie réservée aux méthodes des RNA est divisée en deux grands volets: les réseaux de neurones supervisés et les réseaux non supervisés. Dans le volet des réseaux supervisés, les méthodes présentées sont : la Back-Propagation (BP) avec momentum, la BP avec momentum et adaptation du taux d'apprentissage, la méthode de Levenberg-Marquardt, la méthode de la régularisation bayésienne et les réseaux à fonction de base radiale (RBF) [For1-97, Gro-77, Hag-96, Har-94, Jan-97, Kos-92, Lam-01, Nea-96, , Pol-94, Pow-87, Zur-92]. Les réseaux supervisés sont appliqués dans le diagnostic, la classification et la prédiction.

Dans le volet des réseaux non supervisés, les réseaux présentés sont: la couche compétitive et la carte auto-organisatrice de Kohonen (self-organization maps SOM). Ces réseaux sont appliqués dans la classification et la prédiction [Jan-97, Koh1-82, Koh2-01, Kuh-97, Poi-99, Zur-92].

Dans le deuxième chapitre, les mécanismes de vieillissement thermique dans les isolants solides et liquides ont été présentés. Les isolants choisis dans ce travail sont respectivement le PVC utilisé dans les câbles à moyenne tension et l'huile de transformateur à haute tension utilisée par SONELGAZ et nommée BORAK22. L'étude expérimentale de cette partie réalisée totalement par des chercheurs de l'ENP [Bou2-97, Bou3-97, Bou4-99, Ham-97, Nej-91] nous a servi après comme base de données pour l'entraînement des RNA. Les mécanismes de rupture des intervalles d'air pointe-barrière-plan ont été aussi présentés utilisant des résultats expérimentaux pris des travaux de A.Boubakeur [Bou6-79].

Le troisième chapitre synthétise les méthodes les plus utilisées dans le monde dans la surveillance et le diagnostic et présente les différentes applications des réseaux de neurones dans le diagnostic et la classification de la BORAK22. La méthode de diagnostic adoptée dans ce travail est celle des tests physico-chimiques, adoptée par SONELGAZ dans son Laboratoire de Contrôles Mesures et Essais d'El-Hamma (Alger). Les tests physico-chimiques utilisés sont au nombre de six: couleurs, viscosité cinématique, rigidité diélectrique, facteur de pertes diélectriques, acidité et teneur en eau.

L'interprétation des résultats des essais est une opération lourde et la présence d'analystes experts est nécessaire dans l'étape du diagnostic. Pour construire un système de diagnostic robuste et fiable, l'expérience humaine doit être conservée dans un système intelligent pour l'aide à la décision, ce qui offrirait une meilleure maintenance aux transformateurs. Cela nous a poussé à réfléchir à la conception d'un système intelligent pour le diagnostic et la classification de l'huile de transformateur utilisant les réseaux de neurones. Les algorithmes utilisés dans ce but (diagnostic) sont: la BP avec momentum, la BP avec momentum et adaptation du taux d'apprentissage, la méthode de Levenberg-Marquardt et les réseaux de la régularisation bayésienne. Nous présentons en fin de cette partie les réseaux compétitifs, les cartes auto-organisatrices de Kohonen et le réseau de régularisation bayésienne utilisés dans la classification de l'huile de transformateur.

Un des problèmes majeurs dans les applications des réseaux de neurones est leurs capacités de généralisation. Ceci est lié en partie à la taille et à la qualité de la base d'apprentissage. Ce problème est présenté dans le troisième chapitre aussi en changeant la base d'apprentissage dans toutes les méthodes d'apprentissage utilisées.

L'influence des paramètres d'apprentissage du réseau tels que le nombre de couches cachées, le nombre de neurones et leurs distributions, le taux d'apprentissage, le seuil de l'erreur, les fonctions d'activation, etc., sur le déroulement et la qualité de l'apprentissage est présentée aussi dans ce même chapitre pour arriver à un critère d'optimisation des RN : l'exactitude des résultats.

Les réseaux de neurones utilisés dans le quatrième chapitre sont employés pour prédire le vieillissement thermique de la BORAK22 et du PVC utilisé par l'ex-entreprise nationale de l'industrie des câbles ENICAB (actuellement CABEL) dans l'isolation des câbles de Moyenne Tension. Les mêmes RNA sont utilisés pour prédire la tension disruptive des intervalles d'air pointe-barrière-plan. Les résultats des études expérimentales sur le vieillissement thermique du PVC et de la BORAK22, et sur la rigidité des intervalles d'air avec barrières, réalisées par des chercheurs du laboratoire de haute tension de l'ENP.

Les expériences du vieillissement thermique de l'isolation du PVC ont été réalisées par M. Nedjar au laboratoire de haute tension de CABEL [Ncj1-91]. Les expériences sur éprouvettes ont été effectuées sous différentes températures s'étendant de 80°C à 140°C avec un temps de vieillissement maximum de 5000 heures. Les essais consistent à vieillir le matériau et mesurer les propriétés électriques, mécaniques et diélectriques à des intervalles de temps réguliers. Des essais de vieillissement thermique de la BORAK22, qui est employé par SONELGAZ ont été aussi réalisés par A. Hamimi et S. Abdi respectivement pour un temps de vieillissement de 2000 et 5000 heures.

Les données concernant les intervalles d'air pointe-barrière-plan ont été prises des travaux de A. Boubakeur [Bou6-79].

Le choix des RBFG est justifié par leur indépendance du modèle mathématique ainsi que leurs capacités universelles d'approximation, de recouvrement et de généralisation [Khe-94].

Les RBFG utilisés sont entraînés par trois algorithmes qui sont, la Back-Propagation (BP), l'algorithme de l'optimisation aléatoire (ROM) et une méthode originale qui est la ROM couplée avec la méthode des K-means clustering. Pour cette dernière méthode, un réseau supervisé (RBFG) est entraîné par un réseau non supervisé (K-means clustering). Dans l'application de ces trois algorithmes, deux techniques d'adaptation des poids sont utilisées: la technique FFN Pattern et la technique Batch Learning.

Les SOM sont dans la pratique flexibles et faciles à implémenter. Ils sont utilisés dans la prédiction des caractéristiques des mêmes isolants utilisés dans l'application des RBFG cités ci-dessus. Pour rendre la prédiction indépendante de la nature de la caractéristique d'isolation à prédire, nous transformons la base de données d'entrée par soustraire la valeur moyenne de la propriété considérée de la valeur réelle et diviser le résultat par la variance des données.

Alors, au lieu d'utiliser directement les vraies données, nous employons de cette façon les prétendus 'profils' [Oso-97]. On doit noter ici que la prédiction de la caractéristique réelle est séparée du réseau de Kohonen, elle utilise seulement la base de données des profils prédits par les SOM.

«... Comment se fait-il que le monde physique semble obéir de manière très précise à des lois mathématiques ? Non seulement cela, mais les mathématiques qui semblent contrôler notre monde physique sont extraordinairement puissantes et fécondes, en tant simplement que mathématiques. Cela est pour moi un profond mystère : y a-t-il une relation entre le monde du physique et du mental ? Cela nous conduira aussi à considérer le mystère du monde de la culture ». pp 113-114

Professeur Emérite de Mathématiques de L'Université d'Oxford **Roger Penrose**

**THE LARGE, THE SMALL AND THE HUMAN MIND
LES DEUX INFINIS ET L'ESPRIT HUMAIN**

avec la contribution du Professeur **Stephen Hawking**

«... La qualité de la compréhension n'est pas quelque chose qui peut être confinée dans un ensemble de règles. D'ailleurs la compréhension est une qualité qui dépend de notre conscience. La conscience consciente semble venir essentiellement dans le jeu quand la `compréhension` est présente. Ainsi, notre conscience semble être quelque chose impliquant des éléments qui ne peuvent pas être encapsulés dans des règles informatiques de n'importe quel genre : elle est non-algorithmique ».

Roger Penrose

**THE EMPEROR'S NEW MIND
L'ESPRIT, L'ORDINATEUR ET LES LOIS DE LA PHYSIQUE**

- 1.1 Historique
- 1.2 Fondement biologique
- 1.3 Fonctionnement du neurone
 - 1.3.1 Physiologie du neurone
 - 1.3.2 Création d'un potentiel d'action
 - 1.3.3 Le corps cellulaire comme sommateur à seuil
- 1.4 Les neurones artificiels
 - 1.4.1 Le neurone formel
 - 1.4.1.1 Fonction discriminant
 - a) Fonction de base linéaire (LBF).
 - b) Fonction de base radiale RBF (radial basis function)
 - c) Fonction de base elliptique EBF (elliptic basis function)
 - 1.4.1.2 Fonction dynamique linéaire
 - 1.4.1.3 Fonction non dynamique non linéaire
 - 1.4.2 Architecture des réseaux de neurones
 - 1.4.3 Apprentissage des réseaux de neurones
- 1.5 Réseaux de neurones a apprentissage supervise
 - 1.5.1 Réseaux de neurones basés sur la décision
 - 1.5.1.1 Le Perceptron
 - 1.5.1.2 Adaline et Madaline
 - 1.5.1.3 Méthode des moindres carrés (Least Mean Square Error)
 - 1.5.1.4 Méthode de descente de gradient
 - 1.5.2 Réseaux multicouches statiques
 - 1.5.2.1 Back-propagation
 - a) Data adaptive learning (FFN pattern)
 - b) Block adaptive learning (batch learning)
 - 1.5.3 Variantes de la Back-propagation
 - 1.5.3.1 Back-propagation avec Momentum
 - 1.5.3.2 Back-propagation à taux d'apprentissage adaptatif
 - 1.5.3.3 Méthode d'optimisation aléatoire(ROM)
 - 1.5.4 Dilemme Précision-Généralisation

- 1.5.6.2 Architecture et fonctionnement des réseaux RBF
- 1.5.6.3 Apprentissage des réseaux RBF
 - a) Méthode de Centrage Adaptative
(Adaptive Centering Method)
 - b) Méthode du regroupement (clustering algorithm)
- 1.5.6.4 Réseaux RBF et Approximation de fonctions
- 1.6 Réseaux de neurones a apprentissage non supervise
 - 1.6.1 Fondement de l'apprentissage non supervisé
 - 1.6.1.1 Principes généraux
 - a) L'architecture
 - b) Le neurone gagnant
 - 1.6.2 L'apprentissage
 - 1.6.2.1 Réseaux du type 'winner takes all'
 - a) L'algorithme LBG
 - b) Les algorithmes adaptatifs
 - 1.6.2.2 Réseaux du type 'winner takes most'
 - 1.6.3 Réseaux à activation compétitive
 - 1.6.3.1. Principe de fonctionnement d'une couche compétitive
 - 1.6.3.2 Règle d'apprentissage dans une couche compétitive
 - 1.6.4 Carte auto-organisatrice de Kohonen
 - 1.6.4.1 Neurone formel de Kohonen
 - 1.6.4.2 Compétition dans une couche de Kohonen
 - 1.6.4.3. Apprentissage et voisinage dans une couche de Kohonen
 - 1.6.4.4 Caractéristiques du modèle physiologique des SOM
 - 1.7 Monographie des travaux publiés dans l'application des RNA en technique de haute tension
 - 1.8 Conclusion
 - 1.9 Bibliographie

1.1 HISTORIQUE

Le cerveau s'enracine dans le passé très lointain des espèces. Il a évolué en complexité neurologique sur plusieurs millions d'années. Cependant, de façon paradoxale, cet organe fabuleux est resté un mystère pour l'homme. Pourtant, tous les grands esprits depuis des milliers d'années ont élaboré des théories à son sujet.

L'augmentation de puissance dans les ordinateurs ne permet pas de résoudre toujours les problèmes d'une application informatique dans un domaine particulier. L'idée s'est donc installée que le problème n'était peut être pas tant le matériel que le logiciel. La construction de logiciels s'appuie sur plusieurs approches. Deux parmi les plus utilisées sont l'approche algorithmique et l'approche basée sur la connaissance.

Une approche algorithmique nécessite l'écriture (avant la transcription dans un quelconque langage de programmation) du processus à suivre pour résoudre le problème. Les ordinateurs sont des machines complètement logiques (et même binaires) qui suivent à la lettre chacune des instructions du programme. C'est un avantage lorsque tous les cas ont été prévus à l'avance par l'algorithmicien.

La seconde approche possible est celle de l'intelligence artificielle (IA), avec pour applications les plus connues les systèmes experts. Ici, la résolution du problème est confiée à un ensemble de règles données par l'expert humain du domaine. Il n'en demeure pas moins que toutes les règles doivent avoir été exprimées préalablement au traitement, et que le programme demeure binaire dans son exécution. Les cas qui n'ont pas été prévus par l'expert ne seront pas correctement traités. L'introduction de la logique floue ne change pas la nature des limitations d'emploi du programme : l'exécution reste totalement déterministe.

En fait, l'approche basée sur la connaissance se limite à des domaines d'application où la modélisation de la connaissance, par exemple sous forme de règles, est possible. Ces domaines sont souvent ceux des sciences dites 'exactes' comme l'électronique, la physique, etc., par opposition aux sciences dites 'humaines' comme la médecine, la psychologie, etc., où la connaissance est plus empirique. L'IA se révèle donc être un moyen commode de stocker de la connaissance sous forme explicite [Tou-92].

Ces deux approches ne suffisent pas à répondre à tous les problèmes existants. Citons les domaines de la reconnaissance de formes (images ou signaux), du diagnostic, du contrôle, de la traduction automatique, de la compréhension du langage, depuis longtemps explorés à l'aide des approches algorithmiques et à base de connaissances, qui n'ont pas rencontré le succès escompté.

Une troisième approche au traitement de l'information semble donc s'offrir à nous, où l'on cherche à s'inspirer du traitement de l'information effectué par le cerveau.

L'hypothèse principale est que le comportement intelligent est sous-tendu par un ensemble de mécanismes mentaux. L'hypothèse proposée par de nombreux biologistes est: pour recréer le comportement intelligent du cerveau, il faut s'appuyer sur son architecture, en fait, tenter de l'imiter.

Les réseaux de neurones artificiels sont dotés de propriétés qui, comme l'apprentissage à partir d'exemples, semblent prometteuses dans certains domaines d'applications. Il est d'usage de citer la reconnaissance de formes, le diagnostic, etc. En fait, est considéré comme éligible, tout problème qui se représente sous la forme d'une fonction de mise en correspondance entre un espace d'entrée et un espace de sortie, dès lors que l'on ne dispose que d'exemples de comportement de cette fonction.

Les premiers pas:

- 1890 : W. James, célèbre psychologue américain introduit le concept de mémoire associative, et propose ce qui deviendra une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones connue plus tard sous le nom de loi de Hebb [Heb-49].
- 1943 : J. Mc Culloch et W. Pitts laissent leurs noms à une modélisation du neurone biologique (un neurone au comportement binaire). Ce sont les premiers à montrer que des réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes (tout au moins au niveau théorique) [Cul-43].
- 1949 : D. Hebb, physiologiste américain explique le conditionnement chez l'animal par les propriétés des neurones eux-mêmes. Ainsi, un conditionnement de type pavlovien, tel que nourrir tous les jours à la même heure un chien, entraîne chez cet animal la sécrétion de salive à cette heure précise même en l'absence de nourriture. La loi de modification des propriétés des connexions entre neurones qu'il propose, explique en partie ce type de résultats expérimentaux [Heb-49].

Les premiers succès :

- 1957 : F. Rosenblatt développe le modèle du Perceptron. Il construit le premier neuro-ordinateur basé sur ce modèle et l'applique au domaine de la reconnaissance de formes. Notons qu'à cette époque les moyens à sa disposition sont limités et c'est une prouesse technologique de réussir à faire fonctionner correctement cette machine pendant quelques minutes [Ros-58].
- 1960 : B. Widrow, un automaticien, développe le modèle Adaline (Adaptative Linear Element). Dans sa structure, le modèle ressemble au Perceptron. Cependant la loi d'apprentissage est différente. Celle-ci est à l'origine de l'algorithme de rétro-propagation de gradient très utilisé aujourd'hui avec les Perceptrons multicouches. Les réseaux de type Adaline restent utilisés de nos jours pour certaines applications particulières. B. Widrow a créé dès cette époque une des premières firmes proposant neuro-ordinateurs et neuro-composants, la "Memistor Corporation" [Wid-60].
- 1969 : M. Minsky et S. Papert publient un ouvrage qui met en exergue les limitations théoriques du perceptron. Limitations alors connues, notamment concernant l'impossibilité de traiter par ce modèle des problèmes non linéaires. Ils étendent ces limitations à tous les modèles de réseaux de neurones artificiels. Après, les recherches dans le domaine furent abandonnées par manque de financement et les chercheurs se tournèrent principalement vers les systèmes à bases de règles [Min-69].

L'ombre :

- 1967-1982 : Toutes les recherches ne sont, bien sûr, pas interrompues. Elles se poursuivent, mais sous le couvert de divers domaines comme : le traitement adaptatif du signal, la reconnaissance de formes, la modélisation en neurobiologie, etc. De grands noms travaillent durant cette période tels: S. Grosberg, T. Kohonen, ... [Gro-77, Koh1-82]

Le renouveau :

- 1982 : J. J. Hopfield est un physicien reconnu, à qui l'on doit le renouveau d'intérêt pour les réseaux de neurones artificiels [Hop-82].

La levée des limitations :

- 1983 : La Machine de Boltzmann est le premier modèle connu apte à traiter de manière satisfaisante les limitations recensées dans le cas du perceptron. Mais l'utilisation pratique s'avère difficile, la convergence de l'algorithme étant extrêmement lente (les temps de calcul sont considérables) [Hin-83, Rum1-86].

- 1985 : La rétro-propagation de gradient apparaît. C'est un algorithme d'apprentissage adapté aux réseaux de neurones multicouches (aussi appelés perceptrons multicouches). Sa découverte réalisée par trois groupes de chercheurs indépendants indique que "la chose était dans l'air". Dès cette découverte, nous avons la possibilité de réaliser une fonction non linéaire d'entrée/sortie sur un réseau en décomposant cette fonction en une suite d'étapes linéairement séparables. De nos jours, les réseaux multicouches et la rétro-propagation de gradient reste le modèle le plus étudié et le plus productif au niveau des applications [Tou-92].

Actuellement:

Au niveau commercial, la figure 1.1 montre que plus de 200 compagnies sont aujourd'hui impliquées dans des développements d'applications connexionnistes [Tou-92].

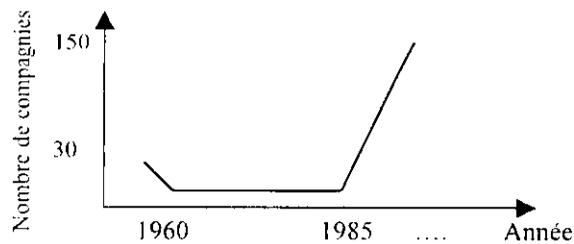


Fig.1.1 Evolution du nombre de compagnies proposant des produits connexionnistes

Les prévisions du marché se chiffrent déjà en dizaines de millions de dollars ; elles ont dépassé les 100 millions de dollars dès 1992 [Tou-92].

En résumé, l'approche connexionniste permet les avantages suivants :

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

Ce furent ces avantages qui poussèrent les chercheurs, dans les années 80, à effectuer une ré-évaluation de l'approche connexionniste. Le regain d'intérêt aux réseaux de neurones reprit alors.

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

1.2 FONDEMENT BIOLOGIQUE

1.2.1 Le neurone

Les cellules nerveuses, appelées neurones, sont les éléments de base du système nerveux central. Celui-ci en possède environ cent milliards. Les neurones possèdent de nombreux points communs dans leur organisation générale et leur système biologique, avec les autres cellules. Ils présentent cependant des caractéristiques qui leur sont propres et se trouvent dans cinq fonctions spécialisées qu'ils assurent (Har-94) :

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

1.2.2 Structure des neurones

Un neurone est constitué de trois parties :

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

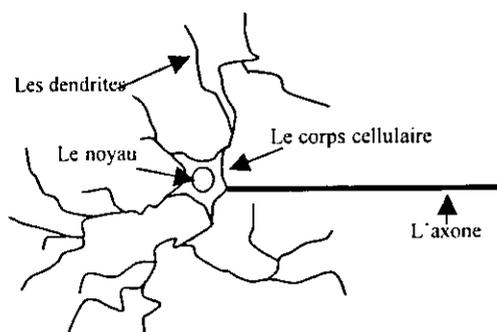
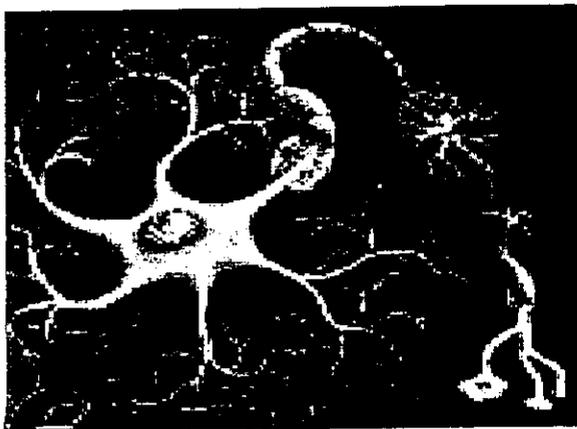


Fig.1.2 Le neurone biologique

a. Le corps cellulaire

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

b. Les dendrites

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

c. L'axone

L'axone, qui est la fibre nerveuse, sert de moyen de transport pour les signaux émis par le neurone. Il se distingue des dendrites par sa forme et par les propriétés de sa membrane externe. En effet, il est généralement plus long que les dendrites, et se ramifie à l'extrémité, là où il communique avec d'autres neurones, alors que les ramifications des dendrites se produisent plutôt près du corps cellulaire.

Pour former le système nerveux, les neurones sont connectés les uns aux autres suivant des répartitions spatiales complexes. Les connexions entre deux neurones se font en des endroits appelés synapses où ils sont séparés par des espaces synaptiques.

1.3 FONCTIONNEMENT DU NEURONE

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

1.3.1 Physiologie du neurone

La physiologie du neurone est liée aux propriétés de la membrane nerveuse et au métabolisme de la cellule. La différence de potentiel mesurée entre le milieu intérieur de la cellule et le milieu extérieur est de -60 mV. Pour maintenir une telle différence de potentiel, la cellule fait appel à des pompes ioniques (Na^+ , K^+ , ..). Cependant, une faible dépolarisation de la membrane entraîne une certaine perméabilité aux ions sodium (Na^+) dont l'effet peut être catastrophique au niveau cellulaire. En effet, à partir d'une certaine valeur seuil de dépolarisation de la membrane, il y a rupture des équilibres ioniques et création d'un potentiel d'action ("spike" en anglais, Fig. 1.3) [Tou-92, Har-94].

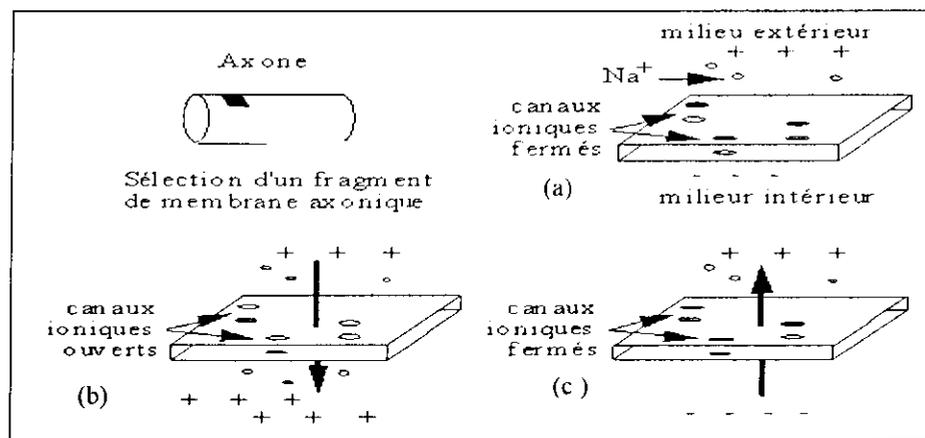


Fig.1.4 Passage d'un potentiel d'action au niveau de la membrane de l'axone
 a) Equilibre ionique (au repos).
 b) Arrivée d'un potentiel d'action (dépolariation).
 c) Après la dépolariation : L'hyper polarisation.

Toute cette série d'évènements au niveau cellulaire n'aura duré que 5 à 10 milli-secondes. Durant la phase d'hyper polarisation, le neurone est très difficilement excitable. Ce qui s'explique par le fait que la différence de potentiel par rapport à la valeur seuil (S) est plus importante que celle au repos [Tou-92, Har-94].

1.3.2 Création d'un potentiel d'action

En fait, à l'arrivée d'un potentiel d'action sur une synapse, un neuromédiateur est libéré dans l'espace synaptique. Il va ouvrir des canaux ioniques sur la membrane post-synaptique, créant ainsi une dépolarisation (aussi appelée potentiel évoqué) qui s'étend jusqu'à l'axone (Fig.1.5).

Les dépolarisations unitaires sont sommées dans l'espace (toutes les synapses du neurone) et dans le temps (sur une période de quelques millisecondes) et génèrent, éventuellement, un potentiel d'action sur le neurone post-synaptique. Ainsi que le montre la figure (1.6), la génération d'un potentiel d'action est le fruit de nombreuses dépolarisations et l'action d'une seule synapse est pratiquement sans effet [Yed-98].

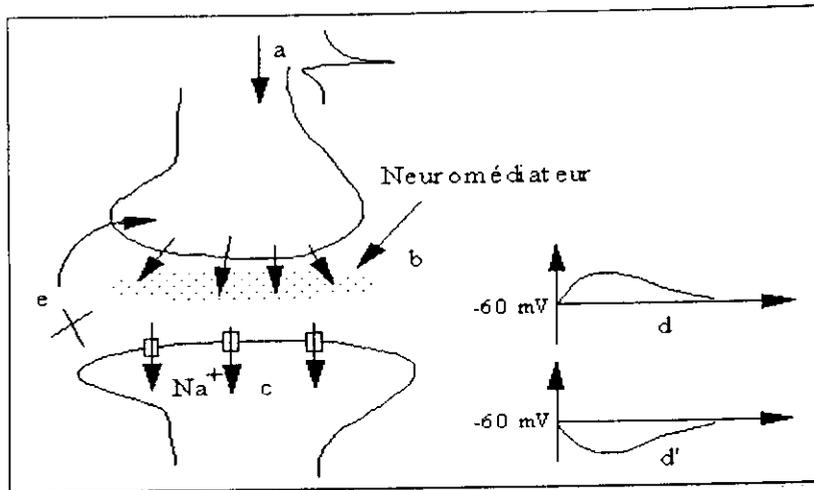


Fig.1.5 Fonctionnement au niveau synaptique
 a) Arrivée d'un potentiel d'action.
 b) Libération du neuromédiateur dans l'espace synaptique.
 c) Ouvertures des canaux ioniques dues au neuromédiateur.
 d) Génération d'un potentiel évoqué excitateur.
 d') Génération d'un potentiel évoqué inhibiteur.
 e) Les synapses inhibitrices empêchent la génération de potentiel d'action.

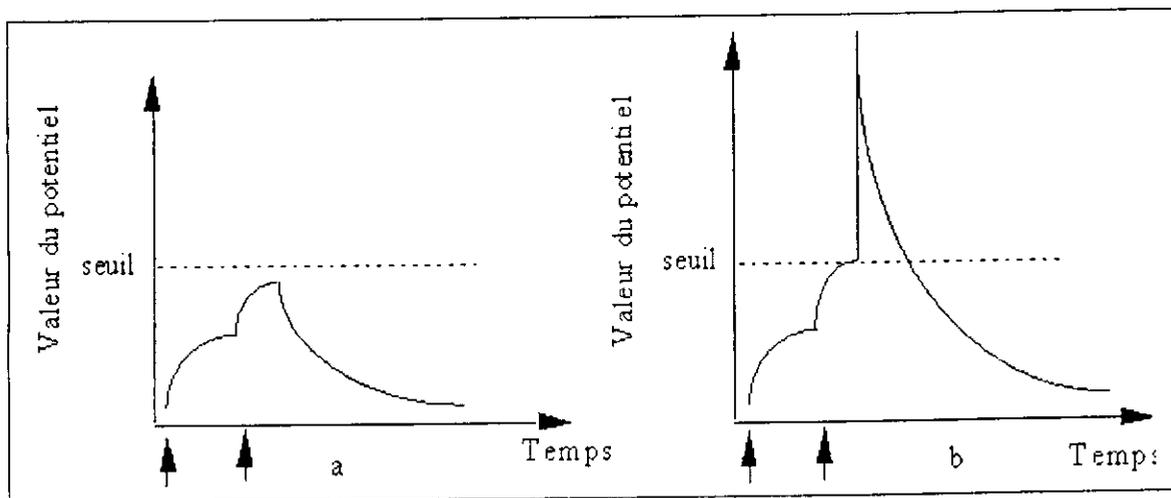


Fig.1.6 Sommation spatio-temporelle :Addition des potentiels évoqués dans l'espace et dans le temps.
 a) 2 potentiels évoqués (reperés par les flèches) ne dépassant pas la valeur seuil.
 b) 2 potentiels évoqués dépassant la valeur seuil, génèrent un potentiel d'action.

1.3.3 Le corps cellulaire comme sommateur à seuil

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

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

L'influx nerveux qui se propage entre différents neurones est, au niveau de ces neurones, un phénomène électrique [Zur-92, Har-94].

1.4 LES NEURONES ARTIFICIELS

1.4.1 LE NEURONE FORMEL

Le premier neurone formel à été modélisé par W. Mc Culloch et W. Pitts, en 1943 [Cul-43, Clé-90]. 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 (Fig.1.7).

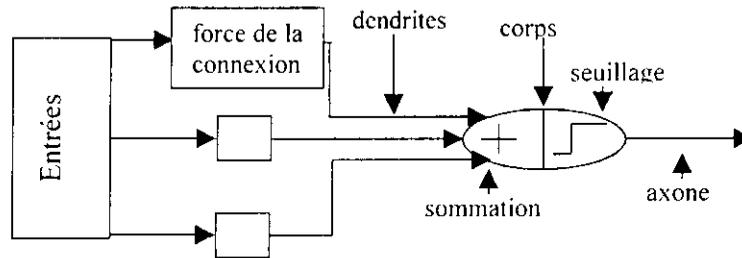


Fig.1.7 Le modèle de Mc Culloch et Pitts

1.4.1.1 Fonction discriminant

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

a) Fonction de base linéaire (LBF).

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

$$y_i^l(t) = f(u_i^l(t)) \quad \text{et} \quad u_i^l(t) = \sum_{j=1}^{n_{l-1}} w_{ij}^l(t) y_j^{l-1}(t) + \theta_i^l(t) \quad (I.1)$$

f est une fonction d'activation, $u_i^l(t)$ représente l'activité du $i^{\text{ème}}$ neurone dans la $l^{\text{ème}}$ couche.

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

l le nombre de couches et n_l le nombre de neurones dans la $l^{\text{ème}}$ couche.

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

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

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

L'apprentissage de ce réseau consiste à minimiser un critère quadratique donnant l'écart E_p entre la sortie du réseau y_i^j et la sortie désirée d^p . Il est défini par :

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

e_i^p est l'erreur du $i^{\text{ème}}$ neurone lié à la $p^{\text{ème}}$ entrée.

b) Fonction de base radiale RBF (radial basis function)

Sa forme générale est :

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

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

c) Fonction de base elliptique EBF (elliptic basis function)

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

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

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

1.4.1.2 Fonction dynamique linéaire

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

$$a_0 u_i(t) + a_1 \dot{u}_i(t) = x_i(t) \quad (I.5)$$

u_i est l'activation du $i^{\text{ème}}$ neurone et x_i est l'entrée du système dynamique décrit par l'équation (I.5) et a_i sont des paramètres.

1.4.1.3 Fonction non dynamique non linéaire

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

Une autre fonction qui est très utilisée malgré qu'elle ne soit pas monotone, est la fonction gaussienne dont le grand avantage est son comportement local (Fig.1.8).

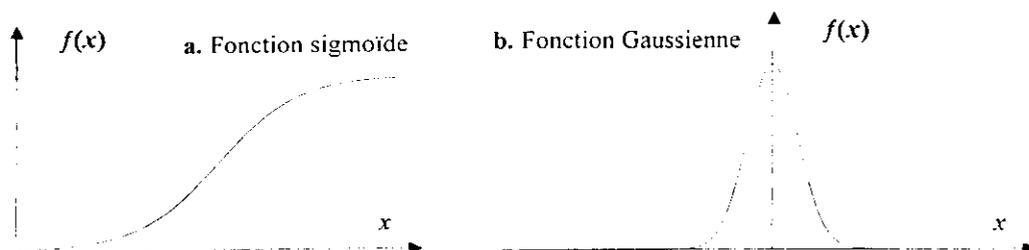


Fig.1.8 Allure de quelques fonctions d'activation

1.4.2 ARCHITECTURE DES RÉSEAUX DE NEURONES

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

1.4.2.1 Réseaux statiques

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

1.4.2.2 Réseaux dynamiques

Contrairement aux précédents, ces réseaux contiennent des retours d'informations (Fig. 1.9). Cet aspect leur confère la possibilité de commander et d'identifier les systèmes dynamiques.

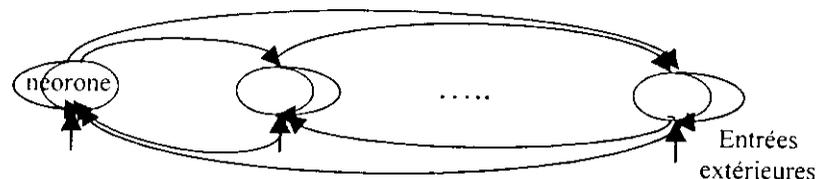


Fig.1.9 Architecture générale d'un réseau dynamique

1.4.2.3 Réseaux auto-organisés

Les réseaux de neurones auto-organisés sont des réseaux qui changent leurs structures internes pendant l'apprentissage. Ainsi les neurones se regroupent topologiquement suivant la représentation des exemples. Ces réseaux sont des dérivées des modèles de Kohonen [Koh1-82].

1.4.3 APPRENTISSAGE ET ARCHITECTURE DES RÉSEAUX DE NEURONES

En général, l'apprentissage désigne tout processus susceptible de modifier un comportement ultérieur. Pour un réseau, il lui permet de s'adapter à des situations désirées tout en modifiant ses poids internes au fur et à mesure qu'on lui présente de nouveaux exemples, et plus il y a d'associations reliant les éléments nouveaux à ce qui est déjà connu, meilleur est l'apprentissage. On peut distinguer deux types d'apprentissage (Fig. 1.10)

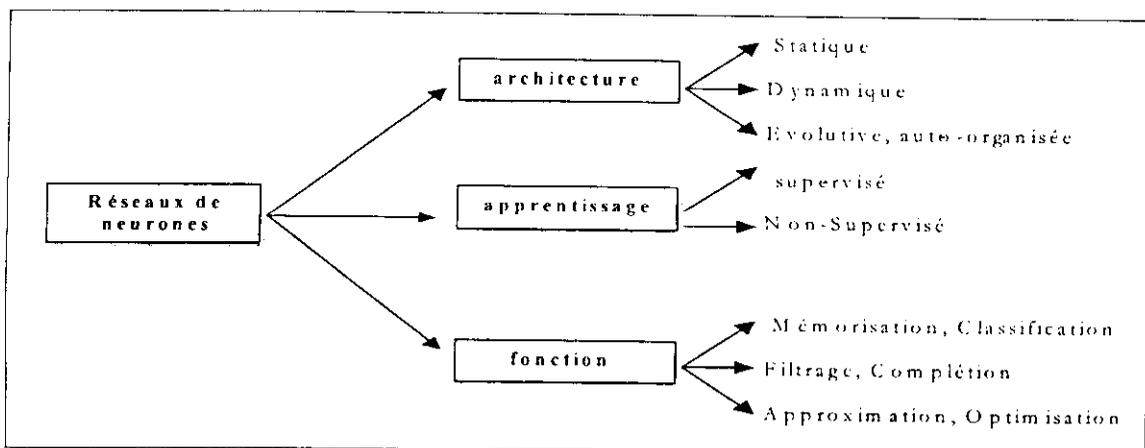


Fig. 1.10 Les différentes possibilités de classification des réseaux de neurones

1.4.3.1 Apprentissage supervisé

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

1.4.3.2 Apprentissage non supervisé

L'apprentissage non supervisé nécessite la présence des entrées seulement. Il est basé seulement sur l'information locale existant au niveau des neurones.

1.5 RESEAUX DE NEURONES A APPRENTISSAGE SUPERVISE

L'intérêt de cet apprentissage est de trouver des poids synaptiques optimaux afin de ramener au réseau un comportement désiré à l'aide d'échantillons aléatoires.

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

1.5.1 RÉSEAUX DE NEURONES BASÉS SUR LA DÉCISION

1.5.1.1 Le Perceptron

Introduit pour la première fois par Rosenblat en 1958 [Ros-58], c'est un réseau statique capable de modifier ses poids en fonction des exemples que l'on veut lui faire apprendre. L'information circule dans un sens unique de la couche d'entrée vers la couche de sortie [Hol-97] (Fig.1.11).

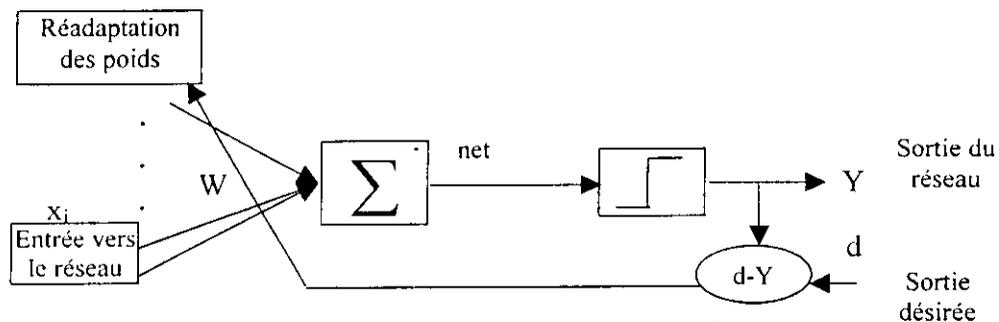


Fig.1.11 Structure du Perceptron linéaire et son entraînement

Tous les réseaux qui traitent de la décision sont basés sur le perceptron. Dans le cas de perceptron linéaire, l'activité du neurone est définie par l'équation linéaire suivante [Har-94, Zur-92, Jan-97, Clé-90] :

$$net = \sum_{j=1}^n x_j w_{ij} + \theta_i \quad i=1 \dots n: \text{nombre d'entrées} \quad (I.6)$$

les x_j représentent les entrées, w_{ij} le poids synaptique en provenance du $i^{\text{ème}}$ neurone de la couche d'entrée vers le $j^{\text{ème}}$ neurone de sortie et θ_i un paramètre qui agit sur le seuil à la sortie que nous avons appelé biais.

La sortie du réseau (perceptron) $s(net) = y$ est définie par la fonction seuil :

$$y = s(net) = \begin{cases} 1 & net_i > 0 \\ 0 & net_i \leq 0 \end{cases} \quad (I.7)$$

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

$$w(t+1) = w(t) + \Delta w(t) \tag{I.8}$$

$w(t+1)$ est la nouvelle valeur adaptée du poids.

Dans le cas de perceptron linéaire on a :

$$\Delta w(t) = \alpha e(t) \cdot x(t) \tag{I.9}$$

$$= \alpha (d(t) - y(t)) \cdot x(t) \tag{I.10}$$

α est le taux d'apprentissage.

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

L'entraînement du réseau se poursuit jusqu'à ce que chaque exemple soit correctement classé. L'inconvénient du perceptron est qu'il n'est efficace que pour les problèmes linéairement séparables. Dans le cas contraire, la combinaison de plusieurs perceptrons linéaires peut être utilisée pour former des zones de décision complexes.

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

1.5.1.2 Adaline et Madaline

Introduit par Widrow et M.Hoff en 1960 [Wid-60], le modèle d'Adaline (Adaptive Linear Element) a la même architecture que le perceptron. Une couche d'entrée est connectée à une ou plusieurs sorties. Dans ce dernier cas, il est appelé Madaline (Multiple Adaptive Linear Elements). La différence entre ce réseau et le perceptron, est qu'il est destiné à l'approximation des fonctions. Sa sortie est linéaire (Fig. 1.12).

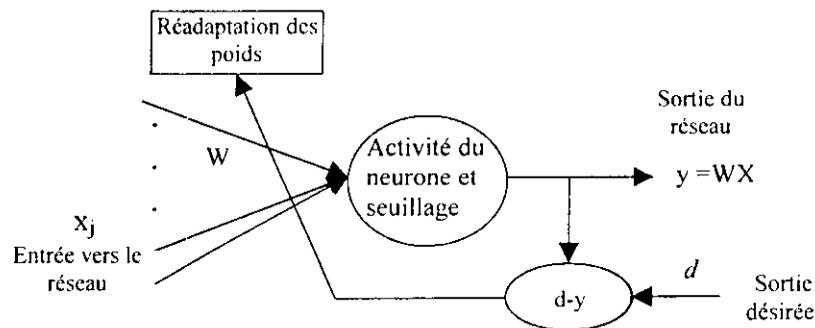


Fig.1.12 Structure d'un réseau Adaline et son principe d'apprentissage

Une fois l'architecture d'un réseau de neurones choisie, il est nécessaire d'effectuer un apprentissage pour déterminer les valeurs des poids permettant à la sortie du réseau de neurones d'être aussi proche que possible de l'objectif fixé. Dans le cas d'un problème de classification supervisée, il s'agit de déterminer une surface de séparation.

L'apprentissage du Madaline s'effectue grâce à la minimisation d'une fonction, appelée fonction coût, calculée à partir des exemples de la base d'apprentissage et de la sortie du réseau de neurones. Cette fonction détermine l'objectif à atteindre. La présentation des entrées x_i produit une sortie correspondante y_j comme suit :

$$y_j = \sum_{i=1}^n x_i w_{ij} + \theta_j \quad j=1 \dots m : \text{nombre de sorties} \quad (I.11)$$

θ_j est le biais de la j^{ème} sortie.

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

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

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

1.5.1.3 Méthode des moindres carrés (Least Mean Square Error)

La méthode des moindres carrés consiste à minimiser l'erreur quadratique définie par [Zur-92]:

$$(e)^2 = (D_s - W^T \cdot X)^2 \quad (I.12)$$

D_s est le vecteur de sorties désirées, W^T est la matrice transposée de la matrice des poids W et X est le vecteur des entrées.

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

$$\frac{\partial}{\partial W} E[(e)^2] = 0 \quad (I.13)$$

Or

$$E[(e)^2] = E[D_s^2] - 2E[D_s X] W^T + WE[X^T X] W^T \quad (I.14)$$

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

$$R_{aut} = E[X^T X] \quad (I.15)$$

$$P_{int} = E[D_s X] \quad (I.16)$$

on obtient :

$$\frac{\partial}{\partial W} E[(e)^2] = -2P_{int} + 2WR_{aut} = 0 \quad (I.17)$$

D'où l'on tire le vecteur de poids optimal par:

$$W^* = P_{int} R_{aut}^{-1} \quad (I.18)$$

1.5.1.4 Méthode de descente de gradient

Cette méthode est destinée à déterminer les adaptations qu'il faut apporter aux poids synaptiques à chaque présentation d'un exemple. On calcule le gradient de la fonction objectif par rapport au vecteur poids W afin d'évaluer la variation de l'erreur. Pour la faire diminuer, il suffit de parcourir la surface d'erreur dans la direction opposée du gradient.

B. Widrow a adopté pour cela la méthode Delta [Har-94, Zur-92], dans laquelle le gradient est estimé en fonction des poids synaptiques; et cela se fait à chaque fois qu'on présente un vecteur à l'entrée. Dans ce cas, l'adaptation des poids fait tendre le gradient

vers un minimum instantané. L'erreur quadratique instantanée est donnée par [Wer-90, Jan-97, Clé-90, Zur-92, Har-94] :

$$\zeta_t = \frac{1}{2}(e)^2 = \frac{1}{2}(d - y)^2 \quad (I.19)$$

Afin de progresser dans le sens opposé du gradient, le taux d'adaptation des poids est donné par:

$$\Delta W = -\alpha \frac{\partial \zeta_t}{\partial W} = e \cdot X \quad (I.20)$$

et les poids adaptés ($w_{ij}(t+1)$) sont alors donnés par :

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(d_i - y_i)x_i \quad (I.21)$$

α est le taux d'apprentissage dont le rôle est de réguler la vitesse de convergence.

Ce réseau étant un approximateur linéaire, l'augmentation du nombre de couches n'apporte rien de nouveau sur l'amélioration de ses performances. Pour le rendre sensible à cette augmentation, on doit introduire des fonctions non linéaires.

La première condition s'assure que le pas choisi n'est pas trop grand (sinon l'algorithme risque d'avoir un comportement oscillatoire), alors que la deuxième s'assure qu'il n'est pas trop petit (sinon l'algorithme a une convergence très lente). Les valeurs habituelles de pas sont comprises entre 0,1 et 0,7 [Yed-98].

1.5.2 RÉSEAUX MULTICOUCHES

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

La topologie de ce type de réseaux est telle qu'avant d'arriver à une sortie, chaque unité d'entrée est suivie par une succession de couches cachées et chaque unité d'une couche donnée reçoit son entrée à partir de la couche précédente seulement (Fig.1.13).

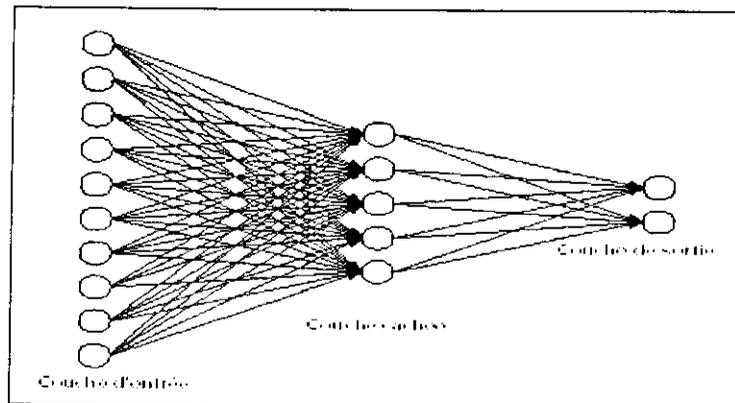


Fig. 1.13. Réseau multicouche classique

1.5.2.1 Back-propagation

Les limitations des modèles neuronaux des années 60 comme l'*Adaline* ou le *Perceptron*, 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. Mis en œuvre en 1972 par Werbos puis Rumelhart [Rum2-90], Hinton et Williams l'ont adapté aux réseaux multicouches dont l'architecture est donnée dans la figure 1.14 [Wer-90].

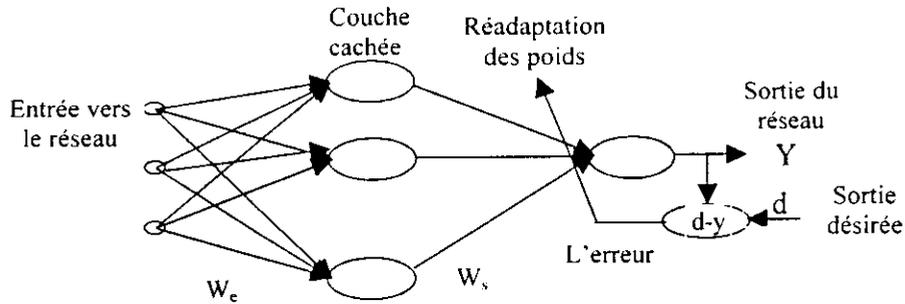


Fig.I.14 Principe de l'entraînement du réseau par rétro-propagation de l'erreur

W_c est la matrice de poids synaptiques liant l'entrée avec la couche cachée, et W_s la matrice de poids synaptique liant la couche cachée avec la sortie.

a) Data adaptive learning (FFN (Feed-Forward Network) pattern)

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

b) Block adaptive learning (batch learning)

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

L'algorithme de base de la Back-Propagation est celui de la descente du gradient. Contrairement au Madaline, dans l'algorithme de la BP avec descente de gradient, les poids directement liés à la sortie ont un lien direct avec l'erreur à minimiser. Ceci introduit des calculs supplémentaires à effectuer. Au début, il faut initialiser les poids synaptiques à de petites valeurs comprises entre -1 et 1.

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

$$E_p = \sum_{i=1}^{n_l} (d_i^l - y_i^l)^2 \tag{I.22}$$

3. Pour le $p^{ème}$ exemple, calculer les valeurs des taux de réadaptation des poids liant les neurones i et j de la couche l par:

$$\left[\Delta w_{ij}^l \right]_p = \alpha \cdot \delta_{wi}^l \cdot w_i^l \cdot f^l(x_i^l(t)) \tag{I.23}$$

α :taux d'apprentissage et δ_{wi}^l est le taux de réadaptation partielle des poids.

$$\delta_{wi}^l = (d_i^l - y_i^l) \text{ notée } \delta_{wsi}^l \text{ pour les neurones de la couche de sortie.}$$

$$= \sum_{i=l}^{n_{l+1}} \delta_{wsi}^l \cdot w_{ij}^l \text{ pour les neurones des autres couches} \tag{I.24}$$

- 4.-S'il s'agit d'un data Learning(FFN Patern), réadapter les poids :

$$\bullet \text{poser : } \Delta w_{ij}^l(t) = [\Delta w_{ij}^l(t)]_p \quad (I.25)$$

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

• si $p \neq M$ (M étant le nombre d'exemples d'entraînement), retourner à 1.

$$\bullet \text{sinon : } \Delta w_{ij}^l(t) = \sum_{p=1}^{p=M} [\Delta w_{ij}^l(t)]_p \quad (I.26)$$

5. réadapter les poids

$$w_{ij}^l(t+1) = w_{ij}^l(t) + \Delta w_{ij}^l(t) \quad (I.27)$$

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

Comme le montre l'algorithme ci-dessus, on peut opérer avec cette méthode de deux manières différentes : data adaptive learning, ou block adaptive learning, que nous avons présenté avant [Jan-97, Yed-98].

1.5.3 VARIANTES DE LA BACK-PROPAGATION

1.5.3.1 Back-propagation avec Momentum

C'est une méthode importante et efficace. Actuellement, la back-propagation est pratiquement toujours utilisée avec le Momentum. D.E. Rumelhart 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 [Har-94, Zur-92].

L'équation d'adaptation devient donc :

$$w_{ij}^l(t+1) = w_{ij}^l(t) + \alpha \cdot \Delta w_{ij}^l(t) + \alpha_m \cdot \Delta w_{ij}^l(t-1) \quad (I.28)$$

Le terme ajouté à l'équation (I.27) est appelé Momentum (élan, quantité de mouvement), en analogie avec la mécanique classique, où un objet en mouvement garde l'impulsion acquise grâce à la quantité de mouvement qui lui a été communiquée précédemment pour accélérer son mouvement.

Le paramètre α_m est utilisé pour pondérer l'effet de ce terme. Sa valeur est généralement prise entre 0.1 et 0.8. Des recherches plus poussées ont montré que l'adaptation de sa valeur, comme c'est le cas pour le taux d'apprentissage, donne des résultats très intéressants. L'utilisation de cette méthode permet de faire sortir les poids des minima locaux, afin de chercher d'autres optimums, ce qui donne beaucoup de chances d'aboutir à un minimum global. Le Momentum tend aussi à éviter les grands « sauts » pendant l'apprentissage, générés par le changement de la pente dans la surface d'erreur [Jac-88]. Notre utilisation de cette méthode nous a permis de constater qu'il est intéressant de l'appliquer avec un taux d'apprentissage dynamique.

1.5.3.2 Back-propagation à taux d'apprentissage adaptatif

La version théorique de la rétro-propagation réclame une variation infinitésimale (microscopique) des valeurs des poids à chaque itération. Cette variation est contrôlée par le taux d'apprentissage α . Conventionnellement ce facteur doit être petit et assurant la convergence d'une manière très lente. Un facteur plus grand génère des changements

plus importants dans les poids synaptiques, pouvant accélérer ainsi l'apprentissage. Malheureusement, ce n'est, le plus souvent, pas le cas. En effet, avec une valeur importante de ce facteur, la surface d'erreur est parcourue d'une manière « téméraire », engendrant ainsi des oscillations et une instabilité dans la recherche du minimum, ce qui compromet la convergence. De ce fait, le choix de la valeur de ce paramètre est d'une importance capitale dans l'algorithme d'apprentissage de la BP.

Une technique très simple est utilisée afin d'augmenter la rapidité de convergence, elle consiste à mettre en œuvre un algorithme qui a pour rôle de contrôler l'erreur d'entraînement à chaque étape comme donné dans la figure 1.15. [Hag-96].

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

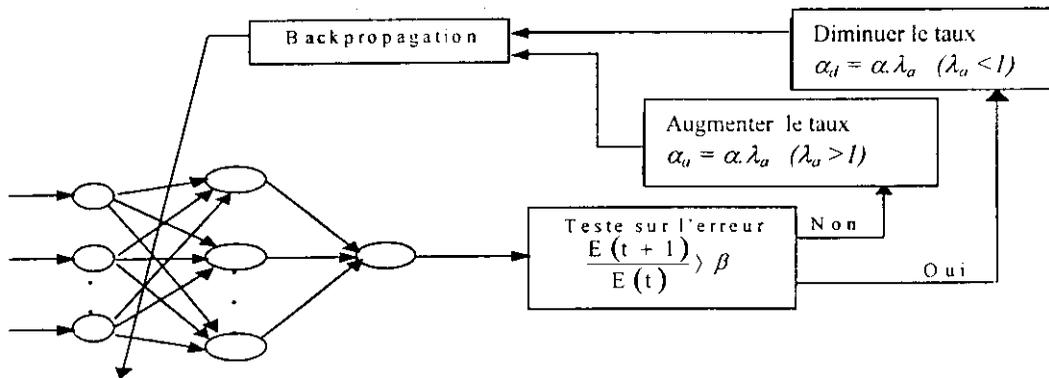


Fig. 1.15 Schéma d'adaptation du taux d'apprentissage

1.5.3.3 Méthode d'optimisation aléatoire (ROM)

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

En 1965, J.Mathias a proposé une méthode aléatoire d'optimisation qui a été développée ensuite par F.J.Sollis et J.B.Wetts en 1981 [Sol-81]. Ces derniers ont montré la convergence du réseau vers un minimum global. Nous présentons ci-dessous l'algorithme de cette méthode.

Algorithme de la ROM

Cette méthode utilise un bruit gaussien pour générer les séquences aléatoires nécessaires pour la réadaptation des poids du réseau suivant les étapes [Yed-98, Sol-81] :

1. Initialiser tous les poids du réseau aléatoirement. Initialiser la variance ν et la moyenne de la séquence $b(0)=0$.
2. Générer une nouvelle séquence aléatoire $\delta(k)$ de moyenne $b(k)$ et de variance $\nu(k)$.
3. Calculer les erreurs pour tous les exemples $E = \sum_{p=1}^M E_p$ à la sortie du réseau pour chaque cas :

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

4. - Si $E(w(k)+\delta(k)) < E(w(k))$, alors

$$\begin{aligned} w(k+1) &= w(k) + \delta(k) \\ b(k+1) &= 0.4\delta(k) + 0.2b(k) \end{aligned} \tag{I.29}$$

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

$$\begin{aligned} w(k+1) &= w(k) - \delta(k) \text{ et} \\ b(k+1) &= b(k) - 0.4\delta(k) \end{aligned} \tag{I.30}$$

- Sinon,

$$\begin{aligned} w(k+1) &= w(k) \\ b(k+1) &= 0.5b(k) \end{aligned} \tag{I.31}$$

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

Le grand avantage de cette méthode est qu'elle est indépendante du critère d'optimisation, contrairement à la méthode de back-propagation où le calcul du Jacobien est nécessaire. L'utilisation de la ROM dépend de la variance du vecteur Gaussien. Une séquence avec une grande variance génère de grands changements dans l'erreur mais jamais des oscillations. Le cas échéant, elle marque des stagnations sur la surface d'erreur jusqu'à l'arrivée d'une séquence favorable. Il est intéressant de commencer l'entraînement avec une variance importante et de la faire baisser au fur et à mesure que l'entraînement avance [Yed-98, Sol-81].

1.5.4 Dilemme Précision-Généralisation

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

Plusieurs travaux ont été menés sur cette question. E.B.Baum et Haussler ont proposé une technique où le nombre de neurones dépend du nombre d'entrées n_x , du nombre de sorties n_y , et de celui d'exemples n_p [Bau-86].

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

$$\frac{n_y n_p}{1 + \log_2(n_p)} \leq n_w \leq n_y \left(\frac{n_p}{n_x} + 1 \right) (n_x + n_y + n_p) + n_y \tag{I.32}$$

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

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

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

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

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

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

1.5.5 RESEAUX BAYESIENS

Dans l'analyse bayésienne de données, toutes les quantités incertaines sont modelées comme des distributions de probabilités. Pour les applications dans les réseaux de neurones, l'approche bayésienne a commencé grâce à MacKay et Neal [Mac3-92, Nea-96]. La difficulté majeure dans la construction du modèle est le contrôle de la complexité de ce dernier. Il est bien connu que le nombre de degrés de liberté (nombre de connexions) dans ce modèle dépend du nombre de vecteurs d'apprentissage, du bruit dans ces vecteurs et de la complexité des fonctions à estimer.

1.5.5.1 Théorème de Bayes

Le théorème de Bayes fournit un cadre théorique pour la problématique de la classification à deux classes, et il intervient également dans l'approche bayésienne. Si l'on considère un problème à deux classes C_1 et C_2 , le théorème de Bayes permet de calculer les probabilités à *posteriori* $P(C_1|\chi)$ connaissant les distributions des observations à *priori* [Lam-01].

$$P(C_1|\chi) = \frac{p(\chi|C_1)P(C_1)}{p(\chi)} \quad (1.33)$$

$P(C_1|\chi)$ est la probabilité à *posteriori* d'appartenir à la classe C_1 connaissant le vecteur des descripteurs χ , $p(\chi|C_1)$ est la densité de probabilité du vecteur χ dans la classe C_1 , $P(C_1)$ est la probabilité à *priori* de la classe C_1 et $p(\chi)$ est la densité de probabilité non conditionnelle définie par $p(x) = p(\chi|C_1)P(C_1) + p(\chi|C_2)P(C_2)$.

Dans le cas d'un problème de classification, cette formule définit une règle de décision : la probabilité de mauvaise classification est minimisée en sélectionnant la classe qui a la plus grande probabilité *à posteriori*.

Ce théorème est au cœur de la problématique de la classification [Bis-95] où les problèmes mal-posés constituent un défi. La régularisation d'un problème mal-posé consiste à le transformer en un problème bien-posé, c'est à dire à définir une solution unique pour toutes les mesures possibles dans l'espace des observations, et à assurer la stabilité de cette solution vis à vis des erreurs sur ces mesures [Int-99].

1.5.5.2 La Régularisation bayésienne

Les résultats décrits par F.D. Foresee et M.T. Hagan [For1-97] appliquent des réseaux de neurones multicouches qui sont employés pour la régression non-linéaire. Les réseaux sont entraînés en utilisant l'apprentissage supervisé, avec un ensemble d'apprentissage d'entrées et de cibles sous la forme $\{x_1, d_1\}, \{x_2, d_2\}, \dots, \{x_n, d_n\}$.

Nous supposons que les cibles sont produites par :

$$d_i = g(x_i) + \varepsilon_i \quad (1.34)$$

où $g(x_i)$ est une fonction inconnue et ε_i un bruit gaussien indépendant. L'objectif initial du processus d'apprentissage est de réduire au minimum la somme des erreurs quadratiques:

$$E = \sum_{i=1}^n (d_i - y_i)^2$$

Où y_i représente la réponse du réseau de neurones et n le nombre d'entrées.

Le but de l'apprentissage du réseau de neurones est de produire un réseau qui donne de petites erreurs sur l'ensemble d'apprentissage, mais qui répondra aussi correctement aux nouvelles entrées. Quand un réseau peut avoir de bonnes réponses aussi bien pour les nouvelles entrées que pour les entrées d'apprentissage, nous disons qu'il généralise bien.

L'objectif est de discuter un algorithme d'apprentissage qui produit uniformément des réseaux avec une bonne généralisation. Cette méthode d'amélioration de la généralisation est contrainte par la taille des poids du réseau. Elle est désignée sous le nom de 'régularisation'. L'idée est qu'on suppose que la véritable fonction fondamentale a un degré de lissage. Quand les poids dans un réseau sont maintenus petits, la réponse du réseau sera lisse.

Avec la régularisation, n'importe quel réseau modestement surdimensionné devrait pouvoir représenter suffisamment la vraie fonction [Mac1-91].

Dans la section suivante, nous présentons des techniques de régularisation. Des techniques bayésiennes de David Mackay [Mac2-92] sont appliquées pour optimiser la régularisation. La technique optimale de régularisation exige le calcul de la matrice hessienne (annexe 1.1). Pour réduire au minimum les coûts informatiques, F.D. Foresee [For1-97] a proposé d'employer une approximation de Gauss-Newton à la matrice hessienne. Cette approximation utilise l'algorithme de Levenberg-Marquardt pour l'apprentissage du réseau.

L'apprentissage vise en général à réduire la somme des erreurs quadratiques E . Cependant, la régularisation ajoute un terme additionnel et la fonction objective devient [For1-97] :

$$F = \beta_p E + \alpha_p E_w \quad (I.35)$$

où E_w est la somme quadratique des poids du réseau, et α_p et β_p sont des paramètres de la fonction objective.

Si :

- $\alpha_p \ll \beta_p$, alors l'algorithme d'apprentissage conduira les erreurs à devenir plus petites.
- $\alpha_p \gg \beta_p$, l'apprentissage soulignera la réduction de la taille des poids aux dépens des erreurs du réseau, produisant de ce fait, une réponse du réseau plus lisse.

Le problème principal avec l'implémentation de la régularisation, est de donner des valeurs correctes aux paramètres de la fonction objective. David MacKay [Mac2-92] a effectué un travail étendu sur l'application de la règle de Bayes à l'apprentissage du réseau de neurones et à la régularisation optimisée.

Dans le cadre bayésien, les poids du réseau sont considérés comme variables aléatoires. Après que les données soient introduites, la fonction de densité des poids $P(w|D, \alpha_p, \beta_p, M)$ pourrait être mise à jour selon la règle de Bayes par [Lam-01]:

$$P(w|D, \alpha_p, \beta_p, M) = \frac{P(D|w, \beta_p, M) P(w|\alpha_p, M)}{P(D|\alpha_p, \beta_p, M)} \quad (I.36)$$

D représente la base de données, M est le modèle particulier utilisé du réseau de neurones, et w représente les poids du réseau. $P(D|\alpha_p, \beta_p, M)$ est un facteur de normalisation qui garantit que la probabilité totale sera égale à 1. $P(D|w, \beta_p, M)$ est la probabilité que les données se produisent, donnant les poids w et $P(w|\alpha_p, M)$ est la densité antérieure, qui représente notre connaissance des poids avant que n'importe quelle donnée ne soit rassemblée. Elles sont données respectivement par:

$$P(D|w, \beta_p, M) = \frac{1}{Z_D(\beta_p)} \exp(-\beta_p E) \quad (I.37)$$

$$P(w|\alpha_p, M) = \frac{1}{Z_w(\alpha_p)} \exp(-\alpha_p E_w) \quad (I.38)$$

avec $Z_D(\beta_p) = (\pi/\beta_p)^{N/2}$ et $Z_w(\alpha_p) = (\pi/\alpha_p)^{N/2}$ où N est le nombre total de paramètres du réseau.

a) Optimisation des paramètres de régularisation

Nous considérons l'application de la règle de Bayes pour optimiser les paramètres α_p et β_p de la fonction objective [For1-97, Mac3-92, Lam-01], et l'équation (I.36) pour le facteur de normalisation peut être résolue :

$$\begin{aligned} P(D|\alpha_p, \beta_p, M) &= \frac{P(D|w, \beta_p, M) \cdot P(w|\alpha_p, M)}{P(w|D, \alpha_p, \beta_p, M)} \\ &= \frac{[\frac{1}{Z_D(\beta_p)} \exp(-\beta_p E)] \cdot [\frac{1}{Z_w(\alpha_p)} \exp(-\alpha_p E_w)]}{\frac{1}{Z_F(\alpha_p, \beta_p)} \exp(-F(w))} \end{aligned} \quad (I.39)$$

Notons que nous connaissons les constantes $Z_D(\beta)$ et $Z_w(\alpha)$. La seule partie que nous ne connaissons pas est $Z_F(\alpha_p, \beta_p)$. Cependant, nous pouvons l'estimer par développement en série de Taylor, $F(w)$ est la fonction objective. Puisque la fonction objective a la forme d'une équation quadratique dans un petit secteur entourant un point minimum, nous pouvons élargir la fonction $F(w)$ autour du point (poids) minimum de la densité postérieure noté w^{MP} où le gradient est nul.

La résolution pour les constantes de normalisation donne :

$$Z_F(\alpha_p, \beta_p) \approx (2\pi)^{N/2} [\det ((H^{MP})^{-1})]^{1/2} \quad (I.40)$$

$H = \beta \nabla^2 E + \alpha \nabla^2 E_w$ est la matrice hessienne de la fonction objective. En remplaçant ce résultat dans l'équation (I.39) nous pouvons trouver les valeurs optimales de α_p et β_p au point minimum w^{MP} . Ceci se fait en prenant la dérivée de chacun des termes du dernier coté gauche de (I.39) et les mettre égaux à zéro. Ceci donne les valeurs optimales α_p^{MP} et β_p^{MP} des paramètres de la fonction objective par :

$$\alpha_p^{MP} = \frac{\gamma}{2E_w(w^{MP})} \quad \text{et} \quad \beta_p^{MP} = \frac{n - \gamma}{2E(w^{MP})} \quad (I.41)$$

avec $\gamma = N - 2\alpha^{MP} \text{tr}(H^{MP})^{-1}$ est le nombre efficace de paramètres du réseau en réduisant la fonction erreur. Il peut s'étendre de zéro à N . H^{MP} est la matrice hessienne au point minimum et n le nombre d'entrées.

b) Approximation de Gauss-Newton à la hessienne

L'optimisation bayésienne des paramètres de régularisation exige le calcul de la matrice hessienne H au point minimum w^{MP} . F.D. Foresee a proposé d'employer l'approximation de Gauss-Newton à la matrice hessienne, qui est aisément disponible si l'algorithme d'optimisation de Levenberg-Marquardt est employé pour localiser le point minimum [Hag-96, For2-96]. Le calcul additionnel exigé pour l'optimisation de la régularisation est minimal.

Voici les étapes exigées pour l'optimisation bayésienne des paramètres de régularisation, avec l'approximation de Gauss-Newton pour la matrice hessienne:

0. Initialiser α_p , β_p et les poids. Choisir $\alpha_p = 0$ et $\beta_p = 1$ et employer la méthode d'initialisation de poids de Nguyen-Widrow [Ngu-90]. Après la première étape d'apprentissage, les paramètres de la fonction objective sont adaptés à partir de l'arrangement initial.
1. Prendre l'étape de minimisation de la fonction objective $F(w)$, de l'algorithme de Levenberg-Marquardt : $F(w) = \beta_p E + \alpha_p E_w$
2. Calculer le nombre efficace de paramètres : $\gamma = N - 2\alpha^{MP} \text{tr}(H^{MP})^{-1}$ se servant de l'approximation de Gauss-Newton à la hessienne disponible dans l'algorithme d'apprentissage de Levenberg-Marquardt: $H = \nabla^2 F(w) \approx 2\beta J^T J + 2\alpha I_N$ où J est la matrice Jacobienne de l'ensemble des erreurs d'apprentissage et I_N la matrice identité.
3. Calculer les nouvelles évaluations pour les paramètres de la fonction objective

$$\alpha_p^{MP} = \frac{\gamma}{2E_w(w^{MP})} \quad \text{et} \quad \beta_p^{MP} = \frac{n - \gamma}{2E(w^{MP})}$$

4. Répéter maintenant les étapes de 1 à 3 jusqu'à la convergence.

Nous remarquons bien que la fonction objective change avec chaque re-estimation de ses paramètres et donc, le point minimum se déplace. La précision sera assez bonne lorsque la fonction objective ne change plus de manière significative dans les prochaines itérations. Ainsi, nous obtiendrons la convergence.

Quand cette approximation de Gauss-Newton est employée dans l'algorithme bayésien de la régularisation (GNBR), de meilleurs résultats sont obtenus si les données d'apprentissage sont d'abord organisées et comprises entre $[-1, 1]$ (ou une certaine région semblable).

Après l'apprentissage, il y a quelques contrôles occasionnels qui devraient être effectués. D'abord, si le nombre efficace final des paramètres γ est très près du nombre réel des paramètres N , alors le réseau de neurones ne peut pas être assez grand pour représenter correctement la vraie fonction. Il faut ajouter dans ce cas des couches cachées et refaire l'apprentissage. Si le réseau élargi a la même grandeur que γ final, alors le réseau le plus petit était assez grand. Autrement, des couches cachées de neurones auraient dû être ajoutées. Ensuite, le deuxième contrôle est pour la consistance des résultats. Si le réseau est suffisamment grand, alors un réseau suivant plus grand réalisera des valeurs comparables à γ , E et E_W .

L'algorithme GNBR a les possibilités pour produire des résultats optimaux dès la première fois. De plus, le coût d'implémenter les changements dans l'algorithme d'apprentissage est minimal puisque dans le calcul, l'approximation de Gauss-Newton pour la hessienne est employée [For1-97]

1.5.5.3 Avantages et inconvénients de l'approche bayésienne

D'un point de vue théorique, R.Neal [Nea-96] a montré que, lorsque les probabilités *a priori* des poids sont convenablement choisies, il n'est pas nécessaire de limiter la taille du réseau pour éviter le sur-ajustement, et le nombre de neurones cachés peut tendre vers l'infini. Selon cette étude, le seul facteur qui doit limiter la taille du réseau est la capacité des ordinateurs utilisés et le temps disponible pour effectuer les calculs nécessaires.

D'un point de vue pratique, la théorie de l'approche bayésienne pour l'apprentissage des réseaux de neurones apporte d'importantes améliorations :

- Le concept de régularisation peut être interprété de façon naturelle dans le contexte bayésien.
- Comme tous les calculs se font à partir de la base d'apprentissage, il n'est plus nécessaire de disposer d'une base de validation. Il est donc possible d'utiliser toutes les données dont on dispose pour estimer les poids du réseau.
- L'incertitude sur les poids peut être prise en considération pour corriger la probabilité calculée par un réseau dans un problème de classification.
- Les entrées peuvent être sélectionnées grâce à la méthode *Automatic Relevance Determination* : un hyper paramètre est associé à chaque entrée et après l'apprentissage, les hyper paramètres avec de grandes valeurs indiquent des entrées non pertinentes.

Pour les inconvénients, nous pouvons dire :

Comme les paramètres utilisés sont maintenant issus de distributions de probabilités, il est nécessaire, pour connaître un paramètre, de calculer des intégrales faisant intervenir les distributions des autres paramètres. Il est, en général, impossible de calculer ces intégrales analytiquement, et plusieurs approches ont été proposées pour effectuer ces calculs. Les méthodes utilisées sont soit très lourdes à implémenter, soit, elles reposent sur des approximations qui peuvent fausser les résultats [Int-99].

1.5.6 LES RESEAUX A FONCTION DE BASE RADIALE RBF

1.5.6.1 Principe de base

Ce sont des réseaux à une seule couche cachée, ils sont utilisés pour la classification et l'approximation de fonctions, à travers une combinaison linéaire de fonctions non linéaires à base radiale. Le fonctionnement de ces réseaux repose sur le principe des estimateurs à noyau, et la généralisation se fait du monovarié vers le multivarié [Yed-98, Ren-95]. Un estimateur à noyau considère des fonctions de R^+ vers R qui sont de la forme [Hér-94] :

$$\phi_i(x) = \phi(\|x - \xi_i\|) \quad (I.42)$$

x représente les entrées du réseau et ξ_i représente les nœuds de l'interpolation pour $i=1, \dots, n_{im}$ (n_{im} : nombre de nœuds d'interpolation).

L'idée principale est introduite par M.J.D.Powell [Pow-87]. Elle est basée sur l'interpolation. Toute fonction $f(x)$ d'une variable $x \in R^n$ peut être approchée par une interpolation composée par la somme de p fonctions noyaux de forme fixée $\phi(x)$ [Hér-94] par:

$$f(x) = \sum_{j=1}^p \lambda_j \phi(\|x - \xi_j\|) \quad (I.43)$$

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

$$f(x_i) = y_i \quad i=1, \dots, n_{im}. \quad (I.44)$$

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

1.5.6.2 Architecture et fonctionnement des réseaux RBF

Broomhead et Lowell [Bro-88] puis Chen et al [Che-91], ont conçu un réseau de neurones à une seule couche cachée dont le fonctionnement est basé sur l'idée des estimateurs à noyaux introduite ci-dessus dans (I.42). Ainsi, à partir de l'équation (I.42) et en la considérant dans le cas multivarié, la sortie que doit délivrer ce réseau est de la forme (Fig.I.16) :

$$f(x) = \sum_{i=1}^k w_j \phi(\|x - c_i\|) \quad j = 1, m \quad (I.45)$$

c_i sont des valeurs, que nous attribuons à chaque neurone de la couche cachée et appelées centres des classes, w_j sont les poids synaptiques reliant ces neurones aux sorties, m est le nombre de sorties du réseau et k est le nombre de centres.

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

Parmi ces fonctions, on peut trouver les formes suivantes [Khe-94, Jan-97] :

Forme Cubique	$\phi(x)=x^3$	} (I.46)
Forme Multiquadratique	$\phi(x)=(x^2+k_d^2)^{-1/2}$	
Forme Logarithmique décalée	$\phi(x)=\log(x(x^2+k_d^2))$	
Forme Gaussienne	$\phi(x) = \exp(-x^2 / \beta_{rd}^2)$	

Ici x représente un réel quelconque. L'application des fonctions noyaux dans les réseaux de neurones, représentera x comme les entrées du réseau. Les paramètres k et β_{rd} décrivant la distribution des centres.

Parmi ces formes, la fonction Gaussienne, que nous avons déjà présenté dans la figure (1.8), est la plus utilisée en raison de ses caractéristiques locales. C'est cette fonction que nous utilisons dans tous les réseaux à base radiale que nous allons présenter dans le chapitre 4. A la figure 1.16, est représentée l'architecture d'un réseau RBF.

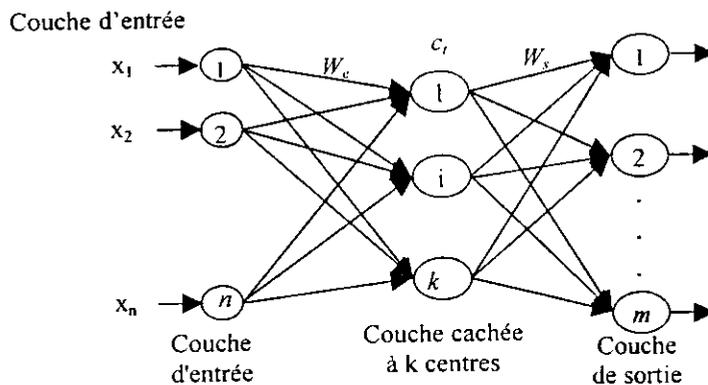


Fig.1.16 Architecture d'un réseau RBF Multi-sorties

Celui-ci est composé de trois couches (Fig.1.16). La première couche reçoit les entrées injectées au réseau. La couche cachée est constituée de neurones, dont le discriminant est à fonction de base radiale. Chacun de ces neurones est doté d'un vecteur c_i appelé centre et chaque entrée du réseau est rangée dans l'une des classes représentées par les neurones de la couche cachée. Les valeurs des centres doivent donc représenter chacune le barycentre de la classe qui lui correspond. Plus le nombre d'exemples d'entraînement est important, plus cette couche est chargée de neurones.

Lorsqu'un vecteur X est présenté à l'entrée du réseau, chaque neurone calcule, à travers son discriminant, la distance euclidienne d_{ii} de cet exemple par rapport au centre de la classe c_i qu'il représente [Khe-94, Jan-97] par l'équation :

$$d_{ii} = \|X - c_i\| \tag{I.47}$$

La sortie de chaque neurone de cette couche cachée notée $s_i(X)$, est calculée en appliquant la fonction noyau. Elle est donné par :

$$s_i(X) = \phi(\|X - c_i\|) \quad i = 1, \dots, k, \tag{I.48}$$

k étant le nombre de neurones cachés.

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

$$s_i(X) = \exp\left[\frac{(X - c_i)^T (X - c_i)}{2\sigma_i^2}\right] \quad (I.49)$$

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

L'utilisation de la fonction gaussienne rend le réseau très puissant. En utilisant cette fonction à caractéristique locale, chaque neurone ne réagit d'une manière significative qu'à une partie restreinte de l'espace d'entrée. En partageant ainsi l'espace des entrées, les aptitudes de ce réseau se trouvent, en approximation significativement améliorées par rapport aux réseaux LBF (Linear Basis Function) dont les sorties sigmoïdes ne possèdent pas cette propriété de calcul local [Lan-01].

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

$$y_j(X) = \sum_{i=1}^k w_{ij} \Phi(\|X - c_i\|) \quad j=1, \dots, m, \quad (I.50)$$

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

1.5.6.3 Apprentissage des réseaux RBF

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

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

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

Quant aux poids synaptiques, leur détermination constitue la partie la plus simple de l'apprentissage. En effet, la sortie étant linéaire, n'importe quel algorithme d'optimisation linéaire peut être aisément utilisé. Dans ces méthodes d'entraînement, il existe trois approches différentes qui peuvent être utilisées [Khe-94, Pol-94] :

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

Nous présentons ci-dessous un algorithme pour chacun des deux premiers types d'apprentissage.

a) Méthode de Centrage Adaptative (Adaptive Centering Method)

Comme pour les paramètres de pondération w_{ij} , à toute valeur attribuée aux centres c_i correspondra une certaine erreur en sortie. Il est aisé de remarquer à partir de l'équation (I.50) que cette fonction est aussi bien dérivable par rapport aux centres qu'aux poids synaptiques.

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

Nous présentons ci-dessous les étapes de cette technique d'apprentissage supervisé [Yed-98, Khe-94].

1. Initialisation des centres $c_i(0)$, les paramètres de la sortie gaussienne de chaque neurone de la couche cachée $\sigma_i(0)$, $i=1, \dots, k$, et les poids synaptiques $w_{ij}(0)$, $j=1, \dots, m$.
2. Choix du taux d'apprentissage $\alpha < 1$ variable ou fixe.
3. Présentation successive de M exemples d'entraînement.
4. Calcul de l'erreur quadratique E à la sortie connaissant la réponse désirée et celle fournie par le réseau.

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

avec m représentant le nombre de sorties du réseau. Les termes y_j^p et d_j^p représentent respectivement la $j^{\text{ème}}$ sortie du réseau calculée et désirée pour le $p^{\text{ème}}$ vecteur d'entrée.

5. Réajustement des centres et éventuellement des paramètres des fonctions noyau :

$$\begin{aligned} c_i(t+1) &= c_i(t) + \alpha \Delta c_i(t) \\ \sigma_i(t+1) &= \sigma_i(t) + \alpha \Delta \sigma_i(t) \end{aligned} \quad (I.52)$$

avec :

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

6. Réajustement des poids synaptiques.

$$w_{ij}(t+1) = w_{ij}(t) + \alpha \Delta w_{ij}(t) \quad (\alpha : \text{taux d'apprentissage}) \quad (I.53)$$

avec

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

7. Répétition des étapes de 3 à 5 autant de fois jusqu'à obtention de la précision désirée en sortie.

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

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

b) Méthode du regroupement (clustering algorithm)

Dans cette méthode, l'algorithme d'apprentissage supervisé comprend une partie d'un apprentissage qui est non supervisé de classification, basé sur l'approche du Nearest Neighborhood. Il est appelé aussi 'méthode des K-means clustering' et c'est l'une des méthodes que nous avons utilisées dans l'entraînement des RBF.

Les étapes de l'algorithme d'apprentissage du RBF dans ce cas, sont les suivantes [Yed-98, Khe-94]:

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

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

$$d_i(p) = \|X^p - c_i(p-1)\| \quad i=1, \dots, k, \quad p=1, \quad (I.54)$$

où p représente le p ème exemple

3. Noter l'argument k_{ar} du centre pour lequel la distance est minimale :

$$k_{ar} = \arg[\min(d_i(p))] \quad i=1, \dots, k \quad (I.55)$$

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

4. Réadapter les centres :

$$\begin{aligned} c_i(p) &= c_i(p-1) & i=1, \dots, k \text{ et } i \neq k_{ar} \\ c_i(p) &= c_i(p-1) + \alpha(p)[X^p - c_i(p-1)] & i= k_{ar} \end{aligned} \quad (I.56)$$

5. Réduire le taux d'apprentissage par:

$$\alpha(p) = \frac{\alpha(p-1)}{1 + \text{int}[\frac{p}{M}]^{1/2}} \quad (\text{int}[\cdot] : \text{partie entière}) \quad (I.57)$$

M : nombre d'exemples

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

Contrairement à la méthode de Centrage Adaptative, celle-ci détermine d'abord les centres, puis les poids synaptiques. L'idée de l'utilisation de l'apprentissage non supervisé dans cette méthode pour la détermination des centres peut limiter le nombre de neurones suivant le nombre de classes auquel cet apprentissage a abouti. Par contre, l'algorithme supervisé nous oblige à fixer le nombre de neurones dès le début.

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

1.5.6.4 Réseaux RBF et Approximation de fonctions

Les réseaux RBF constituent des modèles de réseaux très efficaces, notamment pour l'approximation de fonctions. L'utilisation de la fonction gaussienne permet de bénéficier de sa caractéristique locale pour faciliter l'apprentissage et améliorer l'approximation. Par ailleurs, la procédure d'entraînement, basée sur le fonctionnement de ces réseaux et 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 toujours utilisés avec une seule couche cachée. Ceci libère l'utilisateur du choix du nombre de couches.

En outre, afin d'alléger l'apprentissage, il est nécessaire de fixer les paramètres de dispersion dès le début de l'apprentissage à une valeur fixe commune pour tous les neurones. Dans ce cas, il faut savoir qu'une valeur très petite de ce paramètre nécessite beaucoup de neurones dans la couche cachée, ce qui peut diminuer les capacités de généralisation du réseau. Un paramètre de valeur très importante par contre, peut entraîner un chevauchement entre les classes rendant ainsi l'apprentissage impossible [Khe-94, Jan-97].

1.6 RESEAUX DE NEURONES A APPRENTISSAGE NON SUPERVISE

Ces réseaux peuvent classer un ensemble d'exemples de la base d'apprentissage en un nombre spécifique de catégories sans savoir à l'avance comment les organiser. Cette tâche est réalisée en adoptant la technique du clustering (groupement). Les exemples sont regroupés en fonction de leur « proximité » dans un espace à N_e dimensions où N_e désigne le nombre de descripteurs d'entrées. Le nombre maximum de catégories est fixé par l'utilisateur. Les réseaux de Kohonen font partie de cette famille [Poi-99].

Depuis 1960, ces réseaux ont bénéficié de plusieurs travaux de recherche, et malgré le développement que connaissent certains algorithmes de ce type tel que le Learning Matrix (LM) en 1963 et le Linear Associative Memory (LAM) en 1977 [Yed-98], ils restaient limités puisqu'ils n'avaient pas des modèles mathématiques exacts.

Ce n'est qu'à partir des années 80 et grâce aux travaux de Hopfield et ceux de M.Cohen et S.Grosberg [Hop-82, Coh-83] que le comportement mathématique de ces réseaux a été modélisé, ce qui a donné un grand élan à la recherche dans ce domaine.

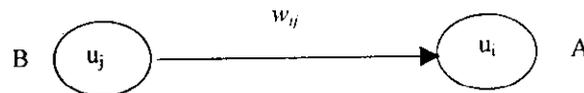
L'apprentissage non supervisé des réseaux de neurones consiste, comme dans le cas des apprentissages supervisés, à modifier les poids des connexions des neurones. Dans ce cas, les exemples de la base d'apprentissage sont des données seules : il n'est pas possible de modifier les poids du réseau en fonction d'une erreur sur les réponses souhaitées, puisqu' aucune réponse n'est connue a priori.

En fait, nous allons voir que ce sont les données elles-mêmes qui servent de réponses car les poids du réseau sont modifiés en fonction de leurs composantes. L'information utile se trouve donc uniquement dans les données.

1.6.1 Fondement de l'apprentissage non supervisé

Cet apprentissage est basé sur la loi de J.Hebb, il consiste à partager l'espace d'entrée en plusieurs classes constitutives de ce dernier. L'énoncé de cette loi est le suivant [Kos-92, Koh2-01] :

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



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

$$\Delta w_{ij} = s(u_j) s(u_i) \quad (I.58)$$

u_i et u_j représentent l'activité de A et B et, $s(u_i)$ et $s(u_j)$ leurs sorties représentées par une fonction non linéaire.

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

1.6.1.1 Principes généraux

La plupart des réseaux de neurones à apprentissage non supervisé font appel aux mêmes notions. Nous allons voir les caractéristiques communes à ces réseaux.

a) L'architecture

Dans ce cas de réseaux, nous utilisons le symbole e pour présenter les entrées. Les réseaux de neurones qui subissent un apprentissage non supervisé sont constitués de deux couches :

- Une couche d'entrée qui comporte autant de neurones que les données ont de composantes. Lorsqu'une donnée est proposée à l'entrée du réseau, chaque neurone d'entrée adopte une activation égale à la composante correspondante de la donnée.
- Une couche de sortie dont chaque neurone (s) est connecté à un neurone x de la couche d'entrée avec la pondération w , si bien que l'on peut associer à chacun des neurones de sortie un vecteur de poids W de même dimension que celle des données. On appelle ces vecteurs les vecteurs de référence.

b) Le neurone gagnant

Le neurone gagnant est un neurone de sortie qui, pour une entrée donnée définie par le vecteur entrée e , a le vecteur associé W le plus proche de e .

Selon le type de réseau, la comparaison s'effectue [Kuh-97, Koh2-01]:

- Par un produit scalaire : dans ce cas, le neurone gagnant est celui qui vérifie :
- Par un calcul de distance (généralement euclidienne). Dans ce cas, le neurone gagnant est celui qui vérifie :

$$\arg \max (W \cdot e)$$

$$\arg \min \|W - e\|$$

Si on raisonne en terme de fonctions d'entrée et de seuil qui ont été définis pour le neurone formel, on peut dire que la fonction d'entrée des neurones de sortie peut avoir deux formes:

$$f = \sum_{i=1}^n e_i \cdot w_i^s \quad (\text{produit scalaire})$$

ou

$$f = \sqrt{\sum_{i=1}^n (e_i - w_i^s)^2} \quad (\text{distance euclidienne})$$

où w_i^s est le poids de la sortie (s) venant du $i^{\text{ème}}$ neurone, e_i est la $i^{\text{ème}}$ entrée et n est le nombre d'entrées.

La fonction de seuil F est linéaire si bien que l'on a : $A = F(f) = f$.

Dans ce formalisme, le neurone gagnant correspondant à une entrée donnée est le neurone le plus activé de la couche de sortie.

1.6.2 L'apprentissage

On peut classer les réseaux à apprentissage non supervisé en deux catégories, suivant la façon dont se déroule l'apprentissage : plus précisément, suivant que chaque exemple de la base d'apprentissage entraîne la modification des poids d'un ou plusieurs neurones à chaque itération ou pas. Nous allons examiner en détail ces deux types de réseaux.

1.6.2.1 Réseaux du type 'winner takes all'

Dans les apprentissages de réseaux de ce type, chaque exemple de la base d'apprentissage influe sur les poids d'un seul neurone lors de chaque itération. Ce neurone est le neurone gagnant qui correspond à l'entrée, d'où la dénomination anglophone 'winner takes all' (WTA) généralement attribuée à ce mode d'apprentissage que l'on peut traduire par 'tout au vainqueur' [Kuh-97, Poi-99, Koh2-01]. Dans les WTA, nous citons :

a) L'algorithme LBG

Cet algorithme qui doit son nom aux initiales de ses auteurs (Linde, Buzo, et Gray 1980) est aussi connu sous le nom d'algorithme de Lloyd généralisé. L'algorithme de l'apprentissage est le suivant [Tou-92]:

1. Initialiser les poids w_i avec différents vecteurs e tirés de la base d'apprentissage.
2. Rechercher le neurone gagnant pour chaque exemple de la base d'apprentissage.
3. Pour chaque neurone s de sortie, réactualiser les poids en remplaçant le vecteur de référence W par la moyenne arithmétique des vecteurs e .
4. Retourner à la deuxième étape tant que le nombre maximal d'itérations n'est pas atteint jusqu'à la convergence du réseau.

Ce type d'algorithme a d'abord été proposé par Lloyd en 1957 et a été publié en 1982 pour effectuer la quantification d'une grandeur scalaire. On peut montrer qu'avec un nombre fini d'itérations, l'algorithme LBG conduit à une convergence vers un minimum local qui dépend de l'initialisation des vecteurs de référence.

b) Les algorithmes adaptatifs

les algorithmes adaptatifs constituent une famille d'algorithmes d'apprentissage qui permettent, contrairement à l'algorithme LBG que nous venons de décrire, de commencer l'apprentissage alors que toutes les données de la base d'apprentissage ne sont pas encore connues. Les données peuvent par exemple provenir d'une expérience qui se déroule, donnant naissance à un flux d'informations qui peuvent alors être traitées immédiatement [Tou-92].

L'algorithme de base dont plusieurs méthodes d'apprentissage découlent est le suivant :

1. Initialiser les poids w_i avec différents vecteurs e tirés aléatoirement de la base d'apprentissage (ou avec des valeurs aléatoires dans le cas où les données sont issues d'un flux).
2. Sélectionner aléatoirement un exemple d'apprentissage dans la base (ou l'exemple qui vient du flux) pour en déterminer le neurone gagnant.
3. Modifier les composantes du vecteur de référence e correspondant selon :

$$w_i(t+1) = w_i(t) + \alpha(t) \cdot (e_i - w_i(t)) \quad (I.59)$$

où $0 < \alpha(t) \leq 1$ est le taux d'apprentissage variant dans le temps.

4. Retourner à la deuxième étape tant que le nombre maximal d'itérations n'est pas atteint.

Plusieurs méthodes d'apprentissage sont envisageables suivant les variations de $\alpha(t)$ dans le temps durant l'apprentissage.

-Taux d'apprentissage constant

On montre que si α conserve la même valeur, comprise entre 0 et 1 durant l'apprentissage, l'influence des entrées précédentes décroît exponentiellement, si bien qu'on assiste à une sorte d'oubli de la partie antérieure de l'apprentissage. En effet, si on écrit les valeurs prises par les composantes w_i du vecteur de référence d'un neurone à des étapes successives de l'apprentissage lors desquelles ce neurone a été le gagnant pour les entrées e_i , on a :

$$\begin{aligned} w(1) &= w(0) + \alpha \cdot (e_1 - w(0)) \\ &= (1 - \alpha) \cdot w(0) + \alpha \cdot e_1 \\ &\vdots \\ w(t) &= (1 - \alpha)^t \cdot w(0) + \alpha \sum_{i=1}^t (1 - \alpha)^{t-i} \cdot e_i \end{aligned} \tag{I.60}$$

Finalement, l'influence au temps t_f des vecteurs d'entrée e_i se trouve négligeable devant celle du dernier vecteur e_{t_f} et ceci d'autant plus que α est proche de 1 et que t est inférieur à t_f . Par exemple, si $\alpha = 0.5$, l'influence d'un vecteur e sur les composantes de $w(t)$ après dix modifications est inférieure à 10^{-6} , on est alors en présence d'un réseau instable dont les vecteurs de référence peuvent varier dans de grandes proportions à cause d'un unique exemple d'apprentissage qui se trouve être suffisamment différent de tous ceux qui l'ont précédé [Tou-92, Kuh-97].

-Décroissance hyperbolique du taux d'apprentissage

Macqueen [Poi-99] proposa en 1967 un système disposant d'un coefficient d'apprentissage pour chaque neurone de sortie [Poi-99]. A chaque fois qu'un nouvel exemple d'apprentissage est présenté au réseau, le taux d'apprentissage correspondant au neurone gagnant noté dans ce cas $\alpha^s(t)$ est modifié selon l'expression :

$$\alpha^s(t) = \frac{1}{t^s} \quad s = 1, m \text{ (} m : \text{nombre de neurones de sortie)} \tag{I.61}$$

De cette manière, le vecteur de référence de chaque neurone de sortie est toujours égal à la moyenne arithmétique des vecteurs descriptifs des exemples d'apprentissage dont il a été le gagnant. En effet, durant l'apprentissage, le vecteur de référence correspondant à chaque neurone de sortie s est modifié selon :

$$\begin{aligned} w(1) &= w(0) + \frac{1}{1} (e_1 - w(0)) = e_1 \\ w(2) &= w(1) + \frac{1}{2} (e_2 - w(1)) = \frac{e_1 + e_2}{2} \\ &\vdots \\ w(t) &= w(t-1) + \frac{1}{t} (e_t - w(t-1)) = \frac{e_1 + e_2 + \dots + e_t}{t} \end{aligned} \tag{I.62}$$

Cette particularité a donné son nom à cet algorithme : si le réseau présente k neurones de sortie, ils sont associés à k vecteurs moyens ; nous disons : les k -moyennes ou **K-Means** en anglais. On peut remarquer que les composantes du vecteur aléatoire attribué à chaque neurone de sortie lors de l'initialisation sont oubliées dès la première modification des poids w_i . Pourtant, cela ne signifie pas que le résultat de l'apprentissage ne dépend pas de l'initialisation car c'est bien sûr des vecteurs initiaux que vont

On peut noter par exemple celle proposée par Ritter en 1991 [Poi-99], également utilisée dans les cartes auto-organisatrices.

$$\alpha(t) = \alpha_{initial} \left(\frac{\alpha_{final}}{\alpha_{initial}} \right)^{\frac{t}{t_{max}}} \quad (I.63)$$

$\alpha_{initial}$ et α_{final} sont les valeurs initiales et finales de α et t_{max} le nombre total d'itérations choisi pour l'apprentissage.

1.6.2.2 Réseaux du type 'winner takes most'

Les réseaux de ce type diffèrent de ceux que nous avons étudié dans la section précédente par le mode d'apprentissage. Alors que dans les algorithmes de type 'winner takes all' (WTA), les composantes d'un seul vecteur de référence sont modifiées à chaque présentation d'un vecteur d'entrée, les algorithmes du type 'winner takes most' imposent quant à eux la modification des vecteurs de références de plusieurs neurones [Koh2-01, Kuh-97, Poi-99].

1.6.3 Réseaux à activation compétitive

Avant de détailler le modèle des cartes auto-organisatrices de Kohonen, nous allons décrire dans un premier temps le principe général d'un réseau à architecture compétitive qui constitue la base du modèle de Kohonen.

Considérons une couche de neurones régie par un mécanisme d'activation compétitive, où une seule unité de sortie est active à un instant donné. Il y a dans ce type de réseau d'une part des liens provenant des entrées vers toutes les unités et d'autre part une interconnexion de chacune des unités d'une même couche qui permettent de garantir qu'une seule unité de la couche sera active à un instant donné. Un tel comportement est appelé "tout au vainqueur" (noté WTA : "Winner Takes All"). Le vainqueur en l'occurrence est l'unité qui présente la plus forte activation par le stimulus d'entrée. Les unités au sein de la couche sont donc en concurrence (ou en compétition) d'où le nom du réseau dit à activation compétitive ("competitive learning").

Le but de ces réseaux est de rassembler ou classer par catégorie des données présentées en entrée. Ce classement peut être utilisé, par exemple, pour encoder des données et les compresser selon une quantification vectorielle (vector quantization), dans laquelle un vecteur d'entrée est remplacé par l'index de l'unité de sortie qu'il va activer. Ce procédé s'applique à des applications comme l'approximation de fonction, traitement d'image ou l'analyse statistique.

Une particularité proche de ce concept réside dans ce que nous appellerons la cartographie de caractéristiques ("feature mapping" en anglais) qui se caractérise par le développement d'une organisation spatiale significative de la couche de sortie, alors que dans les architectures compétitives courantes, nous ne trouvons pas de relation géométrique particulière entre les unités de sortie. Pourtant, l'origine et les concepts des deux idées sont très proches [Kuh-97, Koh2-01, Kos-92].

1.6.3.1. Principe de fonctionnement d'une couche compétitive

Dans la structure de base d'une couche compétitive, on trouve une couche unique G d'unités de sortie, chacune étant repérée par un vecteur position r_i et connectée à un ensemble d'entrées e_i ($i=1$ à n) par des connexions excitatrices (poids) que nous notons ici par $w_{r_i} \geq 0$ (Fig.1.17).

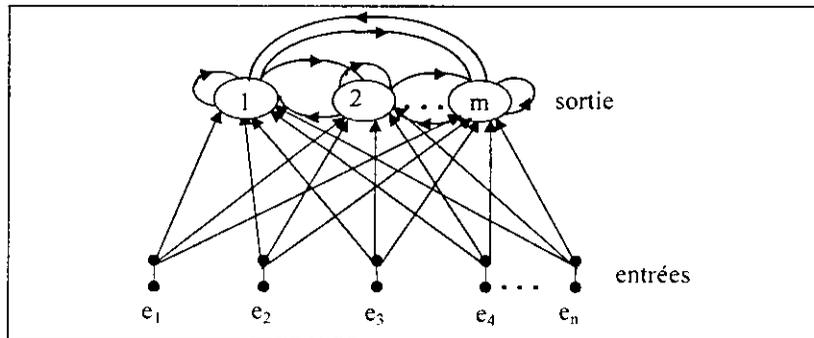


Fig. 1.17 Structure de base d'un réseau à apprentissage concurrentiel. Les connexions en flèche pleine sont excitatrices, les autres inhibitrices

Seulement une seule unité de sortie, appelée vainqueur (gagnant), est activée à un instant donné. Kohonen [Koh1-82] a montré que le vainqueur est l'unité qui a la plus grande activation donnée par (I.64):

$$S_r = \sum_{i=1}^n w_{r_i} \cdot e_i = w_r \cdot e \quad (I.64)$$

S_r est l'activation du neurone gagnant r pour un vecteur passé et w_r représente le vecteur des poids du neurone gagnant r . Dans cette partie des réseaux de neurones non supervisés, et puisque le réseau sera représenté par un neurone gagnant, nous notons le vecteur des poids de ce neurone w_r . Si on change le vecteur d'entrée passé par un autre courant toujours noté e , l'inéquation (I.65) :

$$w_{r'} \cdot e \geq w_r \cdot e \quad (\forall r \neq r' \in G) \quad (I.65)$$

définit alors pour ce nouveau vecteur d'entrée, l'unité gagnante r' avec une sortie correspondante $y_{r'} = 1$ et $y_{r \neq r'} = 0$. Si les poids de chaque unité sont normalisés, c'est-à-dire si $|w_r| = 1 \quad \forall r$, alors (I.65) est équivalent à :

$$|w_{r'} - e| \geq |w_r - e| \quad (\forall r) \quad (I.66)$$

Ce qui signifie que le vainqueur est l'unité comportant le vecteur poids normalisé w_r le plus proche du vecteur d'entrée e . Divers travaux (par exemple ceux d'Obermayer [Poi-99]) ont porté sur la simulation de la compétition en itérant les interactions entre les neurones de la couche. Dans une simulation par ordinateur, on pourra simplement chercher la valeur maximale de s_r et simplifier ainsi l'algorithme.

Plus concrètement, dans un réseau, on pourra implémenter un ensemble d'unités "tout au vainqueur" avec des inhibitions latérales, chaque unité inhibant les autres comme on peut le voir à la figure 1.17. Une connexion auto-excitatrice est alors également requise et les poids latéraux ainsi que la fonction d'activation non-linéaire doivent être choisis correctement afin qu'une seule unité de sortie soit élue [Kuh-97, Poi-99].

1.6.3.2 Règle d'apprentissage dans une couche compétitive

Un réseau de type 'tout au vainqueur' effectue une classification des données à l'aide du critère donné dans l'équation (I.65 ou I.66). Nous présentons le principe que suit le réseau compétitif pour définir ces classes à partir des données d'entrée.

On commence par initialiser aléatoirement les poids avec des valeurs faibles, puis on présente au réseau un ensemble de données d'entrée e . Pour chaque vecteur entrée, on détermine le vainqueur r' parmi les sorties et on effectue la mise à jour des poids $w_{r'}$ du neurone gagnant uniquement afin que le vecteur $w_{r'}$ puisse se rapprocher du vecteur d'entrée courant e .

1.6.4 Carte auto-organisatrice de Kohonen

L'algorithme le plus connu est sans doute celui des cartes auto-organisatrices (Self-Organization Maps abrégées en SOM) ou cartes topologiques de Kohonen. Il représente un type particulier de réseau compétitif qui définit un voisinage spatial pour chaque unité de sortie [Koh1-82, Kuh-97, Poi-99, Hon-97].

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

Les neurones de la couche (appelée couche topologique) sont disposés géométriquement selon une topologie bien définie dans un réseau de dimension p . Ceci signifie que pour $p=2$, le réseau a une forme carrée (un réseau hexagonal est quelquefois utilisé) et devient une grille où chaque neurone possède quatre voisins (excepté pour ceux situés sur les bords). Dans un réseau carré, chaque neurone est ainsi connecté à $2p$ neurones voisins [Koh2-01].

Chaque neurone est relié aux n composants du vecteur d'entrée et pondéré par autant de poids synaptiques. Ces derniers forment le vecteur de poids synaptiques du neurone (Fig. 1.18).

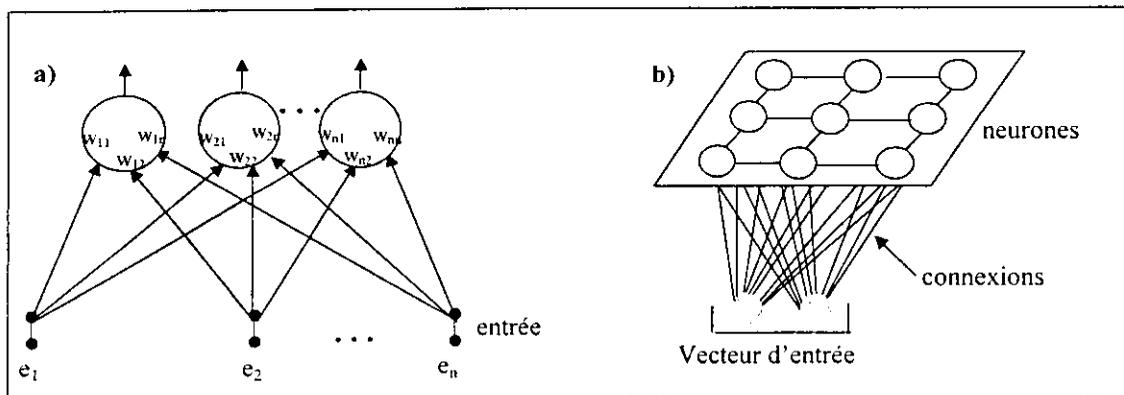


Fig. 1.18 Structure d'un réseau de Kohonen. a) un réseau 1D, b) un réseau 2D

L'espace d'entrées est ainsi de dimension égale au nombre de poids par neurone (ici, n). Les neurones de cette couche sont reliés aux n entrées e_1, e_2, \dots, e_n . Pour chaque entrée, on affecte un nombre de poids égal au nombre de neurones dans la couche. Dans la configuration 1.18-a, le vecteur d'entrée est de dimension n et la couche est de dimension 1 ($p=1$). Sur la figure (1.18-b), on distingue un réseau de Kohonen organisé selon une grille de neurones en 2 dimensions. Dans ce cas, le vecteur d'entrée est de dimension 2 ($p=n=2$) [Kuh-97, Hon-97].

1.6.4.1 Neurone formel de Kohonen

Le neurone formel de Kohonen (Fig.1.19-a), qui est l'élément constituant du réseau de Kohonen, est caractérisé par une variable d'état qui est son activation (et également sa sortie). Cette variable ne prend que des valeurs positives.

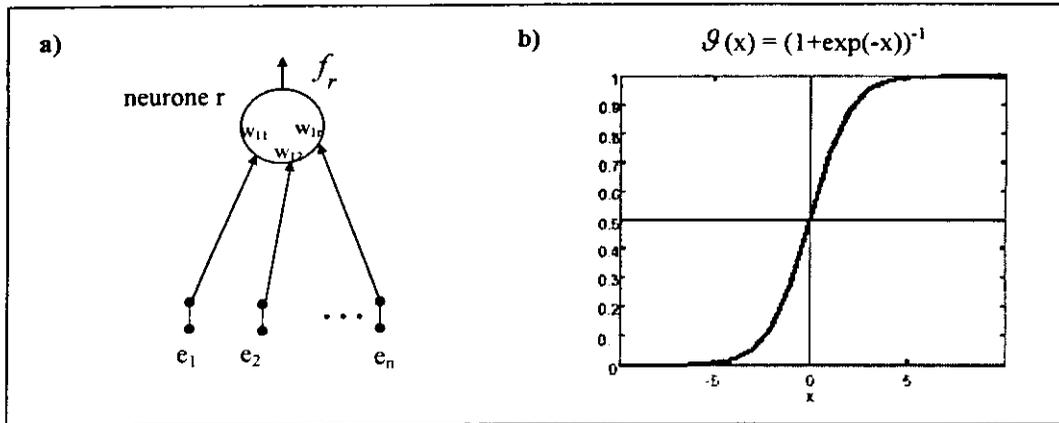


Fig. 1.19 a) le neurone formel de Kohonen b) la fonction d'activation $\mathcal{G}(x)$ du neurone

La loi de fonctionnement d'un tel neurone est donnée par l'équation suivante :

$$f_r(\mathbf{e}) = \mathcal{G}\left(\sum_{i=1}^n (w_{ri} \cdot e_i) - \theta_p\right) \quad (1.67)$$

r est le vecteur position des neurones dans la couche appelée G et n est le nombre d'entrées.

w_{ri} est le vecteur de poids synaptiques reliant chacune des composantes d'entrée e_i au neurone r , \mathcal{G} est une fonction sigmoïde (Fig.1.19-b) et θ_p un terme de perte non linéaire qui permet de rendre compte du phénomène de saturation qui veut que la fréquence à laquelle émet un neurone soit bornée.

La caractéristique la plus importante est que le neurone formel agit comme une unité de traitement non linéaire [Kuh-97, Poi-99].

1.6.4.2 Compétition dans une couche de Kohonen

Les progrès des méthodes électriques dans la neurobiologie ont permis de montrer que certaines cellules nerveuses avaient des rôles bien spécifiques dans les systèmes de perception. On a pu montrer qu'entre les neurones qui recevaient les signaux des cellules réceptrices, il existait un mécanisme d'interaction latérale qui dépendait de la distance entre les neurones concernés. Cette dépendance est représentée par la fonction dite du "chapeau mexicain" [Koh2-01] :

Cette courbe s'interprète de la manière suivante :

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

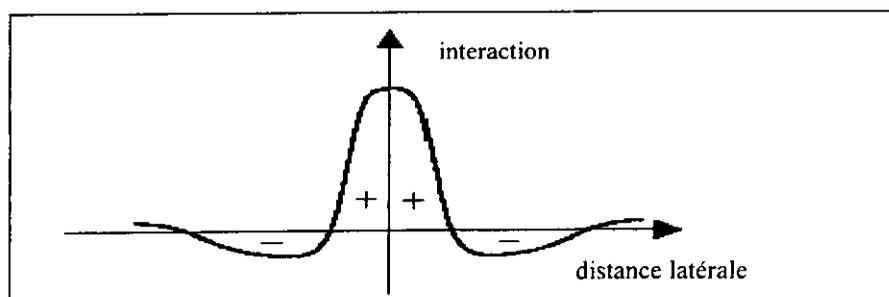


Fig. 1.20 Influence d'un neurone sur les voisins en fonction de l'éloignement
 + : excitatrice ($w > 0$), - : inhibitrice ($w < 0$)

Pour comprendre l'influence de cette zone d'interaction latérale, Kohonen simule un réseau à une dimension (ce qui simplifie le problème qui consiste à déterminer les zones latérales d'excitation et d'inhibition), auquel il envoie un signal d'entrée qu'il laisse évoluer. L'évolution de la compétitivité des neurones de la couche s'écrit alors :

$$f_r = \mathcal{G} \left(\sum_{i=1}^n w_{r_i} \cdot e_i + \sum_r (g_{rr'} \cdot f_{r'}) - \theta_p \right) \quad (I.68)$$

$g_{rr'}$ désigne une fonction d'interaction entre un neurone r et un neurone r' de la couche. Chaque activation $f_{r'}$ du neurone r' fournit une contribution $g_{rr'} \cdot f_{r'}$ au signal d'entrée global du neurone r .

Les résultats des simulations montrent que le réseau s'auto-organise et fournit une répartition des sorties telle qu'il se crée un amas de neurones d'activation élevée autour du neurone qui avait été le plus stimulé par le signal d'entrée. En revanche, les autres neurones du réseau se stabilisent dans un état d'activation faible.

Par rapport à la description très générale effectuée au paragraphe (1.6.3.1), il faut inclure certaines contraintes pour assurer la convergence de la compétition. On peut distinguer les connexions qui réalisent la compétition (souvent intra-couches et inhibitrices) de celles qui sont responsables du traitement des entrées du réseau (souvent inter-couches et excitatrices). Ainsi, le profil d'interaction latérale dite du "chapeau mexicain" est réalisé par des poids fixés d'avance au niveau des connexions intra-couches. La connectivité figée entre les neurones en compétition permet donc d'assurer la compétition au sein de la couche indépendamment d'un effet de l'apprentissage. Ce dernier a pour but de conditionner la réponse des neurones lors de la présentation des entrées et ainsi d'influer sur le comportement final.

L'algorithme de Kohonen intègre l'ensemble de ce mécanisme compétitif dans une simplification algorithmique. Il considère qu'un seul neurone est actif après la compétition : la fonction "tout au vainqueur" est réalisée par un mécanisme global. Il n'y a plus de connexions latérales explicites au niveau de la couche en compétition, c'est la règle d'élection qui tient directement compte des relations de voisinage dans la couche. Ce mécanisme permet d'économiser l'ensemble des itérations déterminant l'état d'activation des neurones pour une entrée donnée [Koh2-01]. La règle de sélection du vainqueur proposée par Kohonen est basée sur une mesure de distance (distance euclidienne) entre le vecteur e d'entrée et le vecteur poids w_r du neurone considéré. Ainsi, lors de la présentation d'un vecteur d'entrée, le mécanisme WTA (Winner takes all) va désigner le neurone vainqueur r' tel que donné par la condition :

$$\|w_{r'} - e\| = \min \|w_r - e\| \quad (I.69)$$

Le vecteur poids de ce neurone est alors celui qui est le plus proche du vecteur d'entrée et qui le représente au mieux [Kuh-97, Poi-99].

1.6.4.3. Apprentissage et voisinage dans une couche de Kohonen

Une autre particularité des réseaux de Kohonen réside dans la structuration topologique de la couche de neurones qui introduit la notion de voisinage et de distance entre neurones dans la couche. L'apprentissage va permettre d'établir un rapport entre la topologie des neurones et celle de l'espace des entrées.

Dans la phase d'apprentissage, le neurone gagnant va modifier ses poids d'entrée pour devenir un peu plus 'sensible' au vecteur présenté, c.à.d que son activation initiale sera

plus forte la prochaine fois que ce même vecteur activera le réseau. Cette adaptation va aussi concerner les neurones voisins du gagnant, mais dans une moindre mesure. Le neurone gagnant influence la plasticité (taux d'apprentissage) des synapses environnantes en fonction de leur distance. L'utilisation de la fonction gaussienne va déterminer la taille du voisinage pendant l'apprentissage, c.à.d le nombre de neurones qui seront affectés par l'étape d'adaptation.

1.6.4.4 Caractéristiques du modèle physiologique des SOM

Certaines hypothèses de la théorie des SOM paraissent avoir des aspects biologiques [Koh2-01]. Elles représentent:

- Une implémentation naturelle et robuste de la fonction WTA basée sur un simple modèle dynamique d'un neurone et sur un réseau de neurones latéralement interconnecté.
- Une interprétation réaliste de la loi de la plasticité synaptique.
- Une explication naturelle du contrôle latéral de plasticité (la fonction voisinage) entre les neurones durant l'apprentissage.

1.7 MONOGRAPHIE DES TRAVAUX PUBLIÉS DANS L'APPLICATION DES RNA EN TECHNIQUE DE HAUTE TENSION

Dans cette section, nous présentons une synthèse bibliographique des travaux publiés qui traitent de l'utilisation des réseaux de neurones artificiels dans le domaine de la haute tension. Ce n'est qu'à partir de 1994 que les RNA ont commencé à apparaître dans leurs applications en haute tension. Les travaux que nous allons présenter de 1994 jusqu'à 2003 constituent l'essentiel des références à l'échelle internationale.

1994

A.A.Mazroua et al. [Maz1-94, Maz2-95] ont publié dans IEEE deux travaux sur la reconnaissance des tailles d'impulsions de décharges partielles (DP) produites par différentes sources de décharges en utilisant un réseau compétitif : classificateur du plus proche voisin (nearest neighbor classifier), un réseau LVQ (learning vector quantization) et un réseau multicouches dont l'algorithme d'apprentissage est la BP. Les trois RNA ont présenté les mêmes performances pour toutes les données présentées sauf pour le cas des petites cavités. Le réseau LVQ a montré les meilleures capacités de reconnaître la différence entre les DP produites par une seule cavité et celles produites par une arborescence électrique.

1995

K.Engel et D.Peier [Eng-95] ont étudié l'influence du développement des décharges partielles sur l'identification du type de défaut en utilisant les RNA. Le changement d'activité des décharges partielles détermine les données mesurées et leur classification. Dans cette étude, la classification des DP est effectuée en utilisant la carte de Kohonen.

Dans un second travail, P.S.Ghosh, et al. [Gho-95] ont modélisé un arc électrique d'un isolant pollué en utilisant un réseau de neurones artificiels entraîné par l'algorithme d'apprentissage de type « Resilient Propagation Algorithm (RPROP) ». Le RPROP est un nouvel algorithme d'apprentissage qui prend en considération la topologie locale de la fonction erreur E. L'adaptation des poids ou leur modification dépend des signes des dérivées de E et non pas de leurs amplitudes. L'adaptation des poids est basée sur ce qu'on appelle la règle d'apprentissage de Manhattan.

Un autre travail présenté par M.E.Ghourab a porté sur l'utilisation des RN pour identifier l'effet de la moisissure sur le spectre de polarisation de l'isolation d'un transformateur. L'algorithme utilisé était celui de la BP. Il est appliqué pour distinguer la différence entre les spectres de polarisation sous les différentes conditions de vieillissement et déterminer la quantité de la moisissure [Ghr-95].

Un autre travail a été publié dans une revue japonaise par T. Nogami [Nog-95] sur l'utilisation des RN pour une pré-détection d'anomalies dans l'huile d'un transformateur. Pour cela, quatre détecteurs de gaz dissous et de leurs mélanges dans l'huile ont été utilisés. Les résultats de la base de données sont comparés à ceux des mesures chromatographiques. Aucun détail n'a été donné concernant le RN utilisé. Les résultats qui en découlent sont présentés donnant les types de défauts pouvant apparaître dans l'huile en fonction des gaz dissous captés et leurs concentrations.

1997

Z.Wang et P.Griffin [Wan1-97] ont publié leur travail dans IEEE sur la combinaison d'un RN et d'un système expert pour le diagnostic de défauts d'un transformateur en utilisant la méthode de l'analyse des gaz dissous. Le type du RN a été seulement cité et utilisé comme classificateur des entrées qui sont les concentrations des gaz dans l'huile et les taux de leur génération.

Un autre travail publié dans IEEE par O.Venegas et al. [Ven-97] concerne l'utilisation des RN pour le diagnostic de l'huile de transformateur. Le diagnostic est basé sur la composition chimique des gaz dissous dans l'huile. Aucun détail n'a été donné concernant le RN utilisé.

T.Kexioug, Z.Deheng [Kex-97] ont utilisé les RNA pour la reconnaissance des modèles de DP tridimensionnels. Ils ont étudié l'influence du nombre de neurones des couches d'entrées et cachées afin de choisir la meilleure structure du réseau. La capacité des réseaux de neurones à reconnaître les DP est testée par une comparaison entre les résultats de simulation et ceux de l'expérience.

S.Happe, H.G.Kranz [Hap-97] ont développé un système de suppression en temps réel des signaux d'impulsions des DP en appliquant les RN. Les pulsations sonores et les taux de répétition stochastique peuvent être particulièrement nuisibles et causent une mauvaise interprétation des activités des DP. Tous les systèmes de mesure des DP connus sont affectés par ce problème. Donc, la suppression des pulsations non souhaitables conduit à l'évaluation des DP.

Le RN auto-organisé de Kohonen classe les signaux inconnus. Cette classification est utilisée pour l'apprentissage du réseau. Pour son adaptation, l'algorithme de BP a été utilisé.

L.Tang, M.R.Raghveer [Tan-97] ont étudié la classification du vieillissement des transformateurs de puissance dans des conditions normales, de surtension et d'arc électrique, en utilisant les RN. Ceci est obtenu par l'analyse des gaz dissous. Le RN assume parfaitement les tâches de reconnaissance des perturbations. Il est capable de classer les différents types de défauts cités.

1999

S.Chowdhary [Cho-99] a présenté un travail dans le Symposium International de Haute Tension (ISH'99) sur la détermination de la tension de rupture du SF₆ sous un champ non-uniforme. Le RN a été entraîné par la PPLN (project pursuit learning network). Ce genre de réseau se construit au fur et à mesure que l'apprentissage se fait. Il commence

par un neurone caché. En ajoutant un autre neurone, l'apprentissage ne se fait pas pour tout le réseau, mais seulement pour le nouveau neurone. Ce réseau a de bonnes capacités de généralisation.

Dans un autre travail, Y. Tian [Tia-99] a utilisé les RNA pour l'identification des DP, à partir de la mesure des émissions acoustiques (EA). Les RN utilisés pour cette identification sont la BP, le réseau de Kohonen (SOM) et le Learning Vector Quantization (LVQ). Une comparaison entre les trois méthodes utilisées a été effectuée en changeant des paramètres d'apprentissage : le nombre de neurones cachés, le nombre de cycles d'apprentissage et les entrées.

Dans un autre rapport [Nun-99], les RNA ont été utilisés pour la détermination de l'intensité du flux électrique et de la valeur critique de la tension disruptive dans les sous stations, en tenant compte des différents facteurs atmosphériques et géométriques.

Le premier objectif est de trouver une méthode perspective pour le dimensionnement des sous stations de 138 kV. Le RN utilisé est composé de deux réseaux en cascade, le premier calcule la tension disruptive qui devient l'une des entrées du second réseau. Ce dernier calcule l'intensité du champ électrique entre les jeux de barres.

2000

A la conférence Internationale IEEE sur les Isolations Electriques et les Phénomènes de Décharges (CEIDP'2000), un article sur l'utilisation des RN en HT a été présenté par X.H. Jin [Jin-00] qui a utilisé deux RN pour classer les défauts provoqués par les DP ayant lieu dans l'isolation d'un transformateur. Les deux RN utilisés sont la BP et le LVQ. Ils ont tous les deux prouvé leur capacité de généralisation.

Nous avons présenté une communication à la conférence internationale Quality, Reliability & Maintenance QRM [Bou1-00] des résultats de l'application des réseaux RBF entraînés par la ROM dans la prédiction de l'effet du vieillissement thermique du PVC sur ses caractéristiques mécaniques. Ceci est effectué sous différentes températures de vieillissement. La durée de vie du PVC a été prédite aussi en fonction du temps de vieillissement.

2001

A la CEIDP 2001, notre article [Mok1-01] a été le seul présenté dans l'application des RN. Le travail se résume dans l'utilisation des RN à base radiale entraîné par l'algorithme de la ROM pour prédire la tension de claquage des intervalles d'air pointe-barrière-plan avec différentes positions de la barrière et différentes tensions appliquées.

Nous avons présenté aussi à la Conférence Internationale 'Advances in Processing, Testing and Application of Dielectric Materials (APTADM'01) un travail sur l'application des réseaux RBF entraînés par la ROM dans la prédiction du vieillissement thermique de l'huile de transformateur utilisée par SONELGAZ (BORAK22) en utilisant deux réseaux : FFN pattern et le Batch learning. Pour chacun des réseaux, nous avons utilisé deux temps d'apprentissage. Une prédiction à 3000 h a été faite aussi [Mok3-01]. Utilisant les mêmes réseaux, nous avons présenté la prédiction d'autres caractéristiques dans les conférences (SIMI, PES, SEIAW) données par les références [Mok10-01, Mok11-01, Mok12-01].

A l'ISH'01 les mêmes réseaux ont été utilisés dans la prédiction du vieillissement thermique de l'huile de transformateur et du PVC [Mok2-01]. Un autre travail a été présenté par Y. Kawaguchi et T. Shimizu [Kaw-01] sur l'application des RN à la reconnaissance des DP dans les gaz isolants.

2002

A la Conférence Internationale des liquides diélectriques (ICDL'2002), notre article était le seul présentant l'utilisation des RN dans les isolants de HT [Mok4-02]. Il a porté sur l'utilisation d'un réseau RBF entraîné par la ROM pour la prédiction du vieillissement thermique de l'huile de transformateur et sur l'influence du pas et du temps d'apprentissage sur la qualité des résultats. Un autre réseau de BP a été appliqué pour le diagnostic de l'huile entraîné par l'algorithme de Levenberg-Marquardt (L.M) et par celui du momentum. Une comparaison entre les deux méthodes a été faite.

A la 4^{ème} Conférence QRM [Mok13-002], à la Conférence IEEE- Systems-Man & Cybernetics (SMC) [Mok16-02], à la Conférence Internationale 'Energy Efficiency-ICEE' [Mok19-03] et à la Conférence Internationale Condition Monitoring and Diagnostic Engineering Management (COMADEM) [Mok21-03] et à d'autres conférences citées dans les références [Mok14-02, Mok15-02], nous avons appliqué les RNA dans le diagnostic et la classification de la BORAK22 utilisant la BP en trois variantes, les réseaux bayesiens et les réseaux de Kohonen. Nous avons présenté aussi des études sur l'influence de l'architecture et de quelques paramètres d'apprentissage sur la qualité des résultats.

A la Conférence Nationale de Haute Tension (CNHT), les premiers travaux qui ont utilisé les RN dans les systèmes HT ont été présentés en 2002. L'un est celui de A.Ailam et al. [Ail-02], présentant l'utilisation des RN à base radiale pour prédire des caractéristiques non-linéaires d'un isolateur pollué. L'autre article était le nôtre [Mok5-02], dans lequel nous avons présenté un réseau à BP entraîné par la méthode L.M pour le diagnostic des huiles de transformateur utilisées par SONELGAZ. Le diagnostic utilisé était basé sur les tests physico-chimiques.

Au Symposium International des Isolations Électriques (ISEI), nous avons présenté un travail [Mok6-02] sur la prédiction de propriétés diélectriques dans les isolants solides, liquides et gazeux. Deux paradigmes des réseaux RBF ont été présentés : entraînement par la ROM et par la BP. Nous avons présenté l'application de ces réseaux sur d'autres caractéristiques dans la référence [Mok17-02].

Dans les références [Mok7-02, Mok26-03], nous avons effectué une comparaison entre quatre méthodes dans les RNA utilisés dans le diagnostic de l'huile de transformateur utilisée par SONELGAZ : la méthode de L.M, la régularisation bayésienne, la méthode du momentum et celle avec momentum et adaptation du taux d'apprentissage.

2003

Nous avons présenté une communication à l'ISH03 sur l'application des réseaux de neurones de Kohonen (SOM) pour la prédiction du vieillissement thermique du PVC et de la BORAK22 [Mok8-03]. Pour rendre la prédiction indépendante de la nature de la caractéristique d'isolation, nous avons transformé la base de données des entrées en retranchant la valeur moyenne de la valeur réelle et en divisant le résultat par la variance des données. Au lieu des données réelles, nous employons de cette façon les prétendus 'profils'. Nous avons présenté d'autres travaux [Mok20-03, Mok22-03] utilisant ce même réseau pour la prédiction de la tension disruptive des systèmes pointe-barrière-plan en fonction, respectivement, de la longueur de l'intervalle et de la largeur de la barrière.

Nous avons publié nos travaux portant sur les réseaux RBFG utilisés dans la prédiction du vieillissement thermique à partir de différentes caractéristiques du PVC et de la BORAK22 et sur le diagnostic de la BORAK22 dans les revues IEE, Journal of

COMADEM et Archives of Electrical Engineering (AEE) [Mok23-03, Mok24-03, Mok25-03]

A la CNHT03 [Mok9-03] un réseau RBF et un SOM ont été appliqués dans la prédiction du vieillissement thermique du PVC et de la BORAK22. Le réseau RBF est entraîné par la ROM couplée avec la méthode des K-means. Ce dernier réseau a été détaillé dans [Mok18-03].

1.8 CONCLUSION

Nous avons présenté dans ce chapitre plusieurs types de réseaux de neurones classés en réseaux supervisés et non supervisés :

Pour les réseaux supervisés, nous avons décrit au début la méthode de la BP qui est très souvent utilisée dans les réseaux de neurones et quelques unes de ses variantes les plus connues.

Les réseaux bayesiens ont été aussi présentés avec leur fondement. La régularisation bayésienne qui sera appliquée dans toutes les applications faites dans ce travail a été décrite et nous avons remarqué sa grande capacité d'optimisation des paramètres du réseau et de généralisation.

En ce qui concerne les réseaux RBFG et leurs méthodes d'entraînement, ils constituent des modèles efficaces pour l'approximation. Grâce à la caractéristique locale de la fonction gaussienne, son application donne aux réseaux RBFG la capacité d'améliorer l'approximation.

En ce qui concerne les réseaux à apprentissage non-supervisé, nous avons vu qu'il était possible de les classer dans deux catégories : 'winner takes all' WTA et 'winner takes most' WTM. La première catégorie contient les réseaux compétitifs que nous avons utilisés dans la classification et la seconde contient les cartes auto-organisatrices de Kohonen (SOM) que nous avons utilisées dans la classification et la prédiction.

1.9 BIBLIOGRAPHIE

- [Ail-02] E.Ailam, M.Teguar, A.Mekhaldi, 'Utilisation des réseaux de neurones dans le domaine de la pollution des isolateurs', Numéro spécial AJOT, Série B-, 4ème CNHT, pp: 136-139, Ghardaia, Algérie, Jan. 2002
- [Bau-86] P.Baldi, E.B.Baum, 'Bounds on the size of ultrametric structures', Physics Revue Letters, 56, pp: 1598-1600, 1986.
- [Bis-95] C.M. Bishop, 'Neural networks for pattern recognition', Oxford: Clarendon Press, 1995.
- [Bou1-00] A. Boubakeur, L. Mokhnache, A.L. Nemmour, A. Khelifi, 'Neural network application in the field of HV cable insulation thermal ageing', 3rd Int. Conf. on Quality, Reliability and Maintenance (QRM), Professional Publishers, IMecE, pp:122-126, Oxford, UK, Apr. 2000.
- [Bro-88] D.S. Broomhead, D. Lowe, 'Multivariable functional interpolation and adaptive networks', Complex systems, 2, pp:321-355, 1988.
- [Che-91] S.Chen, C.F.N Cowan, P.M. Grant, 'Orthogonal least squares learning algorithm for radial basis function networks', IEEE Trans. on neural networks, 2(2), pp:302-309, Mar. 1991.
- [Cho-99] S.Chowdhury, M.S.Naidu, 'Application of Artificial Neural Network (ANN) in SF6 Breakdown Studies in Non-Uniform Field Gaps', 11th ISH, London, UK, paper 5.204.P5, Aug. 1999.
- [Clé-90] F.M. Clément 'Réseaux de neurones', Eyrolles, 1990.
- [Coh-83] M.A. Cohen, S.Grosberg, 'Absolute stability of global pattern formation and parallel memory storage by competitive neural networks', IEEE Trans. on Systems, Man and cybernetics, Vol. SMC-13, pp: 815-826, 1983.
- [Cul-43] W.S.McCulloch, W.H. Pitts, 'A logical calculus of the ideas imminent in nervous activity', Bull. Math. Biophy, 5, pp:115-133, 1943.
- [Eng-95] K.Engel, D.Peier, 'Influence of PD Fault Development on Fault Type Recognition Using an Artificial Neural Network', 9th ISH, paper N° 5861, Graz, Austria, Aug. 1995.
- [For1-97] F. D. Foresee, M.T. Hagan 'Gauss-Newton approximation to Bayesian learning', Proceedings of the 1997 Int. Joint Conference on Neural Networks, pp: 1930-1935, 1997.
- [For2-96] F. D. Foresee, 'Generalization and Neural Networks', Ph.D. Dissertation, Oklahoma State University, 1996.
- [Gho-95] P.S.Ghosh, S.Chakravorti, N.Chatterjee, 'ANN with RPROP Algorithm Applied to Modelling of Flashover Voltage Characteristics of Polluted Insulators', 9th ISH, Graz, Austria, Aug. 1995.
- [Ghr-95] M.E.Ghourab, 'Application of Artificial Neural Networks for Insulation Dignosis', 9th ISH, , paper 5595, Graz, Austria Aug. 1995.
- [Gro-77] S. Grosberg, 'Classical and instrumental learning by neural networks', in progress theoretical biology, New York academic press, Vol.3, pp:51-141, 1977.
- [Hag-96] M. T. Hagan, H. B. Demuth, and M. H. Beale, 'Neural Network Design', Boston, MA: PWS Publishing, 1996.
- [Hap-97] S.Happe, H.G.Kranz, 'Real-Time Partial Discharge Disturbing Pulse Suppression with a Neural Network Hardware', 10th ISH, Montréal, Canada, Aug. 25-29, 1997
- [Har-94] R.L. Harvey, 'Neural network principles', Printice Hall International INC, 1994.

- [Heb-49] D.O. Hebb, 'The organization of behavior, a neuropsychological theory', New York : John Wiley, 1949.
- [Hér-94] J.Hérault, 'Réseaux neuronaux et traitement du signal', Traité des nouvelles technologies, Ed. Hermes, 1994.
- [Hin-83] G.E.Hinton, T.J. Sejnowski, 'Optimal perceptual inference', Proceeding of IEEE-Comput-Soc. Conf. on Computer Vision and Pattern Recognition, pp: 448-453, 1983.
- [Hol-97] A. Holst, 'The use of Bayesian neural network model for classification tasks', PHD thesis, Dept. of numerical analysis and computer science, Royal institute of technology, S-100 44 Stockholm, Sweden, Trita-NA-P9708, 1997.
- [Hon-97] T.Honkela, 'Self-Organization maps in natural language processing', PHD thesis, Hilsinky University of technology, neural networks center, Finland, Dec. 1997.
- [Hop-82] J.J. Hopfield, 'Neural networks and physical systems with emergent collective computational abilities', Proc.Natl.Acad.Sci. 79: 2554-58, 1982.
- [Int-99] 'Chapitre 6 Apprentissage des réseaux de neurones et régularisation', voir 'www.neurones.espci.fr/Theses_PS/Stricker_M/CHAP6.pdf', thèse de Doctorat 1999.
- [Jac-88] R.A.Jacobs, 'Increased rates of convergence through learning rate adaptation', IEEE-Neural networks 1: 295-307, 1988.
- [Jan-97] J.S. Jang, C.T. Sun, E. Mizutani 'Neuro-Fuzzy and soft computing', Prentice Hall, USA, 1997.
- [Jin-00] X.H Jin, C.C.Wang & al., 'Comparison of PD classification capabilities for transformer failure and typical noise models with neural networks application', CEIDP'2000, IEEE Annual report Conf. on Electrical Insulation and Dielectric Phenomena, Vol.1, pp: 288-291, Victoria, BC, Canada, Oct. 2000.
- [Kaw-01] Y. Kawaguchi, T. Shimizu 'Neural network application to the partial discharge pattern recognition in GIS', 12th Int. Symposium on High Voltage Eng., ISH'2001, Volume 5, paper 8-7, Banagalore, India, 2001.
- [Kex-97] T.Kexiong, Z.Deheng, W.Zhenyuan, 'Recognition of Partial Discharge Based on Artificial Neural Network', 10th ISH, Montréal, Canada, Aug. 1997.
- [Khe-94] S. Khemaissia, A.S. Morris 'Review of Networks and choice of radial basis function Networks for systems identification', Technologies avancées, pp: 55-85, 1994.
- [Koh1-82] T. Kohonen, 'A simple paradigm for the self-Organized formation of structured feature maps', in Competition and cooperation in neural nets. Ed. S.Amari, M.Arbib, Vol. 45 Berlin: Springer Verlag, 1982.
- [Koh2-01] T.Kohonen, 'Self-Organizing maps', 3rd Ed. Information Sciences – Springer-Verlag, Berlin, 2001.
- [Kos-92] B.Kosko, 'Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence', Prentice Hall, Englewood Cliffs, 07632, 1992.
- [Kuh-97] D. Kuhn 'Une Approche Neuronale pour l'Asservissement Visuel d'un Robot Manipulateur', thèse de doctorat, Université de Haute Alsace, Sept. 1997.
- [Lam-01] J. Lampinen, A. Vehtari 'Bayesian Approach for Neural Networks – Review and Case Studies', Journal of Neural Networks, 14(3):7-24, Apr. 2001 (Invited article).
- [Lan-01] A. Lendasse, J. Lee & al., 'Approximation par réseaux à fonctions radiales de base – Application à la détermination du prix d'achat d'une option', ACSEG 2001 proceedings, Connectionist Approaches in Economics and Management Sciences, Rennes (France), pp: 275-283, 2001.

- [Mac1-91] D.J.C. MacKay, 'Bayesian methods for adaptive models', Ph.D. thesis, Computation and Neural Systems, California Institute of Technology, Pasadena, USA., 1991.
- [Mac2-92] D. J. C MacKay, 'Bayesian interpolation', Neural Computation, Vol. 4, N° 3, pp: 415-447, 1992.
- [Mac3-92] D.J.C MacKay 'A practical Bayesian framework for Back-propagation networks', Neural Computation, 4(3), pp: 448-472, 1992.
- [Maz1-94] A.A.Mazroua, R.Bartnikas, M.M.A. Salama, 'Discrimination between PD pulse shapes using different neural networks paradigms', IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 1, N° 6, pp: 1119-1131-Dec.1994
- [Maz2-95] A.A.Mazroua, R.Bartnikas, M.M.A. Salama, 'Neural networks systems using the multiplayer perceptron technique for the recognition of PD pulse shapes due to cavities and electrical trees', IEEE Trans. on Power Delivery, Vol.10,N°1, pp: 92-96, Jan. 1995
- [Min-69] M. Minsky, S. Papert, 'Perceptrons', Cambridge, Mass: MIT Press, 1969.
- [Mok1-01] L. Mokhnache, A. Boubakeur, 'Prediction of the Breakdown Voltage in a Point-Barrier-Plane Air gap using Neural Networks', CEIDP'2001, IEEE-Annual report Conf. on Electrical Insulation and Dielectric Phenomena, pp: 369-372, Ontario- Canada, Oct. 2001
- [Mok2-01]. L.Mokhnache, A.Boubakeur, 'Neural networks application to predict high voltage insulation thermal ageing', 12th Int. Symposium on High Voltage Eng., ISH'2001, Banagalore (India), Volume 5, paper N° 8-8, 2001.
- [Mok3-01] L.Mokhnache, A.Boubakeur, B.Ould Noureddine, M.A.R. Bedja, 'Prediction Of Long Thermal Ageing Of Transformer Oil Using Neural Networks', Proceedings of the International Conference on Advances in Processing, Testing and Application of Dielectric Materials APTADM'2001, N° Spécial PRZEGLAD ELEKTROTECHNICZNY, pp: 277-280, Wroclaw, Poland, Sept. 2001.
- [Mok4-02] L.Mokhnache, A.Boubakeur, N.Nait Said, 'Application of Neural Networks Paradigms in the Diagnosis and Thermal Ageing Prediction of Transformer Oil', Int. Conf. on Dielectric Liquids-ICDL'2002 proceedings, IEEE-DEIS, Paper N° 048, pp: 258-261, Graz, Austria, Jul. 2002.
- [Mok5-02] L. Mokhnache, A. Boubakeur, N.Nait Said, 'Application Of Neural Networks In The Diagnosis Of Transformer Oil Used By Sonelgaz', Numéro special AJOT- Série B, 4^{ème} CNHT, pp: 51-54, Ghardaia, Algérie, Jan. 2002
- [Mok6-02] L.Mokhnache, A.Boubakeur, 'The Use of some Paradigms of Neural Networks in Prediction of Dielectric properties for High Voltage liquid-solid and gas Insulations', Int. Symp. Of Electrical Insulation ISEI'200, IEEE-DEIS proceedings, pp: 306-309, Boston, Massachusetts, USA, Apr. 2002.
- [Mok7-02] L.Mokhnache, A.Boubakeur, 'Comparison of Different Neural Networks Algorithms Used in the Diagnosis of Transformer Oil', IEEE-DEIS proceedings CEIDP'2002, pp: 244-247, Cancun, Quintana Roo, Mexico, Oct 2002.
- [Mok8-03] L.Mokhnache, A.Boubakeur, 'Self-Organization neural networks in prediction of insulation thermal ageing', ISH2003, IEEE-DEIS proceedings, Delft-Netherlands, Aug. 2003.
- [Mok9-03] L.Mokhnache, A.Boubakeur, 'RBF and Self-Organization neural networks in prediction of insulation thermal ageing', CNHT2003, Oran-Algeria, Dec. 2003.

- [Mok10-01] L.Mokhnache, A.Boubakeur, A.Khelifi, A.L.Nemmour, 'Neural Networks Application in the Thermal Ageing Prediction of the PVC used in the High Voltage Cables', 1^{er} Symp. Int. de la Maintenance Industrielle, SIMI, session 7, Alger, Algérie, Jan. 2001.
- [Mok11-01] L. Mokhnache, A. Boubakeur, A. Feliachi and al., 'Application of neural networks in the thermal ageing prediction of transformer oil', IEEE-PES summer meeting, paper n°01sm059; Vancouver, Canada, Juil. 2001.
- [Mok12-01] L. Mokhnache, A. Boubakeur, 'The Use of Neural Networks in Prediction of Dielectric properties for High Voltage liquid-solid and gas Insulations' CD-ROM, 3^{ème} symposium des Industries Electriques du monde Arabe, Alger, Algérie, Nov. 2001.
- [Mok13-002] L. Mokhnache, A. Boubakeur, N.Nait Said, 'Diagnosis Of Transformer Oil Using Some Neural Networks Algorithms', 4th QRM, Professional Publishers, IMecE, pp:191-194, Oxford, UK, Apr. 2002.
- [Mok14-02] L.Mokhnache, A.Boubakeur, 'The use of some back-propagation paradigms in the classification of the insulating transformer oil', Proceeding of the Int.. Federation of Operational Research Societies-IFORS'2002, session MB22, pp: 22, Edenburgh, UK, Juil. 2002.
- [Mok15-02] L.Mokhnache, A.Boubakeur, N. Nait Said, 'Some Neural networks Paradigms In Insulating Transformer Oil Maintenance decision', 2nd Int. Conf. On Electrical and Power Engineering EPE, Iasi, Romania, Nov. 2002.
- [Mok16-02] L.Mokhnache , A.Boubakeur, N.Nait Said, 'Comparison of Different Neural Networks Algorithms Used in the Diagnosis and thermal ageing prediction of Transformer Oil', Int. Conf. on Systems, Man & Cybernetics, IEEE-SMC CD-ROM, paper ID: SMC02-REG-485, Hammamat, Tunisia Oct.,2002.
- [Mok17-02] L.Mokhnache, A.Boubakeur, 'Insulation properties prediction applying artificial neural networks', 5^{ème}-Conf. Nat. sur la Physique Appliquée, pp: 261-262, Batna, Algeria, Oct. 2002
- [Mok18-03] L.Mokhnache , A.Boubakeur, 'RBF Neural Networks for Insulation Thermal Ageing', Int. Conf. On Artificial Neural Networks and Genetic Algorithms, ICANNGA'2003 Proceedings, pp: 144-148, Springer-Verlag, Roanne-France, Apr.2003
- [Mok19-03] L.Mokhnache, A.Boubakeur, 'Classification Of Transformer Oil Using Self-Organizing Networks And Bayesian Neural Networks', Int. Conf. on Energy Efficiency ICEE proceedings, Algiers-Algeria, May 2003.
- [Mok20-03] L.Mokhnache, A.Boubakeur, 'Self-Organization Neural Networks In Prediction of the breakdown voltage in a Point Barrier-Plane Air Gap using different barrier widths', GCC, IEEE CD-ROM, paper ID: 8, Bahrain, May 2003.
- [Mok21-03] L.Mokhnache, A.Boubakeur, N.Nait Said, 'Influencing training parameters in some supervised neural networks for transformer oil diagnosis', Int. Conf. on Condition Monitoring and Diagnostic Engineering Management (COMADEM), Växjö, Sweden, Aug. 2003.
- [Mok22-03] L.Mokhnache, A.Boubakeur, 'Self-Organization Neural Networks In Prediction of the breakdown voltage in a Point Barrier-Plane Air Gap', Trans. of 4th Int. Conf. on Neural Networks and Applications of World Scientific And Engineering Academy And Society (WSEAS) , ID: 454-187, Athena-Greece, Jun. 2003.

- [Mok23-03] L. Mokhnache, A. Boubakeur, A. Feliachi, 'Thermal Ageing Prediction of Transformer Oil and PVC of High Voltage Cables Using Neural Networks', IEE Proceedings - Science, Measurement and Technology, pp: 107-112, Vol. 150, Issue 03, May 2003.
- [Mok24-03] L. Mokhnache, A. Boubakeur, 'Use of Neural Networks in the Monitoring of High Voltage Insulation Thermal Ageing', acceptation finale pour publication dans Journal of COMADEM, Oct. 2003.
- [Mok25-03] L.Mokhnache, A.Boubakeur, 'Application of Many Neural Networks Paradigms in the High Voltage Insulation', Archives of Electrical Engineering (AEE), Polish Academy of Science, Vol.3, Oct. 2003.
- Mok26-03] L.Mokhnache, A.Boubakeur, N.Nait Said, Ali Feliachi, R. Kattan, 'Supervised and Unsupervised Neural Networks Used in the classification and Diagnosis of Transformer Oil', soumis à IEEE-Energy Conversion Transactions 2003.
- [Nea-96] R. Neal, 'Bayesian learning in neural networks', Springer, 1996.
- [Ngu-90] D. Nguyen, B. Widrow, 'Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights', Proceedings of IJCNN, Vol. 3, pp: 21-26, Jul. 1990.
- [Nog-95] T. Nogami & al., 'Gas discrimination method for transformer faults by neural network', Electrical Engineering in Japan, Vol.115, N° 1, pp: 93-102, 1995.
- [Nun-99] I. Nunes Da Silvan, A. Nunes Da Souza, 'Analysis of High-Voltage Substations Design Using Artificial Neural Networks', 11th ISH, London, UK, Aug. 1999.
- [Poi-99] P. Poinçot 'Classification et recherche d'information bibliographique par l'utilisation des cartes auto-organisatrices, applications en astronomie', thèse de Doctorat, Observatoire astronomique de Strasbourg-Université de Louis Pasteur, Déc. 1999.
- [Pol-94] R. Poli, R.D. Jones 'A neural net model for prediction', Journal of the American Statistical Association, Vol. 89, N°. 425, pp: 117-121, Mar. 1994.
- [Pow-87] M.J.D.Powell, 'Radial basis functions for multivariable interpolation :A review', J.C. Mason and M.G. Cox, Editors, Algorithms for Approximation, pp:143-167, Oxford University Press, 1987.
- [Ren-95] J.M Renders, 'Algorithmes génétiques et Réseaux de Neurones. Application à la commande des processus', Hermes, 1995.
- [Ros-58] F. Rosenblatt, 'The perceptron: a probabilistic model for information storage and organization in the brain', Psych. Rev. 65:386-408, 1958.
- [Rum1-86] D.E. Rumelhart, G.E.Hinton & al., 'Learning internal representations by error propagation', D.E. Rumelhart, J.L. McClelland, Editors, parallel distributed processing, Vol.1, Chap.8, pp: 318-362, MIT Press, Cambridge, M.A, 1986.
- [Rum2-90] D.E. Rumelhart, 'Brain style computation: learning and generalization', in introduction to neural and electronic networks, New York: Academic Press, 1990.
- [Sol-81] F.J. Sollis, R.J.B. Wetts, 'Minimization by random search techniques', Math. Operational Research, 6, (1), pp: 19-29, 1981.
- [Tan-97] L.Tang, M.R.Raghuveer, 'Classification of Transformer Faults by Using DGA Based Artificial Neural Network', 10th ISH, Montréal, Canada, Aug. 1997.
- [Tia-99] Y. Tian, P.L.Lewin, A.E.Davies, 'PD Pattern Identification Using Acoustic Emission Measurement And Neural Networks', 11th ISH, London, UK, Aug. 1999.

- [Tou-92] C. Touzet 'les réseaux de neurones artificiels introduction au connexionisme cours, exercices et travaux pratiques' , EC2 Edition, Paris, 1992.
- [Ven-97] O.Venegas & al. ,'Diagnosis of oil-insulated power apparatus by using neural network simulation', IEEE Trans. on Dielectrics and Electrical Insulation, Vol.4, N° 3, pp: 290-299, Jun 1997.
- [Wan1-97] Z.Wang, Y.Liu, P.J. Griffin,'A combined ANN and expert system for transformer fault diagnosis', IEEE Trans. on Power Delivery, Vol.13, N° 4, pp:1224-1229, Oct. 1997.
- [Wer-90] P.J. Werbos 'Back-propagation through time, what is does and how to do it?', proceedings of the IEEE Vol. 78, N° 10, Oct. 1990.
- [Wid-60] B.Widrow, M.E.Hoff,'Adaptive switching circuits', 1960 IRE Western Electric Show and Convention Record, part 4, pp: 96-104, Aug.1960.
- [Yed-98] Y.M.Yeddou 'Etude de synthèse sur les Réseaux de Neurones et leurs applications, thèse de magistère en Automatique à l'ENP, Alger 1998.
- [Zur-92] J.M. Zurada,'Introduction to artificial neural systems', West publishing Company, USA, 1992.

Chapitre II

Mécanismes diélectriques dans les matériaux isolants

2.1 Introduction

2.2 Caractéristiques de vieillissement du PVC

a) Propriétés mécaniques des polymères

b) Propriétés électriques des polymères

2.2.1. Vieillissement thermique des polymères

2.2.1.1 Processus de dégradation des isolants électriques solides

2.2.1.2 Vieillissement des isolants électriques

a) Vieillissement thermique

b) Vieillissement électrique

c) Couplage de contraintes (thermique et électrique)

2.2.1.3 Endurance thermique du PVC

2.2.2. Vieillissement thermique du PVC

2.2.2.1 Présentation du PVC

2.2.2.2 Propriétés du PVC

a) Propriétés physico-chimiques

b) Propriétés électriques

c) Propriétés mécaniques

2.2.2.3 Résultats expérimentaux du vieillissement thermique du PVC

a) Variation de l'allongement à la rupture en fonction du temps de vieillissement

b) Résistance à la traction

c) Endurance thermique

d) Perte de masse

2.3 Propriétés et caractéristiques de vieillissement de l'huile de transformateur

2.3.1 Propriétés électriques

2.3.2 Propriétés physiques

2.3.3 Propriétés chimiques

- 2.3.4 Caractéristiques de service
 - 2.3.4.1 Aspect, couleur et caractéristiques optiques
 - 2.3.4.2 Stabilité chimique
 - 2.3.4.3 Stabilité électrique aux décharges partielles
- 2.3.5 Sécurité d'emploi
- 2.3.6 Propriétés de la BORAK22
 - 2.3.6.1 Rigidité diélectrique
 - 2.3.6.2 Acidité
 - 2.3.6.3 Teneur en eau
 - 2.3.6.4 Couleur
 - 2.3.6.5 Viscosité cinématique
- 2.4 Vieillessement des huiles de transformateurs
 - 2.4.1 Influence de la température sur les propriétés de l'huile
 - 2.4.1.1 Pertes diélectriques
 - 2.4.1.2 Permittivité relative (ϵ_R)
 - 2.4.1.3 Résistivité transversale
 - 2.4.1.4 Viscosité
 - 2.4.2 Influence du vieillissement thermique sur les propriétés de l'huile
 - 2.4.2.1 Facteur de pertes diélectriques
 - 2.4.2.2 Permittivité relative (ϵ_R)
- 2.5 Rupture des intervalles d'air pointe-barrière-plan
 - 2.5.1 Influence de la barrière isolante sur la tension de décharge disruptive
 - 2.5.1.1 Influence de la position de la barrière dans l'intervalle d'air
 - 2.5.1.2 Influence de la largeur de la barrière
 - 2.5.1.3 Influence de la forme de tension
 - 2.5.1.4 Influence de la conductivité de la barrière
 - 2.5.1.5 Influence de la longueur de l'intervalle
- 2.6 Conclusion
- 2.7 Bibliographie

2.1 INTRODUCTION

Lorsqu'un isolant gazeux, liquide ou solide est soumis à une contrainte électrique permanente ou exceptionnelle, il peut se produire un claquage ou une perte des propriétés isolantes du matériau [Fou1-86]. D'autres contraintes fonctionnelles (thermique, mécanique, exposition à des décharges ou à des rayonnements ionisants) contribuent à la diminution de la durée de vie de ces isolants et systèmes.

Dans ce chapitre, nous nous intéressons essentiellement aux mécanismes diélectriques dans les isolants solides, liquides et gazeux. Respectivement, les isolants choisis pour étude sont le PVC utilisé dans les câbles fabriqués par CABEL (ex. ENICAB), la BORAK22, huile de transformateur utilisée par SONELGAZ, et les intervalles d'air pointe-barrière-plan. En effet, ces matériaux ou systèmes d'isolation ont fait l'objet de travaux de recherche au laboratoire de haute tension (HT) de l'ENP et dont les résultats constituent notre banque de données dans l'application des réseaux de neurones.

La contrainte la plus sévère sur les isolants solides et liquides est bien la contrainte thermique [Min-99, Bou2-97, Bou3-97, Bou4-99]. C'est pour cela que la présentation d'une étude théorique et expérimentale [Bou2-97, Bou3-97, Bou4-99, Bou5-00, Bou6-01, Bou7-01, Ham-97, Abd-02] succincte est nécessaire.

Aussi, pour les isolants liquides, un travail de diagnostic de l'huile de transformateur en service sera présenté dans le chapitre 3. Ceci implique une connaissance de propriétés électriques, physiques et chimiques d'une telle huile. Nous nous limitons aux propriétés utilisées dans les essais faits au laboratoire de Contrôles, Mesure et Essais (CME) d'El-Hamma/SONELGAZ. Ces propriétés serviront aussi comme base de données pour l'entraînement des réseaux de neurones utilisés dans le diagnostic et la classification de la BORAK22 et seront présentés dans le chapitre 3.

L'étude des phénomènes de claquage des intervalles d'air a permis de prévoir le dimensionnement des lignes à haute tension [Fou1-86]. Le grand nombre d'essais de laboratoire, réalisés sur les systèmes d'électrodes les plus courants a permis d'établir des formules empiriques de calcul des tensions d'amorçage [Roy-84]. D'un autre côté, les modèles décrivant la décharge dans un intervalle d'air ne peuvent pas être combinés ensemble dans un modèle général, capable de prévoir le comportement d'un espace donné sujet à une tension donnée [Nas-90, Mee1-90, Mok34-97].

Tout cela nous a poussé à réfléchir à l'utilisation des RN dans la prédiction des caractéristiques de décharges pour des paramètres différents de ceux de l'expérience. Les arrangements utilisés sont des systèmes pointe-barrière-plan et les résultats expérimentaux sont donnés essentiellement dans la thèse de doctorat de A. Boubakeur [Bou8-79].

Trois parties sont donc présentées dans ce chapitre :

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

2.2 CARACTERISTIQUES DE VIEILLISSEMENT DU PVC

Avant de présenter les caractéristiques de vieillissement du PVC, nous décrivons succinctement les propriétés et les caractéristiques de vieillissement des polymères.

a) Propriétés mécaniques des polymères

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

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

b) Propriétés électriques des polymères

Ces propriétés sont : la rigidité diélectrique, la résistance à la décharge superficielle et la résistance transversale. D'une façon générale, les propriétés électriques sont liées aux structures moléculaires du polymère, mais il faut quand même tenir compte de :

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

2.2.1. Vieillissement thermique des polymères

2.2.1.1 Processus de dégradation des isolants électriques solides

Au contact des liquides ou des gaz, le comportement des polymères varie suivant leur nature et leur composition. Il peut y avoir diffusion, absorption avec gonflement ou gélatinisation, dissolution selon l'agressivité des produits en présence. Ceci peut causer des contraintes qui peuvent être catastrophiques pour le matériau. Selon les conditions d'exploitation et différents régimes d'application de tension, les contraintes appliquées (champ électrique, température, etc) peuvent conduire à différents types de claquage. Nous citons les plus importants.

a) Claquage électrique

Le mécanisme de claquage électrique est basé sur le phénomène d'avalanche électronique. Un électron issu de l'électrode acquiert suffisamment d'énergie pour libérer un second électron dans l'isolant. L'effet d'avalanche commence puisque chacun des deux électrons peut à nouveau libérer d'autres électrons et ainsi de suite jusqu'à la perforation du matériau. Les champs nécessaires à un tel mécanisme sont élevés mais la présence d'hétérocharges présentes à proximité des électrodes peut le favoriser.

b) Claquage thermique

Un important accroissement de la conductivité électrique est dû à une température trop importante de l'échantillon. La densité de courant augmente et occasionne plus de pertes, donc un échauffement par effet Joule. Ce processus peut engendrer une détérioration locale ou plus importante du matériau. Ce phénomène peut être rencontré lorsqu'il existe des points chauds dans un système.

c) Claquage électromécanique

Ce type de claquage est dû à la pression électrostatique qu'exercent les électrodes sur l'échantillon. Si celle-ci est trop importante, l'épaisseur de l'échantillon peut diminuer, engendrant de ce fait une augmentation du champ électrique. Cet effet peut être accru par l'échauffement du matériau. Cependant, un tel mécanisme est très rare dans les systèmes électriques.

d) Claquage par décharges partielles

La présence de cavités, ayant une permittivité plus faible que celle du matériau, dans l'isolant mène à un accroissement du champ électrique dans ces cavités. Un claquage local peut alors naître, entraînant la dégradation du matériau. En effet, l'accumulation de décharges dans une vacuole provoque généralement l'extension de la vacuole et favorise la formation d'arborescences (treeing). Lorsque les arborescences sont suffisamment développées, l'isolant n'assure plus sa fonction et il y a claquage.

2.2.1.2 Vieillessement des isolants électriques

La plupart des publications traitant du vieillissement des polymères tendent à déterminer la durée de vie du matériau en fonction des contraintes que celui-ci subit. Différents modèles ont été établis selon le type de contraintes appliquées. Nous présenterons les contraintes thermiques et électriques. L'objectif étant de montrer l'influence des contraintes électriques lors du vieillissement thermique du matériau.

Dans grand nombre des études consultées, la propriété du matériau observée lors du vieillissement est la tenue à la rupture ou rigidité électrique. L'observation de cette propriété nécessite une approche statistique poussée, utilisant généralement les distributions de Weibull (2 ou 3 paramètres) pour la détermination des valeurs de la rigidité électrique. D'autres propriétés du matériau peuvent être observées en fixant, arbitrairement ou non, un critère de fin de vie. Ces propriétés peuvent être électriques (ex: seuil d'apparition des décharges partielles) ou physico-chimique (ex: bande d'absorption IR) [Gui-00].

a) vieillissement thermique

Le vieillissement thermique des matériaux polymères est évalué selon le modèle Arrhénien [Gui-00]. Ce modèle est construit à partir de l'observation d'une propriété p du matériau. $F(p)$ est la fonction décrivant l'évolution de p au cours du temps.

Le vieillissement peut être défini par $F(p)=R.t$ ou R et t sont respectivement la vitesse (ou taux) et le temps de vieillissement.

Lorsque la propriété du matériau atteint une valeur limite p_l , définie comme critère de fin de vie, alors : $F(p_l)=R.L$ L étant la durée de vie du matériau.

On peut alors en déduire que le taux de vieillissement et la durée de vie sont inversement proportionnels. Ce modèle est basé sur l'équation d'Arrhénus qui définit la constante de vitesse d'une réaction chimique [Gui-00]:

$$R_t = A. \exp\left(\frac{B}{T}\right) \quad (\text{II.1})$$

R_t est la constante de vitesse de la réaction considérée et dans notre cas le taux de "vieillessement".

A et B sont des constantes (B : énergie d'activation) et T est la température absolue.

Mécanismes diélectriques dans les matériaux isolants

La durée de vie sous contrainte thermique L_t est donnée par :

$$L_t = k_t \cdot \exp\left(\frac{B}{T}\right) \quad \text{(II.2)}$$

$k_t = 1/A$.

A température ambiante (T_0), la constante k_t peut s'écrire:

$$k_t = L_0 \cdot \exp\left(\frac{B}{T_0}\right) \quad \text{(II.3)}$$

et en posant
$$\Delta T = \frac{1}{T_0} - \frac{1}{T} \quad \text{(II.4)}$$

nous obtenons :
$$L_t = L_0 \cdot \exp(-B\Delta T) \quad \text{(II.5)}$$

L_0 durée de vie sous température ambiante.

b) Vieillessement électrique

La durée de vie d'un matériau isolant sous contrainte électrique est prédite principalement par deux modèles:

- la loi de puissance inverse;
- la loi exponentielle.

Loi de puissance inverse : La durée de vie du matériau s'exprime par l'équation [Gui-00] :

$$L_e = c \cdot G^{-n} \quad \text{(II.6)}$$

Loi exponentielle : Comme pour le vieillissement thermique, la durée de vie s'écrit:

$$L_e = k \cdot \exp(-hG) \quad \text{(II.7)}$$

L_e est la durée de vie sous contrainte électrique, G la contrainte appliquée, c , n , k et h sont des constantes.

Cependant, lorsque aucune contrainte électrique n'est appliquée au matériau à température ambiante, sa durée de vie doit être L_0 telle que définie par le vieillissement thermique. Ainsi dans le cas du modèle en puissance inverse, on a le système:

$$\begin{cases} \ln\left(\frac{L_t}{L_0}\right) = -B \cdot \Delta T \\ \ln\left(\frac{L_e}{L_0}\right) = -n \cdot \ln\left(\frac{G}{G_0}\right) \end{cases} \quad \text{(II.8)}$$

G_0 : valeur seuil de la contrainte électrique pour laquelle le vieillissement est effectif.

Pour le modèle exponentiel, nous avons:

$$\begin{cases} \ln\left(\frac{L_t}{L_0}\right) = -B \cdot \Delta T \\ \ln\left(\frac{L_e}{L_0}\right) = -h \cdot G \end{cases} \quad \text{(II.9)}$$

c) Couplage de contraintes (thermique et électrique)

La loi simple d'addition des contributions au vieillissement ($R=R_t + R_e$) n'est pas valide. Le modèle développé par Eyring pour le couplage des contraintes thermique et électrique est le suivant [Gui-00]:

$$R = A \cdot \exp\left(\frac{B}{T}\right) \cdot \exp\left[\left(a + \frac{b}{T}\right) \cdot f(G)\right] \quad (II.10)$$

A, B, a, b : constantes indépendantes du temps, de la température et de la contrainte électrique et $f(G)$: fonction de la contrainte électrique.

Ainsi, la durée de vie du matériau est obtenue en considérant le modèle exponentiel pour le vieillissement électrique:

$$L = L_0 \cdot \exp(-B \cdot \Delta T - h \cdot G + b \cdot \Delta T \cdot G) \quad (II.11)$$

D'après les lois de vieillissement présentées ci-dessus, on note que la détermination de la durée de vie d'un isolant n'est pas une opération précise et unique.

2.2.1.3 Endurance thermique du PVC

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

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

Pour y répondre, il suffit de connaître la relation contrainte – temps : $C = f(t)$ (Fig. 2.1) .

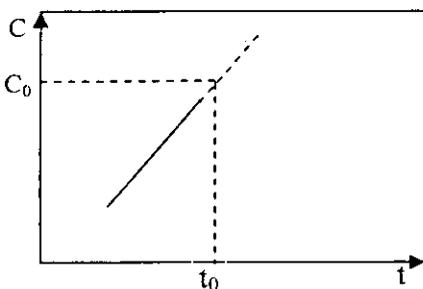


Fig. 2.1 Variation de la contrainte appliquée en fonction du temps de vieillissement

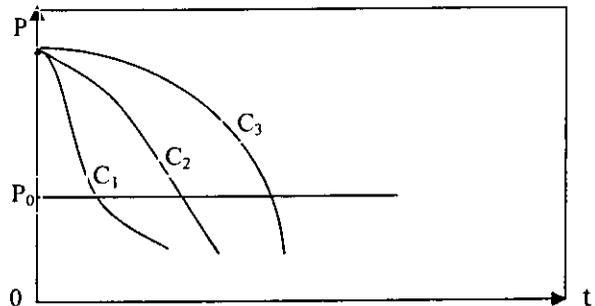


Fig. 2.2 Variation de la propriété en fonction du temps de vieillissement pour différentes valeurs de la contrainte.

Généralement, on détermine les variations d'une propriété du matériau en fonction du temps de vieillissement pour différentes valeurs C_1, \dots, C_n de la contrainte appliquée C (Fig.2.2). Si on fixe une valeur de la propriété P_0 choisie en fonction des critères d'emploi, on obtient la relation $C(t)$ relative à la valeur de la propriété P_0 . Cette courbe est appelée courbe de durée de vie.

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

a) Principe de la détermination de l'endurance thermique des isolants solides

Les travaux relatifs à l'endurance thermique des matériaux isolants solides ont pour objet la mise au point des méthodes générales de détermination de leur stabilité. Le but

Mécanismes diélectriques dans les matériaux isolants

de ces essais est de simplifier les résultats et de les exploiter sous forme d'indices de température attribués aux matériaux pour leurs propriétés électriques.

En ce qui concerne les essais, on soumet les matériaux à l'action prolongée de la chaleur puis on détermine l'évolution d'une ou plusieurs de leurs propriétés en fonction du temps de vieillissement, ce qui permet de tracer les caractéristiques $P = f(t)$ (Fig.2.3)

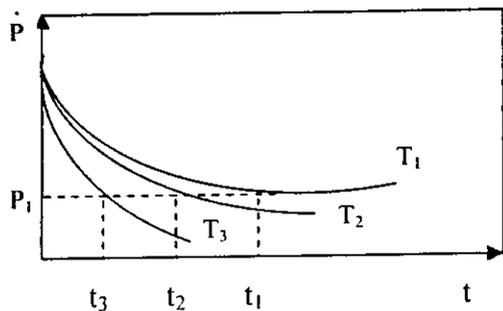


Fig. 2.3 Variation d'une propriété P en fonction du temps de vieillissement pour différentes températures.

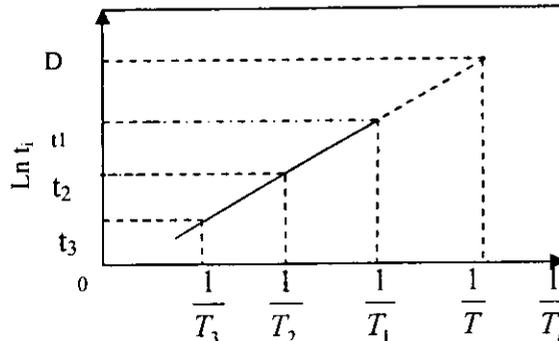


Fig. 2.4 Variation du logarithme de la durée de vie d'un polymère en fonction de l'inverse de la température.

Le choix d'une valeur limite de la propriété P₁ permet de déterminer les temps pour lesquels cette limite (durée de vie) est atteinte et ceci pour plusieurs températures de vieillissement. En répétant les essais, on peut tracer la courbe d'endurance thermique ou durée de vie : $t = f(T)$ (Fig.2.4).

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

b) Choix des températures de vieillissement

La C.E.I. 216-1 fixe un minimum de trois températures différentes telles que le critère de dégradation soit atteint.

- En un temps inférieur à 100 h pour la température la plus élevée.
- En un temps inférieur à 5000 h pour la température la plus basse.

2.2.2. Vieillessement thermique du PVC

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

2.2.2.1 Présentation du PVC

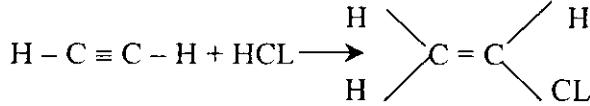
Le PVC est très utilisé dans l'isolation des câbles de haute tension à cause de son faible prix de revient. A l'état pur, le PVC est cassant ; d'où la nécessité de lui ajouter des produits auxiliaires appelés adjuvants. Le type et la quantité de chaque adjuvant sont le grand problème des producteurs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

a) formulation et mise on œuvre du PVC

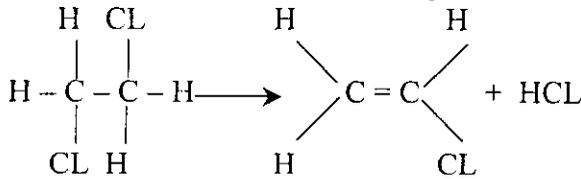
- Préparation du chlorure de vinyle

Le chlorure de vinyle monomère (CVM) $\text{CH}_2=\text{CH-Cl}$ est préparé suivant deux procédés principaux :

- Addition du gaz chlorhydrique à l'acétylène :



- Séparation de l'acide chlorhydrique du dichloréthane :



- Mélange de PVC prêt à l'emploi

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

- Les adjuvants

Les stabilisants

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

Les plastifiants

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

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

Les charges

L'emploi des charges ne vise généralement qu'à baisser le prix de revient ou à améliorer les propriétés diélectriques mais non les propriétés mécaniques [Nej-91]. Les principales charges utilisées sont : le carbonate de calcium, le kaolin calciné, la silice et l'amiante.

Les lubrifiants

L'emploi de lubrifiants a pour but de faciliter l'écoulement du PVC lors de la mise en œuvre. Il a pour effet d'atténuer la tendance qu'a la résine de base à adhérer aux surfaces environnantes. Ceci est réalisé à une température de collage supérieure à la température de travail. Les lubrifiants peuvent être classés en deux groupes d'après la firme ULTRYL PVC :

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

Les pigments et colorants

Ils permettent de réaliser la coloration des mélanges. Parmi les colorants et pigments, il faut choisir ceux qui sont stables aux températures de travail et éventuellement en présence de traces d'acide chlorhydrique.

2.2.2.2 Propriétés du PVC

a) Propriétés physico-chimiques

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

- Le PVC transmet mal la flamme mais sa combustion entraîne le dégagement de gaz corrosifs et de fumées. Ce dégagement peut être réduit par l'emploi d'un mélange de formulation spéciale.
- Le PVC a un bon vieillissement thermique. Sous l'action prolongée de la chaleur, il devient jaune puis brun, puis rouge et enfin noir et forme en général une intumescence.
- Le PVC a une résistance exceptionnelle à l'ozone. Il possède une résistance moyenne aux hydrocarbures aliphatiques, aux acides et bases.
- Le PVC résiste au fluage à température élevée mais il est fragile aux basses températures (demande de précautions pour la pose à temps très froid).
- Le PVC rigide résiste bien aux agents chimiques inorganiques. Il est soluble dans les cétones, le cyclohexanone et le dichloroéthylène .
- La température de transition vitreuse du PVC est élevée, elle se situe entre 75 et 80°C, ce qui le rend impropre à de nombreuses applications (câbles par exemple) sans l'adjonction de plastifiants. Avec l'emploi de plastifiants, la température de transition vitreuse se situe entre -10°C et -5°C.

b) Propriétés électriques

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

- Le facteur de pertes diélectriques à 50Hz et 20°C est de l'ordre de : $\text{tg } \delta = 0.015$.
- La constante diélectrique à 50 Hz et 20°C est de l'ordre de : $\epsilon_r = 3.4$.
- La résistance transversale est de l'ordre de : $\rho_t \approx 10^{16} \Omega \cdot \text{cm}$ à 20°C pour les produits non plastifiés, $\rho_t \approx (10^{10} \text{ à } 10^{15}) \Omega \cdot \text{cm}$ à 20°C pour les produits plastifiés.

c) Propriétés mécaniques

Le PVC a une bonne résistance à la traction de l'ordre de : 5.5 à 6 kgf/mm² à 20⁰C.

- L'allongement à la rupture est de l'ordre de : 100% pour les produits non plastifiés, et 200 à 400% pour les produits plastifiés.

- Au dessus de 140 à 150⁰C, le PVC perd toutes ses propriétés mécaniques [Nej-91].

2.2.2.3 Résultats expérimentaux du vieillissement thermique du PVC

Nous présentons dans ce paragraphe les caractéristiques d'endurance thermique du Polychlorure de Vinyle plastifié obtenues aux laboratoire de HT de l'ENICAB (actuelle CABEL) [Nej-91]. Ces caractéristiques ont été obtenues à partir des variations, en fonction du temps de vieillissement à différentes températures, des propriétés mécaniques. Nous présentons également l'indice de température et mettons en valeur les conséquences de la dégradation du matériau: perte de masse, contraction et changement de couleur.

Les essais de vieillissement thermique ont été effectués par M. Nedjar [Nej-91] sur des éprouvettes soumises à un champ électrique sous haute tension à différentes températures: 80⁰C, 100⁰C, 120⁰C et 140⁰C. Les caractéristiques expérimentales que nous donnerons seront celles utilisées dans le chapitre 4 pour la prédiction du vieillissement thermique.

a) Variation de l'allongement à la rupture en fonction du temps de vieillissement

Pour les températures de 80⁰C et 100⁰C, le PVC présente une augmentation, puis une diminution de l'allongement à la rupture, en fonction du temps de vieillissement. Pour les autres températures, nous n'avons constaté qu'une diminution de ce facteur (Fig.2.5). L'augmentation de l'allongement à la rupture du polymère pourrait être interprétée par un effet de réticulation, et sa diminution par une rupture de chaînes carbonées et une perte de plastifiant [Bou9-97].

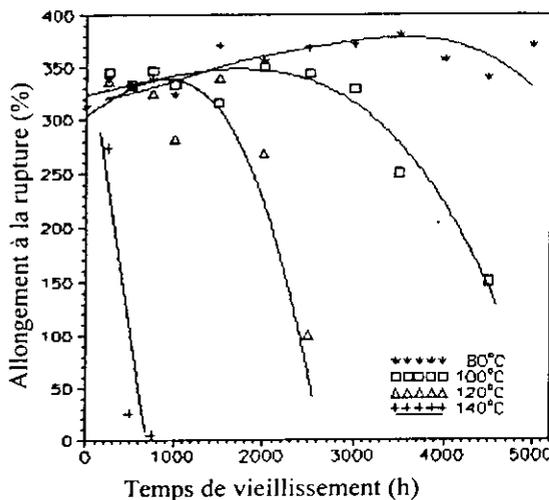


Fig.2.5 Variation de l'allongement à la rupture en fonction du temps de vieillissement.

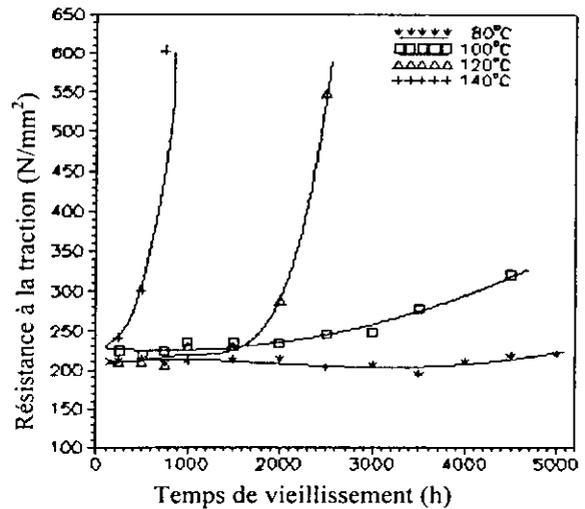


Fig.2.6 Variation de la résistance à la traction en fonction du temps de vieillissement.

b) Résistance à la traction

Pour cette propriété, nous constatons un effet inverse par rapport à l'allongement à la rupture: lorsque ce dernier décroît, la résistance à la traction croît (Fig.2.6). Pour les températures de 120⁰C et 140⁰C, la résistance à la traction croît rapidement en fonction du temps de vieillissement (Fig.2.6).

c) Endurance thermique

En extrapolant les caractéristiques de l'allongement à la rupture, les durées de vie du polymère, relativement aux différentes températures de vieillissement, peuvent être déterminées en supposant un critère de fin de vie de 50% de la valeur initiale de la propriété considérée (Fig.2.7 et Fig.2.8).

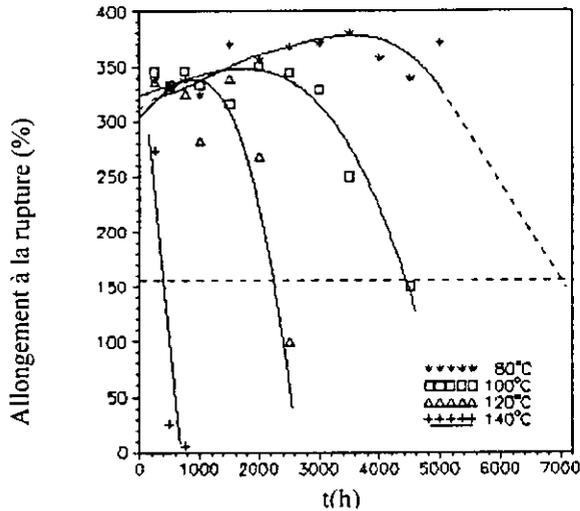


Fig. 2.7 Détermination de la durée de vie à partir des caractéristiques de l'allongement à la rupture.

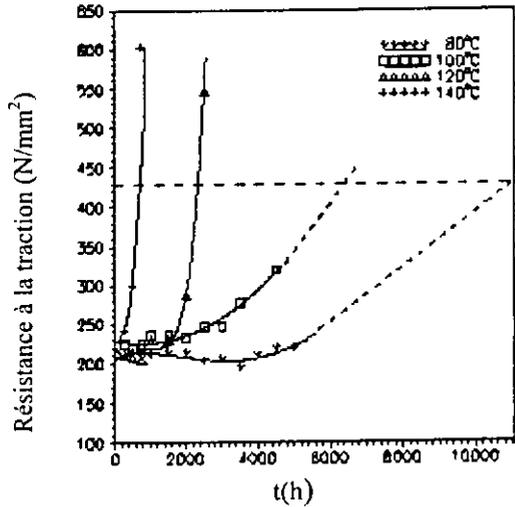


Fig. 2.8 Détermination de la durée de vie à partir des caractéristiques de la résistance à l'attraction.

La figure 2.9 montre le graphe d'endurance thermique correspondant simultanément aux deux propriétés.

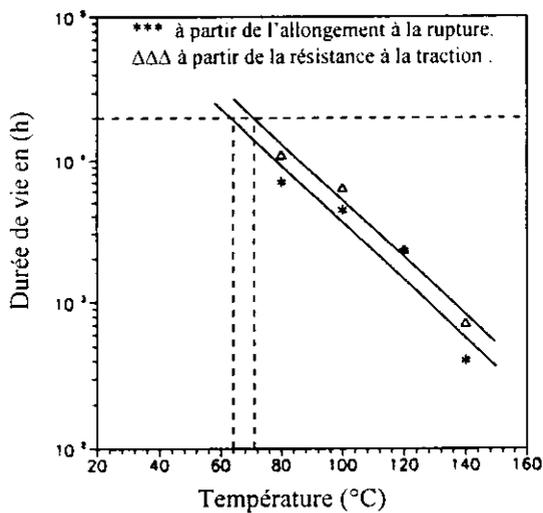


Fig. 2.9 Graphe d'endurance thermique.

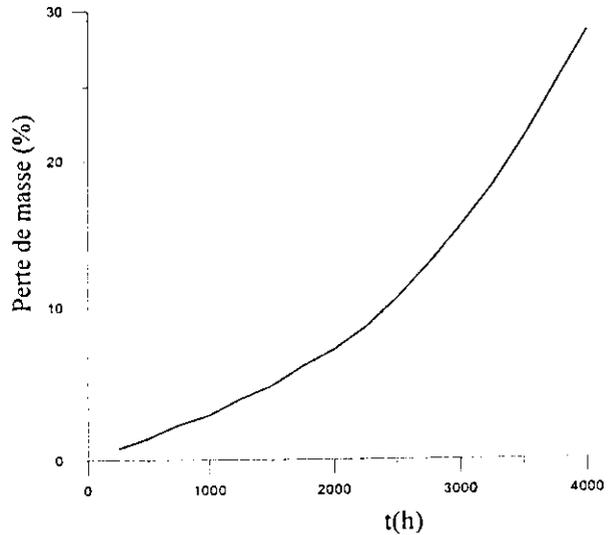


Fig. 2.10 Variation de la perte de masse surfacique en fonction du temps de vieillissement.

d) Perte de masse

D'après M. Nedjar [Nej-91], lorsque le temps de vieillissement augmente, la perte de masse augmente. Elle peut atteindre 30% après 4000h de vieillissement (Fig.2.10).

2.3 PROPRIETES ET CARACTERISTIQUES DE VIEILLISSEMENT DE L'HUILE DE TRANSFORMATEUR

Le choix d'une huile isolante est lié à certaines propriétés telles que les propriétés physiques, les propriétés chimiques et les propriétés électriques [Vua1-90].

2.3.1 Propriétés électriques

2.3.1.1 Permittivité

La permittivité définit la possibilité, sous l'action du champ électrique, de libérer des charges. Elle résulte de diverses contributions de déplacements de charges. La permittivité relative ϵ_r est définie comme le rapport ϵ/ϵ_0 de la permittivité absolue ϵ du liquide à celle du vide. La permittivité relative des huiles varie peu avec la fréquence.

2.3.1.2 Conductivité et facteur de pertes diélectriques

Tout liquide isolant, soumis à une différence de potentiel, donne lieu à un courant de fuite caractérisant la conductivité électrique σ_c du fluide. En pratique, pour mieux caractériser un diélectrique du point de vue pertes, il est usuel d'utiliser la tangente de l'angle de pertes $\text{tg}\delta$. C'est une fonction qui augmente avec la température et dépend de la tension appliquée.

2.3.1.3 Rigidité diélectrique

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

Dans l'industrie, elle est maximisée conventionnellement par la tension efficace nécessaire pour provoquer, dans les conditions normalisées, le claquage d'un volume de liquide compris entre deux électrodes dont la forme, distance et nature sont spécifiées, d'où son nom de « tension de claquage ».

La rigidité diélectrique n'est pas un critère de qualité de fabrication des huiles isolantes, mais un essai conventionnel destiné à déceler leur état de pollution physique par l'eau et d'autres matières en suspension et à conclure ou non de faire un traitement de déshydratation ou filtration avant introduction dans les appareils auxquels elles sont destinées.

2.3.2 Propriétés physiques

2.3.2.1 Viscosité

Les contraintes de cisaillement ou de glissement qui peuvent s'exercer dans un liquide provoquent des réactions du milieu, proportionnelles au gradient de vitesse; le coefficient de proportionnalité est la viscosité absolue η .

La viscosité dynamique s'exprime en pascals-secondes. En pratique, la viscosité d'un liquide se détermine par l'écoulement capillaire d'un volume de liquide sous son propre poids. La viscosité cinématique est donnée par :

$$v_s = \eta / \rho_m \quad (\text{II.13})$$

ρ_m est la masse volumique.

La viscosité diminue avec la température; le nombre de lois empiriques dépasse la cinquantaine et l'on retiendra la loi de Vogel, Fulcher, Tamman et Hesse (relation VFTH), qui donne la viscosité absolue [Vua1-90, Tob-86] :

$$\ln(\eta) = A + [B/(T - T_0)] \quad (\text{II.14})$$

$$\left\{ \begin{array}{l} A \text{ et } B \text{ deux constantes positives caractéristiques des liquides,} \\ T (k) \text{ température thermodynamique,} \\ T_0 = 273,15 \text{ K .} \end{array} \right.$$

Elle est utilisée dans l'étude des phénomènes de transport (conduction électrique) et de relation (pertes diélectriques). La viscosité dynamique d'un liquide augmente avec la pression P et correspond à une relation de la forme :

$$\ln(\eta_p) = \ln(\eta) + \kappa P \quad (\text{II.15})$$

κ est un coefficient qui dépend de η .

L'expression montre que, dans le cas d'une huile minérale, la viscosité double quant la pression passe de la pression atmosphérique à 350 bars.

2.3.2.2 Point d'éclair

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 suite de défauts électriques ou thermiques.

2.3.2.3 Point d'écoulement

Le point d'écoulement, température à laquelle les liquides se figent, est déterminé suivant la norme ISO 3016. Il est défini comme la température la plus basse pour laquelle un liquide peut couler lorsqu'il est refroidi suivant des conditions fixées. Lorsqu'un liquide isolant est refroidi, il acquiert une certaine consistance qui n'est pas la congélation correspondant au passage de l'état liquide à l'état solide d'un corps pur à température fixe.

La notion de point d'écoulement ne sert pas à l'évaluation des performances d'un liquide à basse température. En particulier, dans les transformateurs avec circulation forcée ou dans les disjoncteurs, on considère que les pompes et les contacts mobiles peuvent fonctionner pour des viscosités cinématiques atteignant $850 \text{ mm}^2/\text{s}$.

2.3.2.4 Le point de feu

C'est le point d'inflammation de l'huile. Il indique la température minimale pour laquelle la combustion des vapeurs d'un liquide sera entretenue.

2.3.2.5 La couleur

La coloration d'une huile neuve est une indication de son degré de raffinage : les huiles très raffinées sont blanches. Une 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 une détérioration.

2.3.3 Propriétés chimiques

2.3.3.1 L'acidité

L'acidité d'une huile peut avoir des origines diverses ; elle est minérale, et peut provenir des opérations de raffinage et organique dite de pétrole. 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[mg KOH/g].

2.3.3.2 Teneur en eau

Les propriétés des liquides sont affectées par la teneur en eau. La quantité d'eau dissoute dans un liquide dépend de sa nature. Avant la saturation et à l'état de vapeur, sa teneur en eau 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 100 %, le liquide se sature en eau [Vua2-86].

2.3.3.3 Oxydation des huiles minérales isolantes

On distingue par oxydation, un ensemble de réactions chimiques complexes et lentes au cours desquelles les hydrocarbures réagissent avec l'oxygène. Il se produit tout d'abord des produits de décomposition solubles (aldéhydes, cétones, acides organiques), puis par condensation et polymérisation, des dépôts insolubles. Les conséquences de l'oxydation de l'huile pour les isolations sont :

- Une mauvaise évacuation de la chaleur par suite de l'accroissement de la viscosité et de la présence de dépôts.
- Une diminution des propriétés électriques de l'huile (la conductivité et les pertes diélectriques augmentent).
- L'apparition de l'acidité dans l'huile.

2.3.4 Caractéristiques de service

Les caractéristiques de service correspondent généralement aux propriétés chimiques ou réactionnelles des liquides propres à chaque catégorie.

La première caractéristique de service est l'aspect des liquides, qui concerne le constructeur de matériel, mais davantage l'utilisateur pour qui le contrôle visuel (si cela est possible) est une indication immédiate de l'état du liquide en fonctionnement. Un liquide trouble peut être l'indication d'une saturation en eau [Vua2-86].

2.3.4.1 Aspect, couleur et caractéristiques optiques

- Les normes commencent généralement par définir l'aspect et spécifient simplement que les liquides, examinés en lumière transmise sous une épaisseur spécifiée, doivent être limpides et exempts de matière en suspension ou de dépôts. L'absence de matière en suspension ne peut être jugée à l'œil ; elle fait appel au comptage de particules que pratiquent les constructeurs après les filtrations nécessaires à la bonne utilisation des huiles.
- La couleur est aussi un paramètre souvent considéré. En général, les liquides isolants sont incolores ; cependant, certains, comme les huiles minérales, ont une couleur ambrée ou une fluorescence caractéristique. La couleur traduit l'existence de groupements fonctionnels ou structures chromophores, présentant des bandes

Mécanismes diélectriques dans les matériaux isolants

d'absorption caractéristiques dans l'ultraviolet. Le changement de coloration d'un liquide en service peut provenir de l'oxydation, de diverses réactions ou de la mise en solution de pigments de peinture, de vernis, etc., mal choisis.

- Les propriétés optiques présentent un intérêt particulièrement grand, car elles peuvent caractériser la constitution d'un liquide et un spectre infrarouge peut servir à l'identifier ou à pénétrer dans sa structure.

2.3.4.2 Stabilité chimique

Elle 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 (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, tous 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 du TAN, forment des dépôts insolubles ou des boues d'oxydation qui se déposent sur les isolants solides, réduisant l'évacuation des pertes ; l'huile est régénérée. Les liquides de synthèse ne forment généralement pas de dépôts.

2.3.4.3 Stabilité électrique aux décharges partielles

Sous l'action d'un champ électrique intense et par l'intermédiaire de décharges partielles dans des occlusions gazeuses qui se forment ou préexistent, les liquides se décomposent en donnant des gaz. La stabilité électrique des liquides correspond à la propriété d'absorber ou non ces gaz à l'interface gaz-liquide, sous ionisation.

- Des gaz peuvent se former par l'action momentanée d'un gradient de potentiel élevé (surtension, choc électrique), provoquant un claquage partiel et l'apparition de produits gazeux de dégradation du liquide. Les décharges partielles se poursuivent après retour à la valeur nominale de la tension. Leur extinction ne peut avoir lieu que si le liquide absorbe les gaz créés ; sinon, elles s'amplifient jusqu'au claquage.
- L'interface gaz-liquide sous décharges partielles peut aussi se former par l'apparition de gaz produit par décomposition du liquide et dû par exemple à des forces d'origine mécanique, à des turbulences, à des prises d'air qui, combinées au champ électrique, libèrent des gaz dissous, et initient, dans des zones à fort champ, le phénomène d'ionisation. Ceci peut se rencontrer dans l'isolation compacte des transformateurs. La phase gazeuse est généralement de l'air. A température fixée, la quantité de gaz produite augmente avec la tension.

2.3.5 Sécurité d'emploi

La sécurité d'emploi des liquides isolants concerne les risques d'incendie et d'explosion, la santé des individus et les risques pour l'environnement. Le danger

Mécanismes diélectriques dans les matériaux isolants

d'incendie repose sur des caractéristiques d'inflammabilité (point d'éclair, point de feu, température d'auto inflammation) et certaines caractéristiques de combustion (indice d'oxygène, quantités de chaleur dégagées).

Le danger d'explosion est lié à la nature des gaz produits par décomposition des liquides sous arcs électriques et points chauds ; hydrogène, méthane et acétylène étant les principaux gaz concernés.

2.3.6 Propriétés de la BORAK22

Nous présentons les six propriétés utilisées dans les tests physico-chimiques de la BORAK22, réalisés au laboratoire de Contrôles, Mesures et Essais d'El-Hamma /SONELGAZ. Les tableaux présentés donnent les valeurs de chaque propriété avant et après vieillissement [Ham-97].

2.3.6.1 Rigidité diélectrique

Nous remarquons qu'après vieillissement, la tension de claquage diminue pour $t=120^{\circ}\text{C}$. Cela est dû à la vaporisation de l'eau sous température élevée, ce qui n'est pas le cas pour l'huile vieillie à 80°C (Tab.2).

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

Tab. 2 Tension de claquage de l'huile avant et après vieillissement

2.3.6.2 Acidité

Nous constatons que l'indice d'acidité augmente avec le vieillissement de l'huile ; ceci est lié aux conditions suivantes : température de vieillissement, pression, contact avec l'air, présence d'eau. Ces conditions favorisent les réactions d'oxydation, qui donnent naissance à des produits acides (Tab.3).

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

Tab. 3 Indice d'acidité de l'huile avant et après vieillissement.

2.3.6.3 Teneur en eau

Le vieillissement thermique de l'huile a provoqué une augmentation de la teneur en eau, qui est passée de 30ppm pour l'huile neuve, à 31 ppm pour l'huile vieillie à 80°C , et à 35ppm pour l'huile vieillie à 120°C (Tab.4).. La présence de cette eau peut être la conséquence d'une oxydation de l'huile selon la réaction suivante :



Mécanismes diélectriques dans les matériaux isolants

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

Tab. 4 Teneur en eau de l'huile avant et après vieillissement.

2.3.6.4 Couleur

L'évolution de l'indice de couleur de pour l'huile vieillie à 120°C est due à l'oxydation de l'huile en service qui a pour conséquence la formation des produits acides.

2.3.6.5 Viscosité cinématique

La viscosité cinématique est presque constante durant le vieillissement. La pollution de l'huile a pour effet de ralentir l'écoulement de l'huile.

2.4 VIEILLISSEMENT DES HUILES DE TRANSFORMATEURS

Tout au long de leur utilisation, les huiles minérales isolantes peuvent être soumises à plusieurs contraintes électriques, thermiques, chimiques...etc.

Sous l'action combinée ou séparée de ces contraintes pendant la mise en service des transformateurs, les huiles isolantes se dégradent au cours du temps.

L'étude du vieillissement thermique de l'huile de transformateur à deux buts principaux:

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

2.4.1 Influence de la température sur les propriétés de la BORAK22

Nous présentons quelques résultats d'essais concernant l'effet de la température sur les propriétés de la BORAK22, obtenus par A. Hamimi [Ham-97] au laboratoire de Contrôles, Mesures et Essais (CME) de la SONELGAZ/Hamma dans le cas d'une huile neuve, usagée et régénérée [Ham-97].

2.4.1.1 Pertes diélectriques

Pour une huile neuve, usagée ou régénérée, le facteur de pertes diélectriques augmente en exponentielle en fonction de la température (Fig.2.11). Cette évolution est due en même temps à la réduction de 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. Nous remarquons qu'une régénération de l'huile usagée engendre une nette amélioration du facteur des pertes diélectriques.

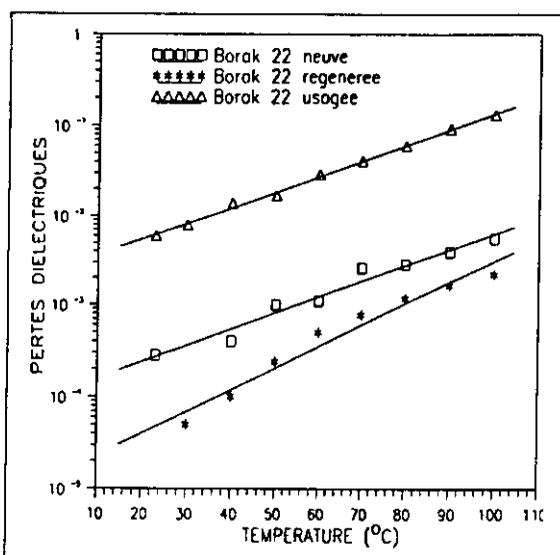


Fig. 2.11 Variation de $\text{tg}(\delta)$ en fonction de la température

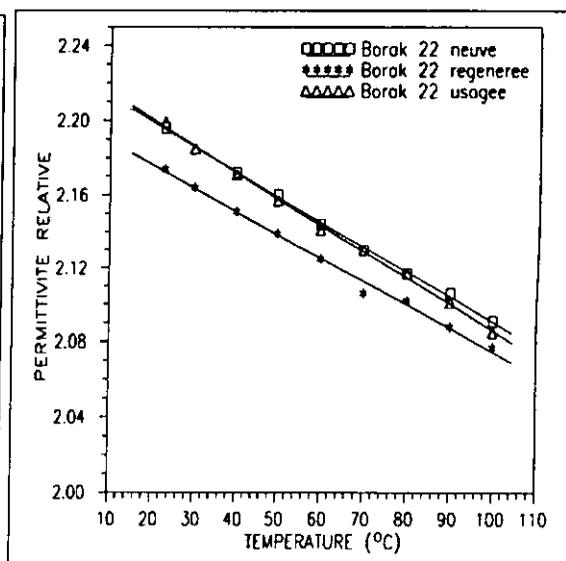


Fig. 2.12 La permittivité relative en fonction de la température

2.4.1.2 Permittivité relative

Nous remarquons que la permittivité relative décroît linéairement avec la température dans tous les cas de type d'huile, et les permittivités de l'huile neuve et de l'huile usagée sont confondues. Ceci permet de dire que la permittivité est un paramètre très affecté par l'usage (Fig. 2.12).

2.4.1.3 Résistivité transversale

La résistivité ρ_R décroît en fonction de la température T selon la loi de type :

$$\ln(\rho_R) = A.T \quad (\text{II.16})$$

où A est une constante.

Une forte amélioration de cette propriété est obtenue après la régénération de l'huile (Fig. 2.13).

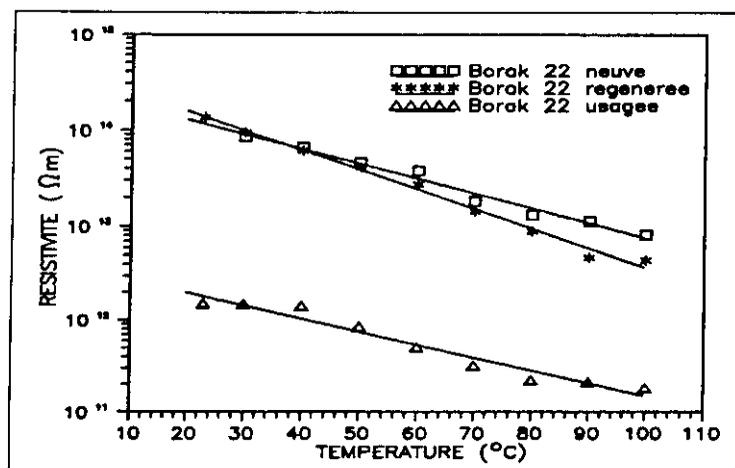


Fig. 2.13 Variation de la résistivité en fonction de la température

2.4.1.4 Viscosité

Nous remarquons que quel que soit le type d'huile, la viscosité diminue quand la température augmente (Fig.2.14), et l'huile usagée ne retrouve pas sa viscosité après sa régénération.

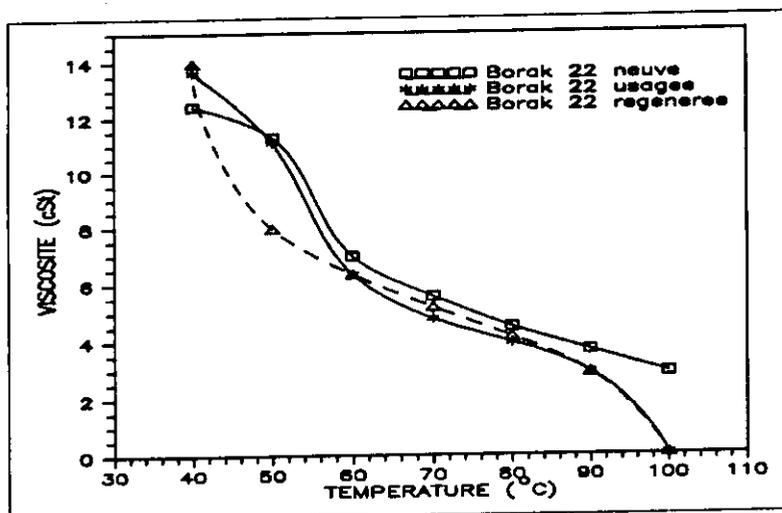


Fig. 2.14 Variation de la viscosité en fonction de la température

2.4.2 Influence du vieillissement thermique sur les propriétés de l'huile

Nous présentons dans ce paragraphe quelques résultats sur l'évolution du facteur de pertes diélectriques et de la permittivité de la BORAK22 sous l'influence du vieillissement thermique. Les essais de vieillissement ont été effectués par A. Abdi [Abd-02] sous deux températures (80°C et 120°C) au laboratoire CME de la SONELGAZ/Hamma-Alger, pour une durée de vieillissement de 1500 heures [Abd-02]. Les mesures ont été faites pour deux températures d'essais : 50°C et 110°C.

2.4.2.1 Facteur de pertes diélectriques

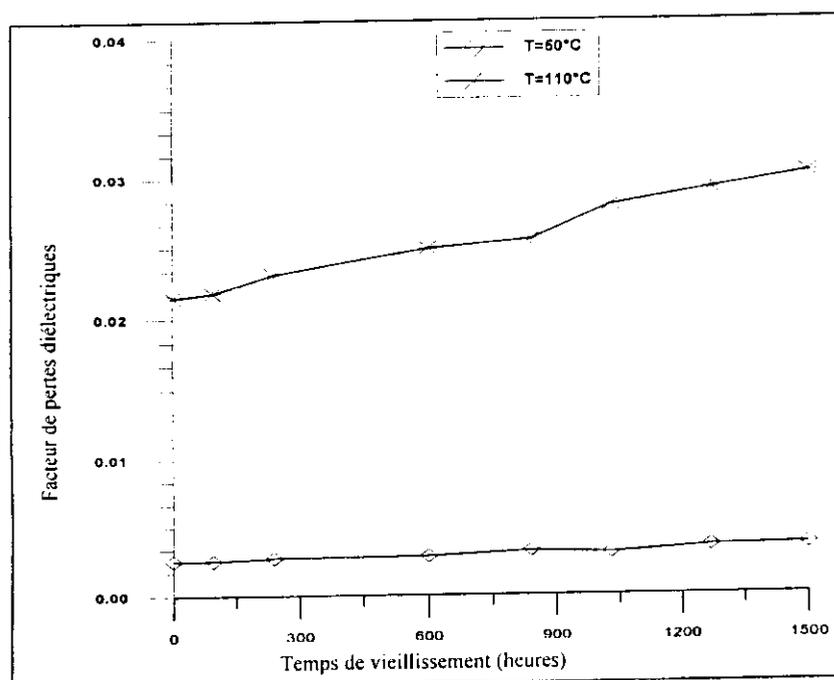


Fig. 2.15 Facteur des pertes diélectriques de la BORAK22 durant le vieillissement à T=80°

Mécanismes diélectriques dans les matériaux isolants

Après 1500 heures de vieillissement sous deux températures 80°C et 120°C, les résultats concernant la variation du facteur des pertes diélectriques $\text{tg}(\delta)$ en fonction du temps de vieillissement sont présentés aux figures (2.15) et (2.16).

D'après ces deux caractéristiques, nous remarquons que le facteur de pertes diélectriques $\text{tg}\delta$ augmente au cours du vieillissement thermique. Le facteur des pertes diélectriques augmente plus rapidement dans le cas de la température $T=120^\circ\text{C}$ que dans le cas, où $T=80^\circ\text{C}$. Cette forte augmentation s'explique par la formation des produits d'oxydation accélérée par des températures élevées au cours du vieillissement.

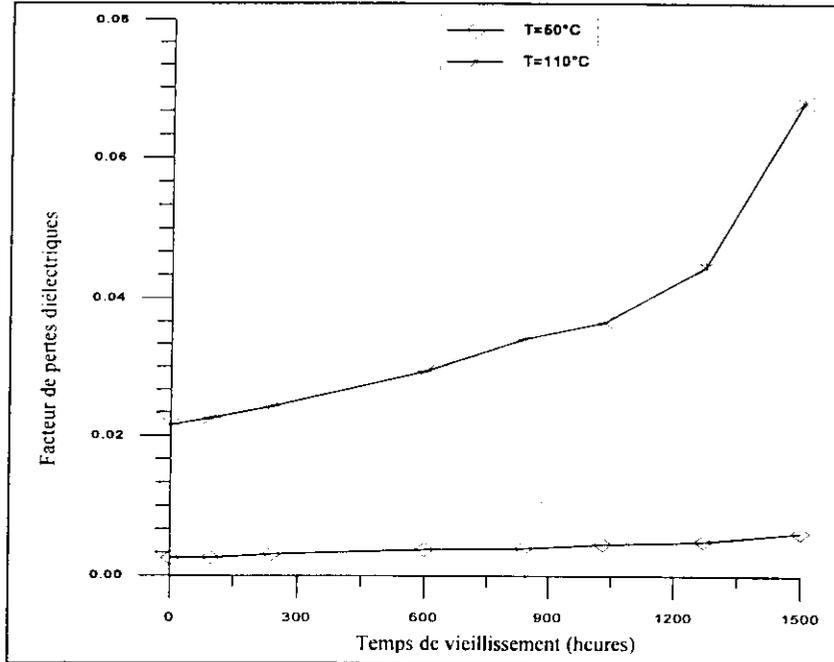


Fig. 2.16 Facteur des pertes diélectriques de la BORAK22 durant le vieillissement à $T=120^\circ\text{C}$

2.4.2.2 Permittivité relative (ϵ_R)

Les résultats concernant les variations de la permittivité relative pour des vieillissements thermiques de 80°C et de 120°C sont présentés respectivement aux figures (2.17) et (2.18).

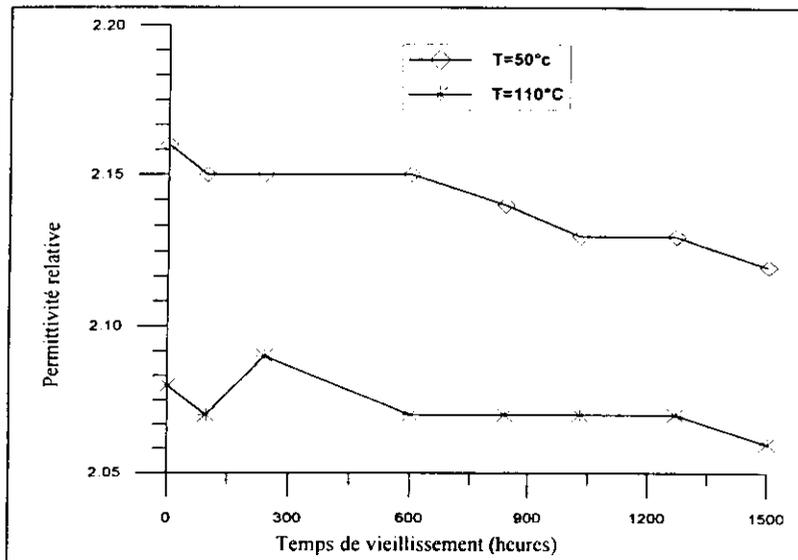


Fig.2.17 Evolution de la permittivité relative (ϵ_r) de la BORAK22 durant le vieillissement à $T=80^\circ\text{C}$

Mécanismes diélectriques dans les matériaux isolants

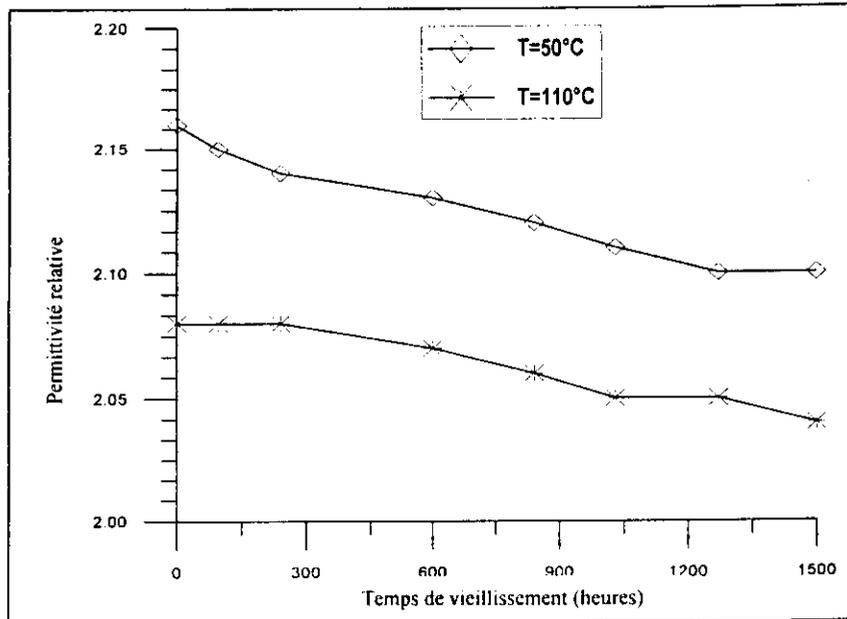


Fig. 2.18 Evolution de la permittivité relative (ϵ_r) de la BORA K22 durant le vieillissement à $T=120^\circ\text{C}$

D'après ces deux figures, nous remarquons que la permittivité relative ϵ_r subit une variation relativement faible durant le vieillissement.

2.5 RUPTURE DES INTERVALLES D'AIR POINTE-BARRIERE-PLAN

L'avantage fourni par l'utilisation des barrières est l'augmentation de la rigidité de l'air. Pour les intervalles relativement petits (<30cm), l'influence des barrières sur la rupture diélectrique de l'air fût déjà remarquée vers les années 1920 par C. P. Steinmetz [Bou8-79]. Les principales recherches furent entreprises ensuite par E. Marx et H. Roser et N.F Wolochtchenko [Bou8-79, Wol-46]. Ces auteurs ont étudié l'influence de différents types de barrières dans différents intervalles d'air à champ non-uniforme aux tensions de choc à front raide, à tension continue et à fréquence industrielle.

L'amélioration serait due surtout à l'uniformisation du champ dans l'espace barrière-plan due à la charge positive déposée sur la surface de la barrière en face de la pointe [Mee2-66]. Cette constatation fut vérifiée ensuite par N.F Wolochtenko et M.P. Verma [Wol-46]. Pour les longs intervalles d'air, les recherches sont rares et l'augmentation de la tension disruptive est obtenue dans le cas des systèmes à champ non uniforme avec l'électrode de faible rayon de courbure sous polarité positive [Dar1-79, Dar2-77].

Dans des travaux de simulation par éléments finis de l'effet de la barrière dans un intervalle d'air pointe-plan [Bou10-98, Bou11-99, Mok27-97, Mok28-97, Mok29-98, Mok32-00, Mok33-00, Mok34-03, Mok34-97], le modèle élaboré a permis de déterminer le champ et le potentiel électrique en présence de la charge d'espace qui est déterminée à son tour après la connaissance du facteur efficace d'ionisation $\bar{\alpha}$. La dépendance de ce facteur du champ électrique E est connue avec une marge d'erreur acceptable à partir d'une base de données ($\bar{\alpha} = f(E)$) communiquée par E. Kuffel et W. Zaengle [Kuf-84].

Le calcul du champ électrique en utilisant les éléments finis a été fait pour une pointe hyperboloïdale en changeant différents paramètres de l'arrangement pointe-barrière-plan. Les résultats et les données expérimentales obtenues par A. Boubakeur [Bou8-79] étaient cohérents.

Ces mêmes données utilisées dans la simulation avec éléments finis sont utilisées dans le chapitre 4 pour entraîner les réseaux de neurones utilisés à prédire la tension disruptive. Nous donnons à la figure 2.19 l'influence de la barrière sur la tension 50% de décharge disruptive de l'intervalle d'air pointe-plan, au choc de manœuvres (+170/2500 μ s).

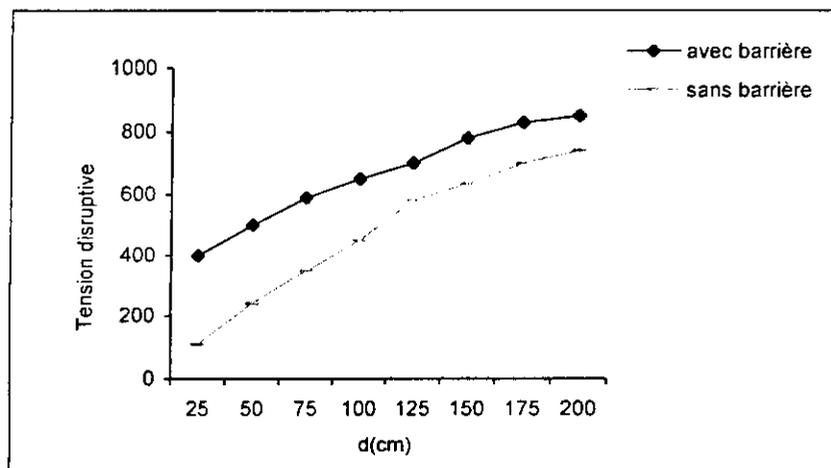


Fig.2.19 Influence de la barrière sur la tension 50% de décharge disruptive d'un intervalle d'air pointe positive-plan sous une tension de choc de manœuvres (170/2500 μ s)

2.5.1 Influence de la barrière isolante sur la tension de décharge disruptive

D'après les résultats expérimentaux [Bou8-79,] et respectivement les résultats théoriques et ceux obtenus par simulation par éléments finis,[Ber-91, Bou10-98, Bou11-99, Mok27-97, Mok28-97, Mok29-98, Mok32-00, Mok33-00, Mok34-03, Mok34-97] l'introduction d'une barrière dans un intervalle d'air pointe-plan uniformise le champ électrique dans l'intervalle barrière-plan. Plusieurs paramètres liés à l'effet de cette uniformisation du champ et que nous exposons brièvement, influent sur la rigidité de l'arrangement pointe-barrière-plan.

2.5.1.1 Influence de la position de la barrière dans l'intervalle d'air

La tension de décharge passe par un maximum pour une position de la barrière à 20% de l'intervalle à partir de la pointe pour des largeurs définies [Bou8-79]. Cette augmentation est due à l'allongement du canal de la décharge (Fig.2.20-2.22) (a étant la distance pointe-barrière. Pour le cas de la tension à fréquence industrielle, l'augmentation de la tension disruptive est de 135% pour la distance de 40 cm et 40% pour $d = 120$ cm.

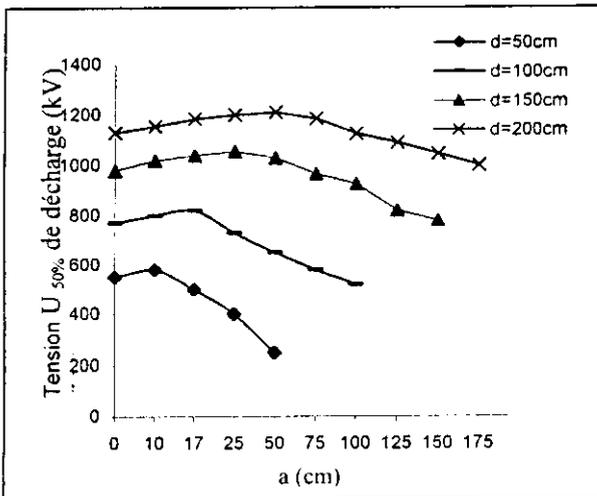


Fig. 2.20 Influence de la position de la barrière sur la rigidité de l'intervalle sous une tension de choc de foudre 1.2/50 µs

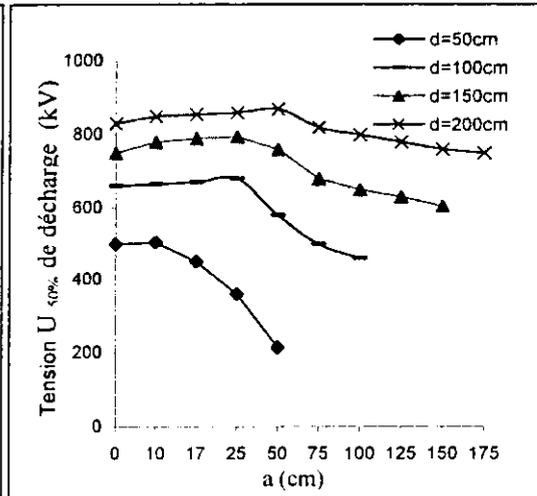


Fig. 2.21 Influence de la position de la barrière sur la rigidité de l'intervalle sous une tension de choc de manœuvres 170/2500 µs

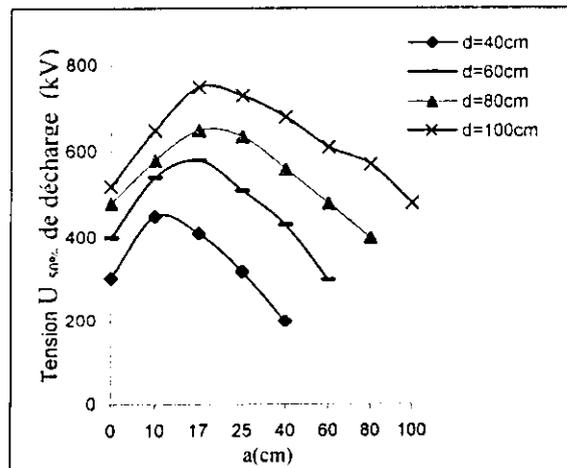


Fig. 2.22 Influence de la position de la barrière sur la rigidité de l'intervalle sous une tension à fréquence industrielle

2.5.1.2 Influence de la largeur de la barrière

Quand la largeur de la barrière n'est pas aussi grande que la distance inter électrodes, la rigidité de l'arrangement total dépend de la longueur du canal de la décharge. Cette longueur est presque égale à la distance pointe-bord de la barrière-plan (Fig.2.23).

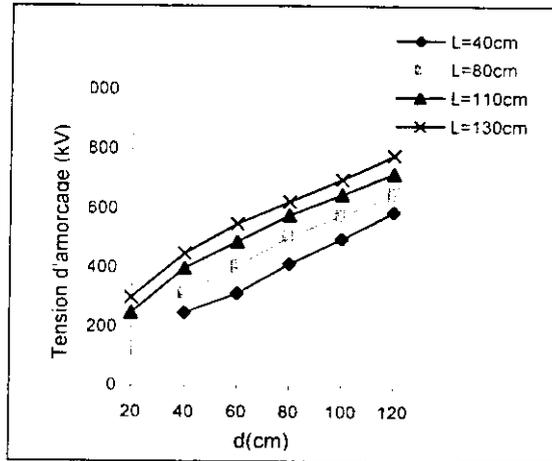


Fig. 2.23 Influence de la largeur de la barrière sur la rigidité de l'intervalle d'air à 50 Hz

2.5.1.3 Influence de la forme de la tension

Avec la polarité négative, la décharge se déroule par étapes pour toutes les positions de la barrière et aucune augmentation de la tension disruptive n'a été obtenue. Pour ce cas, les décharges apparaissent des deux cotés de la barrière. Les résultats donnés à la figure 2.24 montrent l'influence du type de la tension appliquée (tension de choc de manœuvres, de foudre, tension à fréquence industrielle) sur la tension disruptive de l'intervalle d'air.

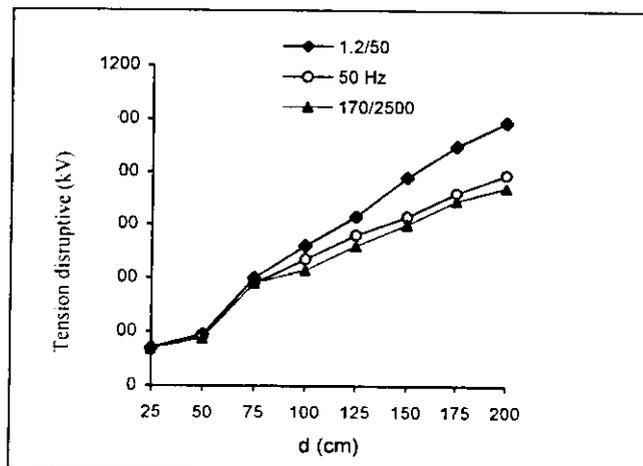


Fig.2.24 Influence du type de la tension appliquée sur la rigidité d'un intervalle d'air pointe-barrière-plan

2.5.1.4 Influence de la conductivité de la barrière

Avec une barrière à surface semi-conductrice, la tension disruptive des intervalles d'air pointe-plan varie entre les valeurs obtenues avec la barrière isolante à surface propre et

celles obtenues avec une barrière métallique [Bou8-79]. Avec cette dernière, la tension disruptive de l'intervalle d'air a la valeur la plus petite (Fig.2.25).

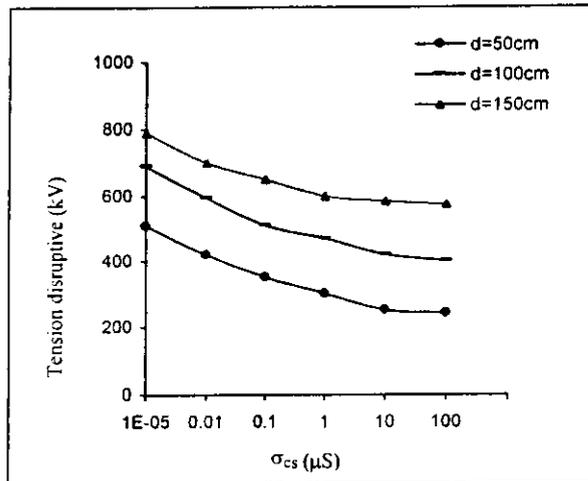


Fig.2.25 Influence de la conductivité de la barrière sur la rigidité de l'intervalle d'air sous un choc de manœuvres (+170/2500 μs)

2.5.1.5 Influence de la longueur de l'intervalle

Pour les petits et moyens intervalles, la rigidité diélectrique de l'air s'améliore surtout pour des positions relativement loin des deux électrodes [Bou8-79]. Pour les longs intervalles et pour des positions de la barrière équivalentes à celles dans les cas des petits et moyens intervalles, l'effet de la barrière n'est plus le même (Fig.2.26).

Les variations de la tension disruptive à 50 Hz en fonction de la distance pointe-plan (d) pour les différentes positions de la barrière a/d , les tensions disruptives restent très proches des valeurs obtenues avec le système d'électrodes sans barrière. Pour les différentes positions a/d de la barrière, nous constatons qu'aux distances d pointe-plan supérieures à 40 cm, la pente des caractéristiques reste très proche de celle de la caractéristique tension-distance du système pointe-plan sans barrière.

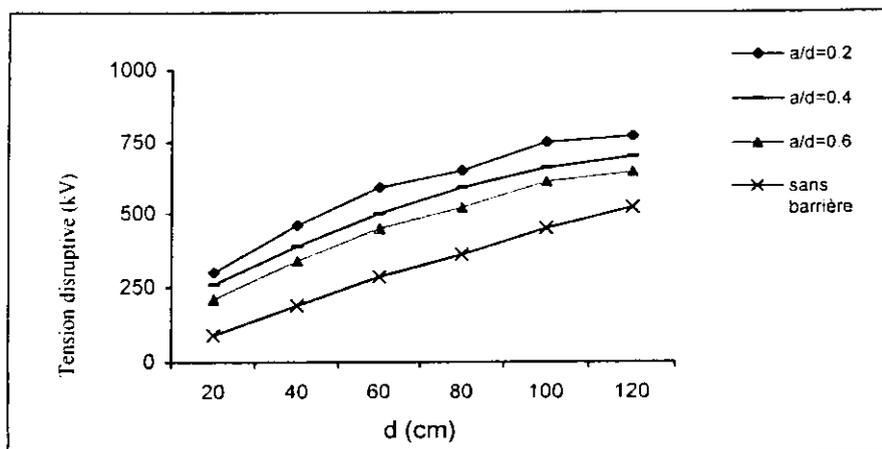


Fig. 2.26 Influence de la longueur de l'intervalle d'air sur la rigidité à 50 Hz pour différentes positions de la barrière

2.6 CONCLUSION

L'étude du phénomène de vieillissement thermique des isolants permet d'évaluer la durée de vie de ces isolants sous l'effet des contraintes thermiques et de trouver une éventuelle corrélation entre le processus de vieillissement et les contraintes qui le provoquent. Cette corrélation pourra nous informer sur l'état de l'isolant après un temps et une contrainte donnés. Ainsi, une bonne maintenance des isolants exige une connaissance exacte de leurs propriétés.

Les essais de vieillissement ont pour objectif de définir les variations des propriétés de l'isolant sous l'effet des contraintes appliquées durant une longue période, de préférence jusqu'à la dégradation définissant la limite de durée de vie. De plus, pour l'huile de transformateur, l'objectif est de définir une meilleure maintenance de cette huile. Le diagnostic de l'huile en exploitation pourra se faire en se basant sur ses propriétés physico-chimiques. Ceci sera présenté dans le prochain chapitre.

Le recours, pour des raisons économiques, à des tensions de plus en plus élevées pour le transport de l'énergie électrique a nécessité et nécessitera encore des distances d'isolement dans l'air de plus en plus grandes.

L'effet des barrières dans l'augmentation de la rigidité des systèmes pointe positive-plan, est connu depuis 80 ans. Il est exploité aussi dans la fabrication des câbles de haute tension pour augmenter leur durée de vie. Le dimensionnement des intervalles d'air pointe-barrière-plan exige beaucoup d'essais tenant compte de tous les paramètres influant sur la rigidité du système. Une bonne base de données de la tenue de ces intervalles d'air serait très utile pour l'interpolation des caractéristiques décrivant l'état du système d'étude.

2.7 BIBLIOGRAPHIE

- [Abd-02] S.Abdı, 'Influence du vieillissement thermique sur les propriétés de l'huile de transformateur', thèse de Magister, ENP, 2002.
- [Ber-91] A. Beroual, A. Boubakeur, 'Influence of Barriers on the lightning and switching impulse strength of mean air gaps in point/plane arrangements', IEEE Trans. On Electrical Insulation, Vol.26, N° 6, pp: 1130-1137, Dec. 1991.
- [Bou2-97] A.Boubakeur, M.Nedjar, 'Effect of thermal ageing on the electrical properties of PVC insulating cables', 5th international middle east power system conference MEPCON, Vol.2, pp: 699-702, Alexandria, Egypt, Jan. 1997.
- [Bou3-97] A.Boubakeur, A. Medjoub, M.Boumerzoug, 'Influence of thermal ageing on the properties of cross-linked polyethylene used as electrical insulation of medium voltage cables', 10th ISH, paper 3054, Montréal, Canada, Jun.1997.
- [Bou4-99] A.Boubakeur, Y.Mecheri, M.Boumerzoug, 'Influence of continous thermal ageing on the properties of XPLE used in Medium Voltage cables', 11th ISH, IEE conference Publication N°467, Vol.4, Session P2, pp: 236-239, London, UK, Sept. 1999.
- [Bou5-00] A. Boubakeur, Y. Mecheri, M. Boumerzoug, 'Comportement diélectrique et mécanique du polyéthylène réticulé chimiquement soumis à un vieillissement thermique continu', Ann.Chim.Sci.Mat, Elsevier, Paris, pp : 457-470, 2000.
- [Bou6-01] A.Boubakeur, L.Boukezzi, M.Lallouani, 'The Effect of Thermal Ageing Influence on the Electrical and Physical Behaviour of XLPE", Proceedings APDATM'2001, Special Issue of Przegląd Elektryczny, edited by SEP - Polish Electrical Society, pp.202-204, Wroclaw, Poland, 2001.

- [Bou7-01] A. Boubakeur, L. Boukezzi, M. Lallouani, D. Abdesselam, 'Thermal ageing influence on the electrical and mechanical behaviour of XLPE', 12th ISH'2001, Bangalore (India), Volume 4, paper No 6-40, Aug. 2001.
- [Bou8-79] A. Boubakeur, 'Influence des barrières sur la tension de décharge disruptive des moyens intervalles d'air pointe-plan', thèse de Doctorat, E.P Varsovie, 1979.
- [Bou9-97] A. Boubakeur, M. Nedajr et R. Khaili, 'Thermal endurance of polyvinyle chloride from mechanical properties', Congrès Arabe de Mécanique, Damas, Syrie, 1997.
- [Bou10-98] A. Boubakeur, L. Mokhnache, S. Boukhtache, 'theoretical investigation on barrier effect in point- plane air gap using finite elements method in presence of space charge' Int. conf. POWERCON'98, IEEE-proceedings, PP 116-121, Beijing, China, Aug. 1998.
- [Bou11-99] A. Boubakeur, L. Mokhnache, S. Boukhtache, ' Numerical Model of Electrostatic Barrier Effect in a Point-Plane Air Gap in Presence of Positive Space Charge', Archives of Electrical Eng., Polish Academy of Science (AEE), VOL. XLVIII N°189, 3./ 1999.
- [Bou12-03] A. Boubakeur, L. Mokhnache, A. Feliachi, 'Theoretical Investigation on Barrier Effect on Point- Plane Air Gaps Breakdown Voltage Based on Streamers Criterion' déposée pour publication dans IEE Proceedings - Science, Measurement and Technology, Fev.2003.
- [Dar1-79] M. Darveniza, B. Holcombe, 'The influence of a thin floating barrier on the switching impulse strength of a long air gap with a plastic sheet barrier', Proc. IEEE letters, Vol.64, nr 6, pp: 1017, 1979.
- [Dar2-77] M. Darveniza, B. Holcombe, 'The switching impulse strength of a long air gap with a plastic sheet barrier', Congrès Electrotechnique Mondial, section 2, rap. 104, Moscou, 1977.
- [Fou1-86] R. Fournié, 'Les isolants en électrotechnique, concepts et théories', Eyrolles, 1986.
- [Fou2-90] R. Fournié, ' Les isolants en électrotechnique, Essais, Mécanismes de dégradation, Applications industrielles', Eyrolles, 1990.
- [Gui-00] C. Guillermin 'Modèles de vieillissement des isolants électriques' rapport LEMD - Schneider Electric SA, guiller@labs.polycnrs-gre.fr, 2000.
- [Ham-97] A. Hamimi, A. Boubakeur, M. Lebid, 'Influence de la température sur certaines propriétés de l'huile de transformateur utilisée par la SONELGAZ', 2^{ème} CNHT pp : 102-104, Tizi-Ouzou, Nov. 1997.
- [Hoc-88] B. Hochard, 'Transformateur de puissance', Editions Eyrolles, 1988.
- [Kuf-84] E. Kuffel, W. Zaengle, 'High voltage engineering' Pergamon press, 1984.
- [Mee1-90] J.M Meek, J.D. Craggs, 'Electrical breakdown of gases', Wiley-Interscience, 1990
- [Mee2-66] J.M. Meek, C. Collins, 'Measurment of field changes predicting impulse breakdown of rod-plane gaps', Proc. Of the 7th Int. Conf. On Phenomena in ionised gazes, Vol.1, pp: 581, Beograd, 1966.
- [Mej-97] A. Mejdoub, 'Influence des vieillissements thermique et électrique sur les propriétés du Polyéthylène réticulé chimiquement utilisé dans les câbles de moyenne tension', Thèse de Magister ENP d'Alger 1997.
- [Min-99] M.S.A Minhas, J.P Reynders, P.J. de Klerk, 'Failures in power system transformers and appropriate monitoring techniques', 11th ISH, IEE conference, N° 467, pp: 1.94.S23-1.97.S23, Aug. 1999.
- [Mok27-97] L. Mokhnache, A. Boubakeur, S. Boukhtache ' Modèle numérique du calcul du champ électrique dans un système pointe-barrière-plan en présence de

- la charge d'espace', 2^{ème} conférence nationale sur la HT, CNHT'97, Tizi Ouzou, Algérie, Nov.97
- [Mok28-97] L. Mokhnache, A. Boubakeur, S. Boukhtache,'Calcul Numérique du champ et du potentiel dans un intervalle d'air pointe-barrière-plan en présence de la charge d'espace', 1st Int. Conf. Of Electronique, Electrotechnique & Automatique, IEEA'97, pp. 244-248, Batna-Algérie, Dec. 1997.
- [Mok29-98] L. Mokhnache, A. Boubakeur, S. Boukhtache,'Application of finite elements method to electric field calculation in point-barrier-plane arrangement in presence of space charge',4th International Workshop on Electric and Magnetic Fields, WEMF, pp 111- 116, Marseille-France, 1998.
- [Mok32-00] L. Mokhnache, A. Boubakeur, S. Boukhtache 'Numerical Model Describing the Effect(s) of a Barrier and the Space Charge Fields on the Electrical Strength of a Point-Plane Air Gap using the finite elements', pp 466-469, CEIDP'2000, proceedings IEEE, Victoria (BC)- Canada, Oct. 2000.
- [Mok33-00] L. Mokhnache, A. Boubakeur, S. Boukhtache 'Numerical Model Describing the Effect of a Barrier and the Space Charge Fields on the Electrical Strength of a Point-Plane Air Gap' UPEC'2000, Belfast, UK Sept. 2000.
- [Mok34-97] L.Mokhnache,'Contribution à l'étude de l'influence des barrières dans les intervalles d'air pointe-plan par le calcul numérique du champ à l'aide de la méthode des éléments finis avec et sans charge d'espace', thèse de Magister, Université de Batna.
- [Nas-90] E. Nasser,'Fundamentals of gaseous ionization and plasma electronics', Wiley-Interscience, 1990.
- [Nej-91] M.Nedjar,'Influence du vieillissement thermique sur les propriétés du polychlorure de Vinyle utilisé dans l'isolation des câbles de haute tension', Thèse de Magister Université de Tizi-Ouzou, 1991.
- [Roy-84] G.L. Roy, C. Gary & al.,'Les propriétés diélectriques de l'air et les très hautes tensions', Editions Eyrolles, 1984.
- [Tob-86] R.Tobazeon,'Liquides diélectriques, claquage et pré-claquage', Technique de l'ingénieur, D226, 1986.
- [Vua2-86] P.J.Vuarchex,'Huiles minérales isolantes', Techniques de l'ingénieur, D231, 1986.
- [Vua1-90] P.J.Vuarchex,'Huiles et liquides isolants', Techniques de l'ingénieur, D-230, Paris 1990.
- [Wol-46] N.F.Wolochtchenko,'Influence des écrans sur la rigidité électrique des différents intervalles d'amorçage d'air',(Russe) Elekritchstwo,nr3, pp: 21, 1946.

Chapitre III

Surveillance et diagnostic de l'huile de transformateur

et Applications des RNA

- 3.1 Introduction
- 3.2 Diagnostic
 - 3.2.1 Méthodes internes de diagnostic
 - 3.2.2 Méthodes externes de diagnostic
 - 3.2.3 Réseaux de neurones artificiels et diagnostic
 - 3.2.4 Diagnostic d'une isolation
- 3.3 Mesures et surveillance comme outils de diagnostic de l'isolation d'un transformateur
 - 3.3.1 Défaillances et problèmes de transformateurs de puissance
 - 3.3.2 Gestion de la vie d'un transformateur
- 3.4 Surveillance et méthodes de diagnostic d'une isolation de transformateur
 - 3.4.1 Méthodes de diagnostic traditionnelles
 - 3.4.1.1 Diagnostic avec analyse de gaz dissous
 - 3.4.1.2 Qualité d'une huile de transformateur (ou diagnostic physico-chimique)
 - a) Critères de fin de vie basés sur les tests physico-chimiques
 - b) Efficacité du diagnostic physico-chimique
 - 3.4.2 Méthodes non traditionnelles pour le diagnostic des transformateurs
 - 3.4.2.1 Essai en service des décharges partielles (DP)
 - 3.4.2.2 Mesure de tension de rétablissement
 - 3.4.2.3 Logiciels et systèmes intelligents et experts pour le diagnostic
- 3.5 Surveillance-maintenance de l'huile de transformateur
 - 3.5.1 Surveillance du transformateur en service
 - 3.5.1.1 Echauffement
 - 3.5.1.2 Protection du diélectrique contre l'humidité
 - 3.5.2 Surveillance de l'huile en service
 - 3.5.3 Remplacement et régénération des huiles de transformateurs

- 3.6 Application des RNA au diagnostic de la BORAK22
 - 3.6.1 La Back-propagation (BP)
 - 3.6.1.1 Back-propagation avec momentum
 - a) Réseau à une couche cachée
 - b) Réseau à deux couches cachées
 - 3.6.1.2 Back-propagation avec momentum et adaptation du taux d'apprentissage (MADPTA)
 - 3.6.1.3 Méthode de Levenberg-Marquardt (L.M)
 - 3.6.2 Régularisation bayésienne (RB)
 - 3.6.3 Paramètres influant l'apprentissage des réseaux du diagnostic
 - 3.6.3.1 Méthode du momentum
 - 3.6.3.2 Méthode MADPTA
 - 3.6.3.3 Méthode L.M
 - 3.6.3.4 Méthode de régularisation Bayésienne (RB)
 - 3.6.4 Influence de la base de données sur l'apprentissage
 - 3.6.4.1 Phase d'apprentissage
 - 3.6.4.2 Phase de généralisation
- 3.7 Architectures et apprentissages des RNA utilisés dans la classification de la BORAK22
 - 3.7.1 Les Réseaux Compétitifs
 - 3.7.2 Les cartes auto-organisées de Kohonen (Self-Organization Maps -SOM)
 - 3.7.3 Les réseaux bayésiens
- 3.8 Discussions
- 3.9 Conclusion
- 3.10 Bibliographie

« Personne ne niera que le monde des perceptions ne détermine en fait rigoureusement le système théorique, bien qu'aucun chemin logique ne conduise des perceptions aux principes de la théorie »

Albert Einstein

COMMENT JE VOIS LE MONDE

« Les capacités d'un niveau le plus élevé de regagner ce niveau et d'affecter les niveaux les plus bas sont un genre de tour magique que nous sentons très proche du noyau de la conscience »

**Professeur Douglas Hofstadter
et Professeur Daniel Dennett**

THE MIND'S I

3.1 INTRODUCTION

Il y a un besoin croissant de meilleurs outils de diagnostic et de surveillance non-destructifs pour évaluer l'état interne des transformateurs. S'il y a un problème, le transformateur peut être alors réparé ou remplacé avant qu'il ne tombe en panne. Ce chapitre synthétise les méthodes existantes les plus utilisées de surveillance et de diagnostic et présente les différentes applications des réseaux de neurones dans le diagnostic et la classification de l'huile de transformateur utilisée par SONELGAZ et nommée BORAK22.

La méthode de diagnostic adoptée dans ce travail est celle des tests physico-chimiques, utilisée par SONELGAZ dans son Laboratoire de Contrôle, Mesures et Essais d'El-Hamma (Alger). Les tests physico-chimiques utilisés sont au nombre de six : couleur, viscosité cinématique, rigidité diélectrique, facteur de pertes diélectriques, acidité et teneur en eau.

L'interprétation des résultats d'essais est une opération lourde et la présence d'analystes experts est nécessaire dans l'étape du diagnostic. Pour construire un système de diagnostic robuste et fiable, cette expérience humaine doit être conservée dans un système intelligent pour l'aide à la décision. Ceci offrirait une meilleure maintenance aux transformateurs. Cela nous a poussé à réfléchir à la conception d'un système intelligent pour le diagnostic et la classification de l'huile de transformateur. Plusieurs algorithmes de réseaux de neurones sont utilisés dans ce but.

3.2 DIAGNOSTIC

Le diagnostic industriel a pour objet de trouver la cause d'une défaillance. Le dictionnaire Robert donne une définition étymologique : 'action de déterminer une maladie d'après ses symptômes' et fait référence à sa racine Grecque 'Dia : par, Gnosis : connaissance'. Les industriels et les étudiants en formation technique adoptent quant à eux la définition retenue par AFNOR et CEI [Zwi-95]:

« le diagnostic est l'identification de la cause probable de la (des) défaillance(s) à l'aide d'un raisonnement logique fondé sur un ensemble d'informations provenant d'une inspection, d'un contrôle ou d'un test ».

Cette définition très courte et concise résume les deux tâches essentielles en diagnostic :

- Observer les symptômes de la défaillance.
- Identifier la cause de la défaillance à l'aide d'un raisonnement logique fondé sur des observations.

Si la prise de décision conduit à déclarer le processus défaillant, il convient alors de sélectionner une méthode de diagnostic. Les deux grandes familles de diagnostic sont les méthodes internes et externes et les méthodes inductives et déductives [Zwi-95]. Nous décrivons dans ce qui suit les différentes méthodes de surveillance, diagnostic et maintenance de l'huile isolante d'un transformateur.

3.2.1 Méthodes internes de diagnostic

Cette famille de méthodes est principalement dérivée des techniques utilisées par les automaticiens. A partir de modèles physiques ou de comportements validés par les techniques d'identification de paramètres, il devient possible de mettre en œuvre la méthode du problème inverse. Ces méthodes impliquent une connaissance approfondie du fonctionnement sous forme de modèles mathématiques qui devront être obligatoirement validés expérimentalement avant toute utilisation industrielle. Les

méthodes de diagnostic internes se regroupent en trois familles : la méthode du modèle, la méthode d'identification de paramètres et la méthode d'estimation du vecteur d'état.

3.2.2 Méthodes externes de diagnostic

Au sens strict, ces méthodes supposent qu'aucun modèle n'est disponible pour décrire les relations de cause à effet. La seule connaissance repose sur l'expertise humaine confortée par un solide retour d'expérience. Dans cette catégorie, on retrouve toutes les méthodes basées sur l'intelligence artificielle, entre autres celles que nous allons utiliser dans ce travail : les réseaux de neurones.

Les prises de décisions s'exercent pratiquement à chaque étape d'une procédure de diagnostic industriel. Tout d'abord, à partir de la relation de causalité, il faut déterminer les paramètres les plus significatifs du système à diagnostiquer. Ensuite les tests ou mesures,... fournissant les signatures les plus sensibles doivent être retenues. Ces prises de décisions, délicates à prendre, sont du ressort des spécialistes qui possèdent une maîtrise des phénomènes physiques. L'étape suivante consiste à détecter une différence éventuelle entre la signature mesurée à l'instant présent et la signature normale.

3.2.3 Réseaux de neurones artificiels et diagnostic

En diagnostic et maintenance, les spécialistes de maintenance sont conduits à reconnaître et à classer les signatures (paramètres du diagnostic, qui sont dans ce travail les six tests physico-chimiques) le plus souvent de façon visuelle. De façon similaire, lors du fonctionnement normal d'un processus complexe tel qu'une centrale nucléaire, les centaines d'informations visuelles délivrées par les indicateurs réparties sur les différents panneaux de la salle de commande constituent des formes visuelles que les opérateurs utilisent pour identifier les modes opératoires.

Les réseaux de neurones artificiels sont des outils particulièrement adaptés pour aider les spécialistes de maintenance dans les activités de reconnaissance et de classification. Pour cette raison, de nombreuses études pilotes sont entreprises dans le secteur de la maintenance pour évaluer les apports des réseaux de neurones avant leur mise en œuvre de façon opérationnelle.

Les réseaux de neurones présentent des perspectives d'application particulièrement intéressantes pour l'aide au diagnostic avec signatures externes. En effet, ces réseaux possèdent des propriétés similaires à celles de la reconnaissance des formes pour la classification de signatures.

La maintenance préventive et la maintenance prévisionnelle incluent des tâches telles que la surveillance, le diagnostic, les contrôles. Pour la maintenance préventive du matériel électrique, la détection et le diagnostic de défauts sont réalisés à l'aide de l'interprétation des signatures complexes (dans notre cas : couleur de l'huile isolante, acidité, pertes diélectriques, ...) [Zwi-95].

3.2.4 Diagnostic d'une isolation

Des compagnies d'électricité sont confrontées à un défi de plus en plus difficile pour assurer un approvisionnement régulier en électricité à des clients exigeants tout en devant réduire les coûts en fonctionnement et l'entretien pour rester commercialement concurrentielles. La cause la plus importante des arrêts coûteux dans les installations à haute tension est le vieillissement des matériaux d'isolation. L'un des problèmes est le manque d'outils appropriés pour diagnostiquer les matériaux d'isolation d'une façon sûre et non-destructive.

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

Les méthodes de diagnostic disponibles sont valables la plupart du temps seulement dans les conditions de laboratoire et sont difficiles à appliquer sur le terrain. Les résultats de mesure devraient être précis et assez détaillés pour former une base de données pour des décisions à prendre au sujet de l'entretien et du remplacement [Ida-03].

Les transformateurs sont demandés dans tous les systèmes d'alimentation interconnectés. La puissance de ces transformateurs s'étend de quelques kVA à quelques centaines de MVA, avec des coûts de remplacement allant de quelques centaines de dollars à des millions de dollars. Dans la pratique, la durée de vie d'un transformateur peut être plus longue que 60 ans avec un entretien approprié. Cependant, la défaillance en service d'un transformateur est potentiellement dangereuse pour le personnel (explosions, feu) et potentiellement préjudiciable à l'environnement par la fuite d'huile.

Des diagnostics électriques, optiques, physico-chimiques, acoustiques,... sont utilisés pour évaluer l'état de dégradation des matériaux qui conditionne la fiabilité des transformateurs. La maintenance prédictive de ces derniers fait l'objet d'investigations. La dégradation de l'huile isolante des transformateurs de puissance peut diminuer la vie utile de ces équipements et provoquer des arrêts de production coûteux. Pour prévenir de telles situations et pour aider les industries et les réseaux indépendants à assurer à coût concurrentiel la maintenance préventive de leurs équipements en charge, des systèmes de surveillance-diagnostic et maintenance doivent être installés [Kum-00, Mye-98].

Les stratégies basées sur les statistiques d'entretien visent en général, à réduire le nombre de défaillances, le coût de la défaillance et le coût de la réparation associée. L'exécution de la stratégie d'entretien fournit des possibilités pour acquérir de premières indications des changements des équipements [Aln-96]. Ces indications ont pu être de grande importance également en détectant des déviations dans la qualité des tests de l'isolation en question, tôt et avant qu'elles ne constituent des dangers ou causent des pertes.

3.3 MESURES ET SURVEILLANCE COMME OUTILS DE DIAGNOSTIC DE L'ISOLATION D'UN TRANSFORMATEUR

Le diagnostic des isolants est un outil très important et puissant pour augmenter la fiabilité des composants des systèmes électriques et leur sécurité et aussi diminuer le coût de l'approvisionnement en énergie. [Goc-99, Mye-98].

Les tests utilisés dans le diagnostic ont un but bien défini et ils doivent suivre les exigences suivant des normes nationales et/ou internationales. Après avoir réalisé les tests, une évaluation est nécessaire à faire pour classer l'isolant : a-t-il supporté le test ou non ? Le diagnostic d'un isolant a besoin de plusieurs mesures de paramètres pour prendre en considération leurs inférences [Kac-97].

Au début du diagnostic, la décision doit être prise par un expert et les décisions prises seront gardées dans les archives pour servir après à entraîner des systèmes intelligents tels que les réseaux de neurones utilisés dans le présent travail.

Les performances de l'isolation suivent la courbe représentée à la figure 3.1 [Goc-99]. Si le niveau du diagnostic a été atteint, une décision au sujet du temps de vie résiduelle peut être prise. Sans des actions concernant la réduction des performances de l'isolant ou la mise en place d'un système de diagnostic, le point 1 sur la figure 3.1 sera atteint et les performances de l'isolation seront en dessous du niveau de service, ce qui signifie

que l'appareil doit être mis hors service. Dans le cas contraire, le point 2 qui est un prolongement raisonnable de la durée de vie, sera atteint.

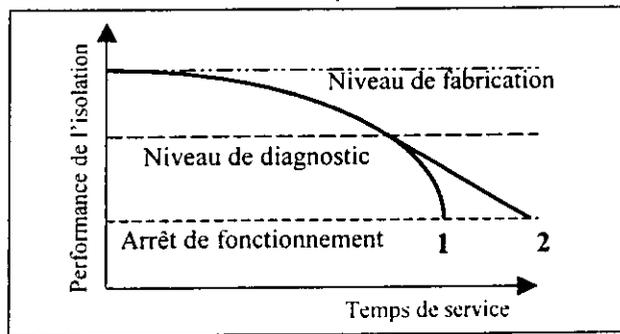


Fig. 3.1 Performance d'un isolant en fonction du temps de service

La forme de la courbe entre le point où la performance de l'isolation coupe le niveau du diagnostic et le point 2 (où la courbe coupe le niveau d'arrêt de fonctionnement) peut être déterminée avec la connaissance du système de diagnostic, ce qui offrira un autre prolongement du temps de service de l'isolant [Goc-99].

3.3.1 Défaillances et problèmes de transformateurs de puissance

La défaillance d'un transformateur peut se produire en raison de différentes causes et conditions. Ces dernières peuvent être généralement définies comme suit [Cig1-83, Kog-88] :

- Toute panne forcée due aux dommages du transformateur en service (dommages d'enroulement,...).
- Problème qui exige le déplacement du transformateur pour réparation, ou qui exige un champ étendu de réparation (réduction excessive de gaz, niveaux d'humidité élevés...).

Un aperçu donné par O.N. Grechko [Gre-96] signale que les causes principales (51% de défaillances de transformateur dans une période de cinq ans) étaient dues aux problèmes suivants :

- Humidité, contamination et vieillissement, qui ont fait diminuer la rigidité diélectrique interne du transformateur.
- Dommages de l'enroulement ou décompression de l'enroulement sous les forces de court-circuit.

Les statistiques de défaillance pour les grands transformateurs qui avaient été en service entre 15 et 25 ans indiquent que le vieillissement de l'isolation, les dommages des enroulements et la contamination, sont les sources principales des défaillances de transformateurs. Le coût et le temps de réparer et remplacer un transformateur de puissance sont très substantiels.

3.3.2 Gestion de la vie d'un transformateur

Pour des raisons économiques, la gestion de la durée de vie d'un transformateur a gagné de plus en plus d'importance dans les 10-15 dernières années [Wam-02, Wan3-00, Din-95, Zha-96]. L'objectif fondamental est de favoriser la plus longue durée de vie possible et de réduire au minimum les frais d'exploitation. L'importance de ce sujet [Cig2-93, Dar3-98, Jar-98, Kaz-981, Lap-95, Pet1-90, Pet2-98] a mené à beaucoup de recherches dans ce secteur. En général, la vie d'un transformateur est égale à la vie de l'isolation [Wam-02, Wan3-00], qui dépend de la rigidité mécanique et électrique des isolants

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

utilisés. La dégradation de l'isolation se compose des dégradations hydraulique, oxydante, et thermique.

La température d'un transformateur a un impact important sur la vie de l'isolation. Dans les guides de l'industrie électrique (CEI, ANSI, IEEE), le facteur principal de la fin de vie est lié seulement au facteur thermique du transformateur. Une méthode classique pour le calcul de la vie restante d'un transformateur consiste à utiliser la formule d'Arrhenius-Dakin [Wam-02]:

$$\text{Durée de vie restante} = Ae^{B/T}$$

où A est la vie initiale, B est une constante dépendant des propriétés du matériau à étudier et T est la température absolue en Kelvin.

Une approche plus complète est nécessaire pour évaluer la vie restante d'un transformateur. Les autres facteurs affectant la probabilité de défaillance ne sont pas aussi faciles à mesurer que le vieillissement thermique. Pour évaluer l'état d'un transformateur, plusieurs techniques de surveillance sont employées. Les méthodes les plus communes de surveillance (monitoring) employées pour l'évaluation de l'état d'un transformateur sont essentiellement données dans les références [Bel1-99, Bel2-99, Bre-92, Dar3-98, Har-97, Mal-00, Sah1-97, Sah2-97, Ten-00, Wan2-00]. L'évaluation de la vie des grands transformateurs est faite pour [Lap-95] :

- Surveiller l'état des transformateurs et fournir une détection précoce des défauts.
- Diagnostiquer les problèmes quand les transformateurs montrent des signes de détresse ou pour suivre l'opération de protection de l'équipement.
- Définir si un transformateur est dans des conditions convenables pour affronter des conditions inhabituelles.
- Obtenir des résultats de référence pour aider dans l'interprétation des essais.
- Aider à projeter la stratégie de remplacement pour une population de transformateurs.

3.4 SURVEILLANCE ET METHODES DE DIAGNOSTIC D'UNE ISOLATION DE TRANSFORMATEUR

D'une manière générale, le terme 'surveillance' (monitoring) décrit une mesure de paramètres de base avec des déclenchements d'alarmes de seuil. Le terme 'diagnostic' indique l'addition d'une analyse sophistiquée, telle qu'un système intelligent capable de fournir une évaluation de l'état de l'équipement et des actions suggérées. Il y a une variété d'outils disponibles pour évaluer l'état des transformateurs [Mal-00, Tan1] et qui peuvent être classées en deux catégories : traditionnelles et non-traditionnelles.

3.4.1 Méthodes de diagnostic traditionnelles

Les essais d'huile isolante sont les plus employés pour évaluer l'état des transformateurs en service. Les défauts thermiques et électriques mènent à la dégradation de l'huile.

3.4.1.1 Diagnostic avec analyse de gaz dissous

Les huiles isolantes sous des contraintes électriques ou thermiques claquent et se décomposent pour libérer de petites quantités de gaz. La composition de ces gaz dépend du type de défaut. Par le moyen d'analyse de gaz dissous, il est possible de distinguer des défauts tels que les décharges partielles (DP), surchauffement et formation d'arcs dans une grande variété d'équipements remplis d'huile. Un certain nombre d'échantillons

doit être successivement considéré sur une assez longue période pour discerner les tendances et déterminer la sévérité et la progression des défauts naissants. Dans ces essais, on évalue généralement la concentration de l'hydrogène, du méthane, de l'acétylène, de l'oxyde de carbone (CO, CO₂), de l'azote, et de l'oxygène. Les rapports relatifs et les quantités de gaz détectés dans l'échantillon sont employés pour trouver les problèmes liés à la structure de l'isolant [IEC-92, IEEE-91, Fal-75].

- *Décomposition cellulosique* : la décomposition thermique de l'isolation de cellulose imprégnée produit des oxydes de carbone (CO, CO₂), un peu d'hydrogène (H₂) ainsi que du méthane (CH₄).
- *Décomposition de l'huile* : les huiles minérales de transformateur sont des mélanges de molécules différentes d'hydrocarbures, et les processus de décomposition de ces hydrocarbures lors des défauts thermiques ou électriques sont complexes. Le chauffage de l'huile produit de l'éthylène (C₂H₄) comme gaz principal.

L'analyse des gaz dissous ne fournit pas toujours des informations suffisantes pour évaluer l'état de l'huile du transformateur. Généralement, il y a trois étapes nécessaires pour une bonne évaluation.

- La première étape est de vérifier si un défaut existe. Les transformateurs en service ont toujours des gaz dissous de défaut dans leur huile. Seulement quand leurs niveaux excèdent une certaine valeur-seuil, un défaut est suspecté. Plusieurs valeurs recommandées ont été publiées [Wam-02].
- La deuxième étape est de déterminer le type de défaut. Les méthodes les plus utilisées sont celles des gaz clés (principaux) et celles des rapports de gaz [Bel2-99, Goc-99, Har-97, Mal-00, Wan2-00]. La première exige de tracer la totalité du gaz dissous combustible en pourcentage dans un histogramme. Chaque type de défaut donnera un modèle distinctif caractérisé par un gaz principal, généralement le plus abondant [Wam-02, Wan3-00]. Par exemple, un niveau élevé d'hydrogène avec des niveaux bas d'autres gaz est une caractéristique de la décharge partielle. La méthode du rapport exige le calcul des rapports de gaz les uns par rapport aux autres, tel que le méthane à l'hydrogène.
- Dans la troisième étape, trois ou quatre rapports pareils sont employés pour le diagnostic. Les rapports les plus répandus sont les rapports de Roger [Wam-02]. La sévérité d'un défaut est établie par la comparaison des niveaux des gaz générés dans l'huile et de leur taux de génération à des niveaux seuil. Au moins deux échantillons consécutifs sont nécessaires pour calculer des taux de génération de défaut. Une liste de gaz principaux (clés) et leurs défauts relatifs sont donnés dans le tableau 3.1. Pour une information plus détaillée, il faut consulter IEEE Standard C57.104-1991 [IEEE-91].

Gaz principaux	Défaut caractéristique
H ₂	Décharge partielle
C ₂ H ₆	Défaut thermique < 300°C
C ₂ H ₄	Défaut thermique 300°C < 700°C
C ₂ H ₂ , C ₂ H ₄	Défaut thermique > 700°C
C ₂ H ₂ , H ₂	Libération d'énergie

Tab. 3.1 Gaz principaux produits par un défaut particulier

L'essence du diagnostic avec analyse des gaz dissous est la détermination de la cause de la formation des composés de produits chimiques dans le milieu liquide d'isolation. Aujourd'hui, ces méthodes de diagnostic permettent non seulement d'énoncer l'état de l'isolation électrique à haute tension remplie d'huile, mais aussi de faire les premiers pas dans la prédiction de la vie résiduelle de l'isolant [Ara-02].

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

Le diagnostic avec analyse des gaz dissous a des avantages et des inconvénients. Les inconvénients sont la perte de substances à diagnostiquer due à de nouvelles réactions chimiques, la perte par diffusion vers l'environnement externe et la perte par absorption des matériaux de construction de l'équipement. Un avantage est que les matériaux du diagnostic, dans la plupart des cas, sont liés aux défauts électriques ou thermiques et corrélient avec la vie résiduelle du matériau. Un autre avantage est que les mesures peuvent fréquemment être faites sans interrompre le fonctionnement de l'appareil.

En soi, La génération des gaz, n'est pas suffisante pour déterminer un caractère d'un défaut parce que les mêmes gaz sont formés soit sous un effet thermique ou électrique ou de cavitation [Ara-02]. Cependant, le taux de concentration de ces gaz aide à différencier entre les dommages thermiques et électriques. En outre, les taux de génération des gaz caractérisent la rapidité du processus de décomposition de l'isolation. La formation des composants non-saturés (l'éthylène, et particulièrement l'acétylène) est spécifique aux décharges électriques d'énergie élevée.

3.4.1.2 Qualité d'une huile de transformateur (ou diagnostic physico-chimique)

L'état de l'huile affecte considérablement les performances et la durée de vie des transformateurs. Une combinaison des essais électriques, physiques et chimiques est effectuée pour mesurer le changement des propriétés électriques et le degré de détérioration de l'huile isolante.

Les résultats sont employés pour établir des procédures d'entretien prédictives pour éviter des arrêts coûteux et la défaillance prématurée de l'équipement, et pour prolonger la durée de vie de l'équipement. Les essais les plus utilisés pour les huiles isolantes et leurs significations sont donnés dans le tableau 3.2. Des niveaux de seuil sont indiqués pour ces tests dans ASTM D3487 pour les huiles neuves et le guide d'IEEE 367-1985 pour les huiles de service [Wam-02].

Type d'essai	ASTM	Signification/effets
Rupture diélectrique	D877, D1816	Humidité, particules, fibres de cellulose/rigidité diélectrique basse
Indice de neutralisation	D644, D974	Produits acides de l'oxydation de l'huile/dépôt, corrosion
Tension interfaciale (IFT)	D971	Présence des contaminants polaires, acides, dissolvants...
Couleur	D1500	Opacité indiquant la contamination ou la détérioration
Teneur en eau	D1533	Décomposition excessive du papier / Rigidité diélectrique basse
Inhibiteur d'oxydation	D2668, D1473	Vieillessement accéléré de l'huile

Tab. 3.2 Essais des huiles isolantes neuves selon les normes ASTM

Les essais et les normes utilisés par la SONELGAZ sont donnés dans le tableau 3.3

Type d'essai	Norme
Couleur	ASTM D 1500
Viscosité cinématique	NF-T-60 100
Acidité totale	CEI 296
Rigidité diélectrique	CEI 156
Pertes diélectriques	CEI 250
Teneur en eau	ISO R 760

Tab. 3.3 Normes utilisées par SONELGAZ pour l'analyse de la BORAK22

a) Critères de fin de vie basés sur les tests physico-chimiques

Une fin de vie est atteinte quand une ou plusieurs des caractéristiques du diagnostic changent jusqu'au point où le fonctionnement de l'équipement n'est plus admissible. La fin de vie est atteinte par exemple si les propriétés électriques de l'huile ($\tan\delta$, conductivité, etc...) excèdent les niveaux recommandés par le concepteur de l'équipement. Il est possible de calculer la valeur de vie restante de chacune des caractéristiques si l'essai d'usine est pris comme valeur initiale et la valeur finale comme valeur limite permise. Si ces limites sont indiquées comme r_0 et r_1 , la valeur courante 'r' est comprise entre r_0 et r_1 . Des valeurs utiles de n'importe quelle caractéristique peuvent être exprimées comme le pourcentage de la valeur utilisée [Ara-02] :

$$R_{util} = 100(r-r_0) / (r_1 - r_0)$$

Et le pourcentage de la valeur résiduelle comme :

$$R_{res} = 100(r_1 - r) / (r_1 - r_0)$$

Ou, si la variation de cette caractéristique avec le temps $\Delta r / \Delta \tau$ est connue, par :

$$R_{res} = (r_1 - r) / (\Delta r / \Delta \tau)$$

La valeur estimée de la vie restante aidera dans la prise de décision. Il est de toute manière obligatoire que l'équipement soit mis hors service si la valeur de la durée de vie de service devient égale ou plus grande que la durée de vie estimée.

À la différence des propriétés électriques, les caractéristiques physico-chimiques ne sont pas étroitement liées à la durée de vie. La fluctuation des caractéristiques de l'huile est faite sur une large gamme. Ceci ne devrait pas devenir une raison pour mettre l'équipement hors service. Si une partie de l'équipement est arrêtée pour d'autres raisons, le remplacement de l'huile ou sa régénération rendra la propriété physico-chimique de l'huile à sa valeur originale. Pour cette raison, il est nécessaire de comprendre lesquelles des caractéristiques physico-chimiques sont critiques, le rapport entre les changements de ces caractéristiques, et l'état de l'équipement. Les changements dans les données physico-chimiques utiles pour l'évaluation de l'équipement sont :

- Pression partielle des gaz et de l'eau, $\tan\delta$ du papier et de l'huile et conductivité du papier et de l'huile sur toute la gamme de température de fonctionnement.
- Vieillessement du papier et de l'huile.
- Température du point d'éclair.
- Viscosité.
- Tension superficielle de l'huile.
- Point de feu de l'huile.
- Acidité de l'huile.
- Teneur d'antioxydant dans l'huile.
- Contamination mécanique de l'huile.
- Couleur de l'huile.

b) Efficacité du diagnostic physico-chimique

Afin de réaliser efficacement le diagnostic physico-chimique, les conditions suivantes sont nécessaires :

- Prévention de la perte de composants dissous dans l'huile pendant le prélèvement, le transport et le stockage de l'échantillon d'huile, et la prévention de l'introduction de composants d'air.
- Utilisation de méthodes d'analyse extrêmement sensibles.

- Prise en compte des phénomènes d'absorption dans les systèmes complexes d'isolation de l'équipement.
- Emploi de logiciels basés sur la connaissance de l'expertise.

3.4.2 Méthodes non traditionnelles pour le diagnostic des transformateurs

Il y a eu, ces dernières années, plusieurs développements nouveaux dans les techniques d'essais et de surveillance qui trouvent une utilisation croissante dans le diagnostic des transformateurs.

3.4.2.1 Essai en service des décharges partielles (DP)

Les DP dans les huiles de transformateurs détériorent les propriétés des matériaux isolants et peuvent mener aux défaillances [Wam-02]. Deux méthodes de détection de DP sont utilisées. Détection des signaux acoustiques et mesure des signaux électriques produits par les DP [Cor-99]. Les DP peuvent être détectées indirectement en utilisant des techniques chimiques telles que la mesure des produits de dégradation produits par les DP. Les limites acceptables de la charge des DP pour de nouveaux transformateurs dépendent de la tension appliquée et de la taille des transformateurs variant de 100 à 500 pC.

Les impulsions des DP produisent des ondes de contraintes mécaniques qui se propagent à travers l'huile environnante (dans la gamme de 100 à 300 kHz) [Wam-02]. Pour détecter ces ondes, des détecteurs d'émission acoustique sont montés sur les parois du réservoir du transformateur ou dans l'huile à l'intérieur du même réservoir. Les méthodes acoustiques de détection des DP sont limitées par l'atténuation du signal et les mesures électriques sont limitées par des problèmes d'interférence électromagnétique. Des investigations sont également en cours pour améliorer la détection acoustique des DP, comme la détection électrique, pour la surveillance en service [Dor-01].

3.4.2.2 Mesure de la tension de rétablissement

La méthode de mesure de la tension de rétablissement (en anglais Recovery Voltage Measurement - RVM) est employée pour détecter l'état de l'isolation huile-papier et la teneur en eau de l'isolation. Cette méthode se fonde sur le principe de la polarisation inter-faciale des matériaux diélectriques stratifiés. C'est l'accumulation des charges d'espace générées par les impuretés et l'humidité aux interfaces de l'isolation huile-papier. Cette méthode est très controversée quant à sa convenance à la mesure directe de la teneur en eau dans l'huile. Les figures 3.2 et 3.3 montrent les courbes typiques de RVM pour les vieux transformateurs qui sont respectivement en bonnes et mauvaises conditions [Wam-02].

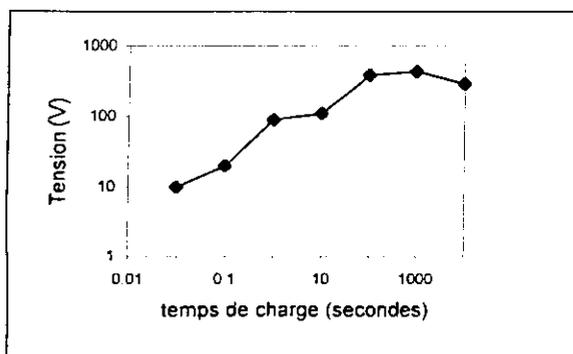


Fig.3.2 Courbe typique de la RVM pour un transformateur en bonnes conditions

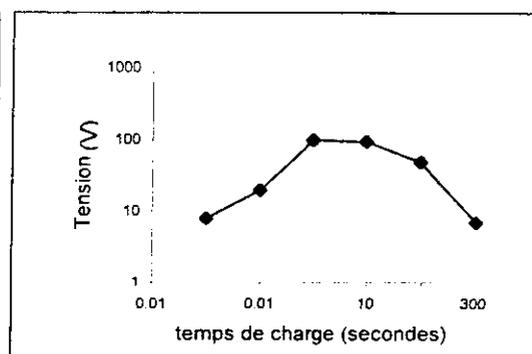


Fig.3.3 Courbe typique de la RVM pour un transformateur en mauvaises conditions

3.4.2.3 Logiciels et systèmes intelligents et experts pour le diagnostic

Les logiciels pour diagnostic, qui donnent des indications plus précises des problèmes des huiles de transformateur que l'analyse conventionnelle, font l'objet d'études de beaucoup de chercheurs et utilisateurs [Tom-93, Wan2-00]. L'utilisation d'un logiciel peut améliorer la fiabilité et la répétitivité de l'analyse des essais. Elle peut également servir à extraire l'information qui n'est pas fournie directement par les bases de données. Beaucoup de recherches ont été faites par logiciel pour interpréter des essais d'huile de transformateur tels que les gaz, la teneur en eau, la rigidité diélectrique, l'acidité et corrélérer les données avec l'état de l'isolation de transformateur.

L'avance actuelle des techniques d'intelligence artificielle (IA) a permis à des ingénieurs et à des chercheurs de développer des logiciels puissants et souples d'IA pour diagnostiquer des défauts de transformateur. L'utilisation d'un système expert offre le potentiel de réduire la main d'œuvre et les frais pour évaluer l'état du transformateur ; cependant, ce potentiel n'a pas été encore totalement réalisé.

3.5 SURVEILLANCE-MAINTENANCE DE L'HUILE DE TRANSFORMATEUR

3.5.1 Surveillance du transformateur en service

La sauvegarde du diélectrique, donc du transformateur, passe par une surveillance vigilante des conditions d'exploitation.

3.5.1.1 Echauffement

La connaissance des températures atteintes par le diélectrique (lue au thermomètre placé sur le couvercle) en fonction des variations de la charge et de la température ambiante doit faire l'objet de relevés systématiques qui, seuls, permettront de déceler d'éventuelles anomalies. Il arrive parfois que des appareils vieillissent anormalement en raison d'échauffements survenant soit à l'occasion de surcharges répétées, soit en raison de mauvaises conditions d'installation favorisant par exemple le recyclage de l'air chaud sur les éléments de refroidissement.

L'essai d'émission infrarouge est employé aussi pour vérifier la température de surface externe du transformateur en service. Il est utile pour détecter des problèmes thermiques dans un transformateur tels que des blocages de systèmes de refroidissement, localisant des problèmes de connexions électriques et les points chauds [Duk-98, Har-97]. La sévérité du surchauffement peut être classée par catégorie comme suit :

Classification	Excès de Température
Attention	0-9°C
Intermédiaire	10-20°C
Sérieux	21-49 °C
Critique	> 50°C

L'excès de température est la différence en température entre le point référence du transformateur sous température normale et un point ayant la température la plus élevée.

3.5.1.2 Protection du diélectrique contre l'humidité

Dans le cas d'appareils de type respirant, le travail imposé à l'assécheur d'air varie dans de très grandes limites en fonction des conditions atmosphériques ainsi que de l'importance et de la fréquence des variations de charge du transformateur.

Seule une surveillance active, effectuée à intervalles rapprochés, de préférence hebdomadaires, permet de déceler une perte d'efficacité des assécheurs et d'en effectuer la régénération à temps.

Il n'est pas rare, en effet, de constater des évolutions inattendues. Tel appareil exploité irrégulièrement à faible charge et même non chargé mais exposé alternativement au chaud et au froid, peut respirer plus que son homologue chargé en permanence. Une rupture de charge survenue par temps d'orage ou de brouillard peut avoir pour effet de saturer presque complètement la charge déshydratante.

Les assécheurs sont composés d'un récipient au moins en partie transparent, contenant une charge de granulés déshydratants, agissant par absorption, auxquels est associé un indicateur coloré bleu à l'état anhydre, virant au rose à saturation. L'ensemble pouvant être réactivé, en cas de saturation, par simple étuvage de quelques heures à une température comprise entre 100°C et 150°C [Hoc-88].

3.5.2 Surveillance de l'huile en service

Il a été constaté [Hoc-88] que le respect strict des consignes suscitées assure une bonne conservation du diélectrique durant un temps relativement long, mais il arrive que la meilleure organisation ait des failles et il est utile de savoir à l'occasion « faire le point ». C'est pourquoi il est indispensable de procéder périodiquement à des examens permettant de se rendre compte de l'évolution d'un diélectrique en service.

Il est difficile d'établir une règle absolue, valable dans tous les cas, et ainsi donne-t-on seulement des directives générales laissant à chacun le soin de les adapter à son cas particulier, en fonction des possibilités d'interruption de service, de la valeur du matériel installé, du prix de revient de ces examens, etc.

Notons également qu'il est indispensable de conserver tous les résultats de chacun de ces examens, car il est souvent plus utile de connaître de quelle façon évolue la situation, que de se référer à des limites établies dans un absolu qui laisse une assez grande part à l'arbitraire.

a) Périodicité des examens

Cette périodicité pourrait s'établir selon B. Hocard [Hoc-88] sur la base de:

- 2 à 3 ans pour des appareils respirants installés à l'extérieur ou dans des locaux humides.
- 3 à 5 ans pour des appareils respirants installés à l'intérieur de locaux secs.
- 10 ans pour des appareils de type étanche (avec ou sans matelas gazeux, avec conservateurs à membrane ou à parois déformables).

Il est certain que l'application de la gamme complète des tests conduirait souvent à des dépenses trop élevées et, en conséquence, c'est en général la valeur du matériel et le degré de fiabilité recherché qui guideront dans chaque cas particulier un choix dans l'éventail proposé.

b) Prélèvement d'échantillons

Quel que soit le test envisagé, la vérification d'un diélectrique passera toujours par le prélèvement d'un échantillon. Des précautions sont nécessaires pour que celui-ci soit représentatif du diélectrique contenu dans le transformateur.

Autant que possible, il est recommandé d'effectuer un prélèvement moins de trois heures après l'arrêt du transformateur, afin de recueillir un échantillon moyen résultant du brassage opéré par la circulation du diélectrique [Hoc-88]. D'autre part, cet

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

échantillon sera encore tiède, donc moins exposé à d'éventuelles condensations de l'humidité ambiante.

En effectuant la prise, il faut éviter de former des tourbillons afin de minimiser le contact du diélectrique avec l'air ambiant (formation de bulles).

La plus grande propreté est de rigueur ; on ne devra pas négliger de nettoyer les orifices et tuyauteries de prélèvement qui sont souvent souillés de poussière agglomérées en raison de suintement. Il est prudent de laisser écouler une certaine quantité de diélectrique, environ 10 fois le volume des tubulures de prélèvement, avant de recueillir l'échantillon, ceci afin de purger et rincer ces tubulures.

Il est conseillé de remplir le récipient de prélèvement à l'aide d'un tube plongeur et de laisser déborder environ une fois la contenance du récipient après remplissage, ce qui assure un ultime rinçage du récipient par le liquide à prélever. Enfin, il faut s'assurer que les matériaux constituant l'outillage ou les récipients de prélèvement sont sans interaction possible avec le diélectrique. De ce point de vue, le verre est le plus recommandable.

Les tests adoptés au laboratoire de contrôles mesures et essais de SONELGAZ- EL-Hamma sont au nombre de six et concernent la BORAK22. Ils sont donnés dans les paragraphes qui suivent.

3.5.2.1 Rigidité diélectrique

En règle générale, un abaissement de la tension de tenue électrique est dû à une pollution du diélectrique par l'humidité dissoute et /ou par des poussières en suspension. Les divers procédés de traitement dont l'application se fait en fonction du degré de pollution et du niveau de qualité à rétablir, visent à assurer simultanément un filtrage et une déshydratation du liquide diélectrique.

On doit noter que lors du premier remplissage d'un appareil neuf, il est souhaitable, par traitement préalable, d'amener le diélectrique à des valeurs de tension de tenue supérieures d'environ 10kV aux valeurs du tableau 3.4 données par la CEI 156.

Tensions nominales	$U_n \leq 36 \text{ kV}$	$36 < U_n \leq 70$	$70 < U_n \leq 170$	$170 < U_n$
Valeurs limites minimales de la tension de tenue électrique de l'huile [kV]	30	35	40	50

Tab. 3.4 Valeurs limites de la tenue diélectrique de l'huile de transformateur pour différentes tensions

3.5.2.2 Teneur en eau

L'humidité présente dans le diélectrique n'a pas nécessairement une action directe et immédiate sur sa tenue électrique, notamment si la concentration reste au-dessous de la saturation mais crée un risque. L'élimination de cette eau fait appel à des procédés de traitement assez complexes (traitement sous vide) surtout lorsque l'on désire abaisser la teneur en eau au dessous de 20 ppm (parties par million). Notons également que lors du premier remplissage d'un transformateur neuf, il est souhaitable de ramener, par traitement préalable, la teneur en eau à une valeur inférieure de 10 ppm environ à celles figurant au tableau 3.5 (ISO R 760).

Tensions nominales	$U_n \leq 36 \text{ kV}$	$36 < U_n \leq 70$	$70 < U_n \leq 170$	$170 < U_n$
Valeurs limites maximales de la teneur en eau de l'huile [ppm]	40	35	30	20

Tab. 3.5 Valeurs limites de la teneur en eau de l'huile de transformateur pour différentes tensions

3.5.2.3 Facteur de dissipation électrique (tgδ)

L'augmentation du facteur de dissipation électrique d'un diélectrique résulte de multiples facteurs. Souvent les traitements de filtrage et de déshydratation ont une action bénéfique sur le facteur de dissipation. Ceci peut en particulier être le cas du traitement préalable au premier remplissage d'un transformateur neuf.

Mais il arrive parfois que des diélectriques fortement pollués, suite au vieillissement, ou à un amorçage interne, ne puissent être ramenés dans les limites acceptables du facteur de dissipation (Tab.3.6) (CEI 250) [Hoc-88], sans recourir à un conditionnement par voie physico-chimique, s'apparentant à un raffinage généralement onéreux.

La valeur du facteur de dissipation variant avec la température, il est recommandé d'opérer à la température de référence 90°C. Mais dans le cas où l'on serait amené à opérer à 25°C, les valeurs de référence doivent être divisées par 15 [Hoc-88].

Tensions nominales	$U_n \leq 36 \text{ kV}$	$36 < U_n \leq 70$	$70 < U_n \leq 170$	$170 < U_n$
Valeurs limites de Tgδ à 90°C et 50 Hz	1.5	0.8	0.5	0.2

Tab. 3.6 Valeurs limites de tgδ de l'huile de transformateur pour différentes tensions

3.5.2.4 Indice de neutralisation

L'indice de neutralisation des huiles isolantes augmente progressivement en fonction du vieillissement. La valeur limite de sécurité est de 0.5mgKOH/g (CEI 296), au-delà de laquelle on peut craindre des modifications inattendues. Lorsque l'indice de neutralisation d'un diélectrique augmente au point de faire craindre pour sa conservation, il est prudent de procéder à un remplacement plutôt qu'à un traitement coûteux (parfois la régénération peut être assurée par le fournisseur du diélectrique).

3.5.2.5 Couleur

L'évolution de l'indice de couleur de 1 à 3 pour l'huile vieillie est due à l'oxydation de l'huile en service qui a pour conséquence la formation des produits acides (Tab.3.7) [Ham-97] (1 c'est blanc pur et 3 c'est citron pâle). Cet indice varie de 1 à 4. La valeur limite est quand la couleur est limpide sans contamination visible.

Etats de l'huile	Indice de couleur	Viscosité cinématique (CST)
Neuve	1	12.54
Vieillie à 80°C	1	12.54
Vieillie à 120°C	3	12.49

Tab.3.7 Indice de couleur et viscosité cinématique de l'huile avant et après vieillissement.

3.5.2.6 Viscosité cinématique

La viscosité cinématique est presque constante durant le vieillissement. La pollution de l'huile a pour effet de ralentir l'écoulement de l'huile (Tab.3.7) [Ham-97]. La valeur limite minimale est 10.5 (norme NF-T-60 100) (voir annexe 3.1).

3.5.3 Remplacement et régénération des huiles de transformateurs

- Le remplacement pur et simple du diélectrique peut s'imposer dans le cas où celui-ci est disqualifié ou même seulement pollué au point de rendre tout traitement plus onéreux que le remplacement.

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

Dans le cas d'un remplacement de diélectrique, on ne devra pas omettre de procéder à un rinçage soigné de la partie active du transformateur, pour éliminer les boues ou dépôts à l'aide de solvants appropriés, ou d'une certaine quantité de diélectrique neuf (qui sera perdue).

Les produits mis au rebut ne doivent en aucun cas être rejetés ni dans la nature ni à l'égout en raison des risques de pollution. Les huiles minérales de composition assimilable à celle d'un fuel peuvent être éliminées par combustion ou renvoyées à des entreprises de récupération

Les cas où le remplacement du diélectrique s'impose de façon absolue sont assez rares. Cependant, il faut signaler le cas d'une huile exagérément vieillie (indice de neutralisation $> 1\text{mg KOH/g}$ et tension interfaciale $< 10\text{mN/m}$).

- Une opération de régénération est nécessaire lorsqu'un diélectrique est sévèrement atteint pour être pollué par des produits de décomposition qui, s'ils sont insolubles, provoquent des dépôts dont l'élimination nécessite un décuvage et s'ils sont solubles, ils contaminent le diélectrique en jouant parfois un rôle de catalyseur dans la poursuite de la décomposition et ne peuvent être éliminés que par une opération chimique ou physico-chimique spécifique assimilable à un nouveau raffinage.

Cette régénération nécessite le plus souvent des moyens dont le coût de mise à disposition et de fonctionnement est hors proportion avec la valeur du diélectrique concerné; c'est pourquoi il est nécessaire, en règle générale, de traiter avec des entreprises spécialisées ou avec le fournisseur du diélectrique.

- L'opération de traitement est aussi nécessaire dans la majorité des cas, lorsque le diélectrique est pollué. Il s'agit essentiellement d'humidité, de poussières en suspension ou de gaz dissous dont l'élimination est possible en recourant à l'arsenal des moyens que nous passerons en revue [Hoc-88] :

- Les tamis

Réalisés de diverses façons : tissus textiles ou métalliques, grilles formées de lames métalliques minces ou de fils enroulés en spirales, ils retiennent les solides de dimensions supérieures à celles de leurs mailles, mais n'ont souvent qu'une efficacité réduite dans le domaine des poussières les plus fines, peut-être les plus nocives pour les transformateurs, en raison de la facilité avec laquelle elles restent en suspension.

- Les filtres –presses

Ces filtres sont constitués d'un empilement de feuilles de papier poreux ou de feutres et d'entretoises métalliques perforées, disposées de façon à réaliser des groupements en série (filtration multiple) ou en parallèle (grandes surfaces) dans lesquels le fluide à traiter traverse le matériau filtrant perpendiculairement à sa surface et dont les caractéristiques de filtration dépendent étroitement de la porosité du papier ou du feutre utilisé.

Le filtre–presse élimine assez bien les poussières et, dans une certaine mesure, assèche le diélectrique en retenant par capillarité l'eau libre et en absorbant l'eau dissoute (tant que la matière filtrante est elle-même sèche). Cependant, il se sature assez rapidement et ne peut fonctionner en régime continu puisqu'il faut périodiquement le démonter pour remplacer les éléments filtrants par des éléments nouveaux ou au moins nettoyés et déshydratés.

- Les cartouches filtrantes

Ces cartouches sont en général composées d'un empilement de rondelles (papier ou tissu), traversées par le fluide de la périphérie vers le centre, donc parallèlement au plan du matériau. Les caractéristiques de filtration dépendent tant du matériau employé que de la pression exercée sur l'empilement de rondelles. Le matériau constituant les cartouches peut aussi jouer un rôle vis-à-vis de l'eau, mais cette action est le plus souvent limitée en raison des pressions plus élevées mises en jeu pour la circulation du diélectrique.

- Les lits filtrants

En association avec les filtres déjà cités, on a parfois recours à des produits très variés (aluminium activé, argile, terre à Foulon, etc.) ayant sur l'eau dissoute et sur certains produits chimiques une action physico-chimique d'absorption plus au moins sélective, améliorant l'épuration. Leur mise en œuvre ne peut cependant se faire que dans les installations assez importantes.

- Les centrifugeuses

La séparation les uns des autres des solides et des liquides mélangés, à condition que leurs densités soient différentes, peut s'obtenir assez lentement par décantation. La centrifugeuse, en soumettant les mélanges, par rotation rapide de son «bol», à des accélérations considérables (plusieurs centaines de la valeur de la gravitation g), réalise cette décantation en des temps très courts. Grâce à des dispositions particulières (bols coniques multiples), elle peut fonctionner en régime continu au moins pour ce qui concerne la séparation des liquides (eau, huile en particulier).

Les impuretés solides restant en général piégées dans le bol, une préfiltration est nécessaire pour éviter l'engorgement; par contre, l'action sur les poussières les plus fines est le plus souvent remarquable. Certaines centrifugeuses équipées pour travailler à chaud et sous vide parviennent même à éliminer de façon satisfaisante une partie de l'humidité dissoute.

- Les colonnes de déshydratation –dégazage

En soumettant un liquide chaud pulvérisé ou dispersé en couches minces (colonnes à plateaux ou à anneaux de Raschig) [Hoc-88] à l'action d'un vide suffisant (limité cependant par l'ébullition du liquide), on parvient à éliminer l'eau ainsi que la plupart des gaz dissous. Ce procédé est très utilisé pour l'huile dont la tension de vapeur est nettement plus basse que celle de l'eau.

3.6 APPLICATION DES RNA AU DIAGNOSTIC DE LA BORAK22

Après avoir présenté les méthodes de surveillance, de diagnostic et de maintenance de l'huile de transformateur, nous présentons dans les sections qui viennent l'application des réseaux de neurones dans le diagnostic et la classification de la BORAK22.

Dans toutes les applications que nous présentons pour le diagnostic, les mêmes échantillons d'apprentissage et de test (généralisation) sont utilisés. La base de données est présentée comme suit : le vecteur entrée $P = [P_1, P_2, \dots, P_6]$ et le vecteur de sortie cible (désirée) présenté comme superviseur est donné par : $D = [d_1, \dots, d_4]$. Les vecteurs sont:

$P = [$ couleur, viscosité cinématique, acidité totale, rigidité diélectrique, facteur de pertes diélectriques, teneur en eau $]$.

$D = [$ changer, traiter, régénérer, garder $]$.

L'huile utilisée est pour des transformateurs de catégorie 'B' (transformateur de puissance de tension de réseau maximale supérieure à 72,5 kV et inférieure ou égale à 170 kV).

D'après les normes présentées dans ce chapitre et la base de données obtenue au laboratoire de Contrôles, Mesures et Essais de SONELGAZ – EL-Hamma, les cinq échantillons de la BORAK22 utilisés dans la phase de généralisation, sont analysés comme suit :

Echantillon 1 : tous les paramètres sont dégradés et sont hors normes. La régénération sera coûteuse et ne va pas améliorer beaucoup l'état de l'huile. Ceci a été décidé par l'analyste après avoir consulté l'historique de cette huile où l'échantillon (1) a déjà subi une régénération ; ce qui justifie la décision de remplacement de cette huile.

Echantillon 2 : l'huile est en bon état surtout que la rigidité diélectrique est très bonne, la décision prise est de la garder.

Echantillon 3 : la couleur est jaune pâle, la viscosité cinématique de cet échantillon est basse et la valeur de la teneur en eau indique la présence d'humidité dans l'huile. La valeur de l'indice de neutralisation n'est pas alarmante, la comparant aux autres tests, mais elle indique une certaine pollution dans l'huile. La valeur de la rigidité diélectrique est dans les normes, mais pas très bonne. L'analyste a pris la décision de traiter cette huile.

Echantillon 4 : tous les tests sont bons et donc la décision prise est de garder cette huile.

Echantillon 5: cette huile est plus polluée que celle de l'échantillon 3 sans dépasser les normes. La même décision prise pour l'échantillon 3 est applicable à cette huile.

Plusieurs algorithmes sont appliqués pour le diagnostic et l'aide à la décision. Ces algorithmes sont présentés ci-dessous. A la figure 3.4, la couche d'entrée est composée de 6 neurones, la(les) couche(s) cachée(s) contient(nent) un nombre variable de neurones et la sortie a 4 neurones. La sortie produit des nombres réels compris entre 0 et 1 indiquant la probabilité de chaque décision qui doit être prise après l'analyse des tests physico-chimiques.

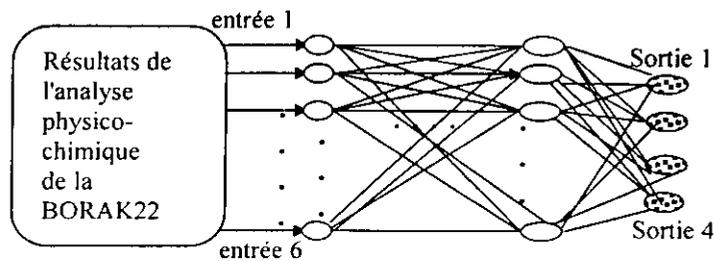


Fig.3.4 Schéma général du réseau de neurones multicouche utilisé

3.6.1 La Back-propagation (BP)

Plusieurs variantes de l'algorithme de la back-propagation sont utilisées. Pour chaque algorithme, plusieurs architectures de réseaux sont proposées. Nous avons remarqué l'influence de plusieurs paramètres :

- Le nombre de couches cachées.
- Le nombre de neurones dans la couche.
- L'ajout et l'inhibition des biais pour chaque couche.
- Les fonctions d'activation pour chaque couche.
- La variation des paramètres d'apprentissage.

Pour l'algorithme de la BP, trois méthodes sont utilisées : la BP avec momentum, la BP avec momentum et à adaptation de taux d'apprentissage (MADPTA) et la méthode de Levenberg-Marquardt (L.M).

3.6.1.1 Back-propagation avec momentum

Nous avons détaillé cette méthode au chapitre 1. Les paramètres les plus influents sur l'application de cette méthode sont le nombre de neurones et le nombre de couches cachées. Cette méthode est lente et moins précise que les autres méthodes que nous allons présenter. Le réseau ne converge qu'à partir d'une erreur fixée égale à 5.10^{-2} .

Nous considérons la sortie calculée en phase de test (généralisation) Y_G inactive (égale à 0) pour les valeurs données par le réseau comprises entre -0.2 et 0.45 et active (égale à 1) pour les valeurs de sortie comprises entre 0.46 et 1.1. Un vecteur de sortie $Y_G = [0 \ 0 \ 0 \ 1]$ ayant la sortie Y_{G4} active indique de garder l'huile. Les résultats présentés ci-dessous sont obtenus après la phase de généralisation.

a) Réseau à une couche cachée

Avant de commencer à présenter les décisions données par les réseaux utilisés, nous présentons la base de données d'apprentissage (Tab.3.8).

Échantillon N°	1	2	3	4	5
P_1	1	2	2	2	4
P_2 (cst)	10	11.45	12.3	12	12
P_3 (mgKOH/g)	0.046	0.11	0.072	0.092	0.3
P_4 (kV)	56	50	58	48	30
P_5	0.015	0.025	0.02	0.022	0.6
P_6 (ppm)	30	35	33	37	41
Y_{L1}	0.0002	-0.0004	-0.0002	0.0009	0.9002
Y_{L2}	0.0004	0.9006	0.0006	0.0000	0.0005
Y_{L3}	0.0001	0.0008	0.0001	0.9991	0.0005
Y_{L4}	1.1005	0.0003	0.8006	0.0004	0.0001

Tab.3.8 Diagnostic dans la phase d'apprentissage utilisant L.M

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

L'utilisation d'une couche cachée avec 20 neurones dans un réseau (Tab.3.9) et 10 neurones dans un autre (Tab.3.10) a donné les résultats ci-dessous. La sortie donnée par le réseau pour l'échantillon 1 dans le tableau 3.9 par exemple, est [1.02 -0.1, 0.36 -0.36] qui correspond à [1 0 0 0] d'après les intervalles des sorties que nous avons définis ci-dessus.

La base de données présentée ici est pour la phase de généralisation.

Echantillon	1	2	3	4	5	1	2	3	4	5
P_1	4	1	3	2	3	4	1	3	2	3
P_2 (cst)	11	12	10	12.5	11	11	12	10	12.5	11
P_3 (mgKOH/g)	0.4	0.04	0.08	0.02	0.1	0.4	0.04	0.08	0.02	0.1
P_4 (kV)	30	60	47	56	42	30	60	47	56	42
P_5	0.5	0.04	0.02	0.02	0.1	0.5	0.04	0.02	0.02	0.1
P_6 (ppm)	40	33	35	32	37	40	33	35	32	37
Y_{G1}	1.020	-0.110	-0.192	0.031	-0.100	1.020	-0.102	-0.002	0.131	-0.108
Y_{G2}	-0.100	0.100	0.463	-0.421	0.863	-0.001	0.100	0.650	-0.420	0.683
Y_{G3}	0.360	0.040	0.270	0.200	0.340	0.107	0.069	0.400	0.200	0.340
Y_{G4}	-0.360	0.857	0.450	0.870	-0.043	-0.005	1.086	0.227	1.025	-0.200

Tab3.9 Diagnostic de la BORAK22 utilisant la méthode du momentum à 1 couche cachée ayant 20 neurones

Tab.3.10 Diagnostic de la BORAK22 utilisant la méthode du momentum à 1 couche cachée ayant 10 neurones

Les deux réseaux ont donné des décisions correctes mais le réseau avec 10 neurones dans la couche cachée présente les meilleurs résultats : l'erreur la plus petite et la convergence la plus rapide. Le réseau contenant 20 neurones a convergé après 1500 itérations (Epochs). De plus, la sortie Y_{G3} a donné une valeur critique. Dans le premier cas le réseau est saturé. Nous donnons à la figure 3.5 l'erreur calculée ($5.12 \cdot 10^{-1}$) et permise (10^{-2}) à la sortie du réseau utilisant la méthode du momentum. Pour les problèmes demandant une haute précision, cette méthode ne pourra pas répondre à cette exigence.

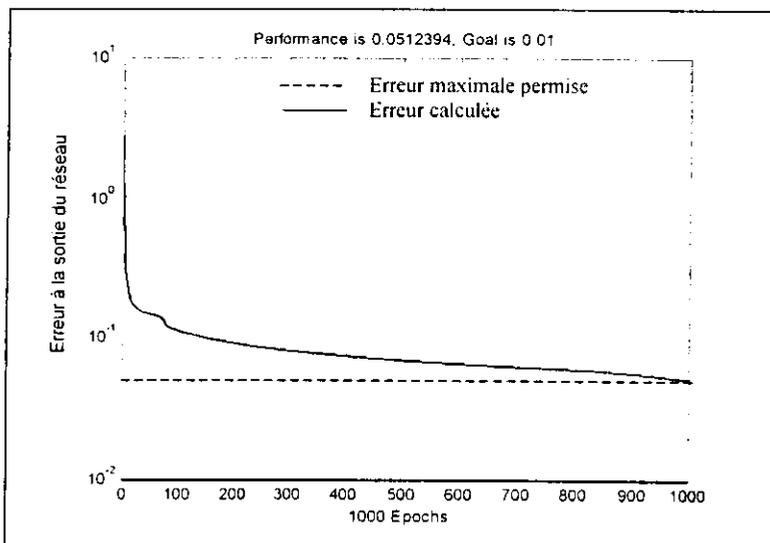


Fig.3.5 Erreur de l'apprentissage utilisant la méthode du momentum (1 couche, 10 neurones)

b) Réseau à deux couches cachées

Le réseau à deux couches cachées est présenté dans les tableaux (Tab.3.11 et Tab.3.12) en changeant le nombre de neurones dans chaque couche. Les résultats présentés dans

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

ces tableaux montrent que le nombre de neurones a plus d'influence que le nombre de couches cachées sur la qualité de l'apprentissage. La valeur du momentum est constante dans tous les cas d'apprentissage : $\mu = 0.69$.

L'influence des paramètres d'apprentissage sur la qualité de l'entraînement du réseau est présentée dans une étude [Mok15-03, Mok21-03] dans la section 3.6.3 pour tous les réseaux utilisés dans le diagnostic.

Echantillon	1	2	3	4	5	1	2	3	4	5
Y_{G1}	0.996	-0.005	-0.192	0.005	-0.060	0.472	-0.002	-0.007	0.004	-0.008
Y_{G2}	-0.150	0.019	0.693	-0.019	0.645	-0.001	0.240	0.459	-0.240	0.459
Y_{G3}	0.160	0.060	0.430	0.070	0.370	0.003	0.239	0.240	0.249	0.340
Y_{G4}	-0.120	1.022	0.162	1.014	-0.053	-0.015	0.765	0.025	0.765	0.083

Tab.3.11 Diagnostic de la BORAK22 utilisant la méthode du momentum à 2 couches cachées : 20 et 10 neurones dans les couches 1 et 2 resp.

Tab.3.12 Diagnostic de la BORAK22 utilisant la méthode du momentum à 2 couches cachées : 10 et 10 neurones dans les couches 1 et 2 resp.

Le réseau ayant 20 et 10 neurones dans les couches 1 et 2 respectivement présente les meilleurs résultats.

3.6.1.2 Back-propagation avec momentum et adaptation du taux d'apprentissage (MADPTA)

L'algorithme de cette méthode est donné au chapitre 1. L'utilisation de deux couches pour la MADPTA a mené le réseau à une stagnation dans l'évolution de l'erreur. Alors nous présentons les résultats donnés par l'utilisation d'une couche cachée seulement (Tab.3.13 et Tab.3.14).

Echantillon	1	2	3	4	5	1	2	3	4	5
Y_{G1}	1.042	0.084	0.030	0.082	0.219	0.814	0.0142	0.004	0.002	0.010
Y_{G2}	0.219	0.360	0.494	0.354	0.589	0.217	0.010	0.627	0.055	0.640
Y_{G3}	0.337	0.169	0.414	0.175	0.233	0.412	0.065	0.274	0.005	0.272
Y_{G4}	0.304	0.644	0.295	1.084	0.072	0.138	0.765	0.051	0.727	0.085

Tab.3.13 Diagnostic de la BORAK22 utilisant la méthode MADPTA à une couche : 30 neurones , $\mu = 0.4$

Tab.3.14 Diagnostic de la BORAK22 utilisant la méthode MADPTA à une couche :

Le réseau à 6 neurones présente la plus petite erreur et donc les meilleurs résultats. Nous donnons à la figure 3.6 l'erreur calculée et l'erreur maximale permise à la sortie du réseau et que nous avons fixée à 10^{-1} . Ce dernier a convergé après 16 itérations (Fig.3.6)

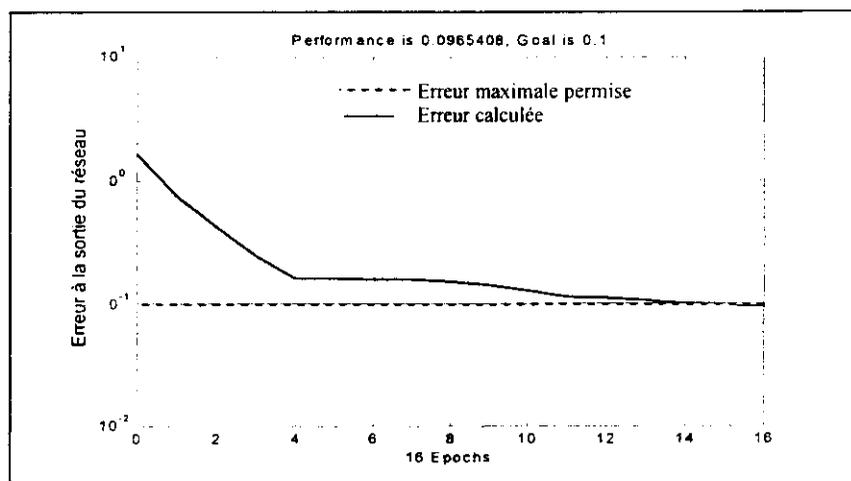


Fig.3.6 Erreur de l'apprentissage utilisant la méthode MADPTA (6 neurones)

3.6.1.3 Méthode de Levenberg-Marquardt (L.M)

En plus des bons résultats obtenus, cette méthode est caractérisée par la rapidité de convergence qu'elle présente par rapport aux deux autres méthodes citées dans les sections précédentes. La méthode de Levenberg-Marquardt est l'une des plus utilisées. Cette méthode consiste à considérer la surface d'erreur quadratique, en se basant sur la fonction d'énergie qui a la forme [Jer-93] :

$$E(k+1) = E(k) + \dot{E}(k) \cdot \Delta W(k) + \frac{1}{2}(\Delta W(k))^T \cdot \ddot{E}(k) \cdot \Delta W(k) \quad (III.1)$$

En posant pour chaque neurone i d'une couche l le vecteur de poids lui parvenant des neurones de la couche qui le précède par w_i^l , minimiser la fonction d'énergie revient à résoudre l'équation :

$$\frac{\partial E(k+1)}{\partial \Delta W(k)} = 0 \quad (III.2)$$

avec $\Delta W(k)$ est le vecteur global des taux d'adaptation de tous les vecteurs de poids w_i^l . Soit F_i^l les vecteurs des dérivées de l'erreur à la sortie du réseau par rapport à ces vecteurs poids w_i^l pour $l=1, \dots, L$, et $i=1, \dots, n_l$ où n_l est le nombre de neurones dans la dernière couche et L le nombre de couches cachées.

$$F_i^l = \sum_{p=1}^K \sum_{i=1}^{n_l} (d_i^p - y_i^{(L)}) \left[-\frac{\partial y_i^p}{\partial w_i^l} \right] \quad (III.3)$$

K représente le nombre d'exemples d'entraînement et n_l le nombre de neurones dans la dernière couche. On définit F le vecteur rassemblant les vecteurs F_i^l qui est le jacobien de l'erreur par rapport aux vecteurs poids. Afin de trouver l'optimum de (III.2), on obtient:

$$H \Delta W = -F \quad (III.4)$$

H représente la matrice Hessienne (ou le Hessien) (voir annexe 3.2).

L'équation (III.4) représente un système d'équations linéaires dont la résolution peut être faite par la méthode de Gauss-Newton.

La méthode de Levenberg-Marquardt remplace le calcul du Hessien par une approximation numérique donnée par :

$$H = FF^T + \lambda \Omega \quad (III.5)$$

λ est un coefficient positif, la matrice résultante de FF^T étant définie semi-positive, le coefficient λ est utilisé afin de mieux conditionner la matrice H . Ainsi, la matrice Ω doit être d'un choix approprié. Le choix d'une matrice diagonale dont les éléments sont égaux aux éléments diagonaux de la matrice FF^T donne de bons résultats. Le choix d'une matrice identité diagonale peut aussi être bien fait. Les valeurs que doit prendre le coefficient λ doivent être convenablement choisies.

En effet, la recherche doit être orientée dans une région de l'espace permettant de représenter adéquatement le problème non-linéaire par le modèle quadratique, afin de garantir la convergence de l'algorithme vers un minimum local.

Il existe quelques techniques, qui ont été mises au point pour cela. La plus simple consiste à commencer avec une petite valeur et la faire augmenter jusqu'à ce que l'erreur diminue, la valeur maximale de ce paramètre doit être cependant bornée.

L'algorithme de cette méthode est donné à la figure 3.7.

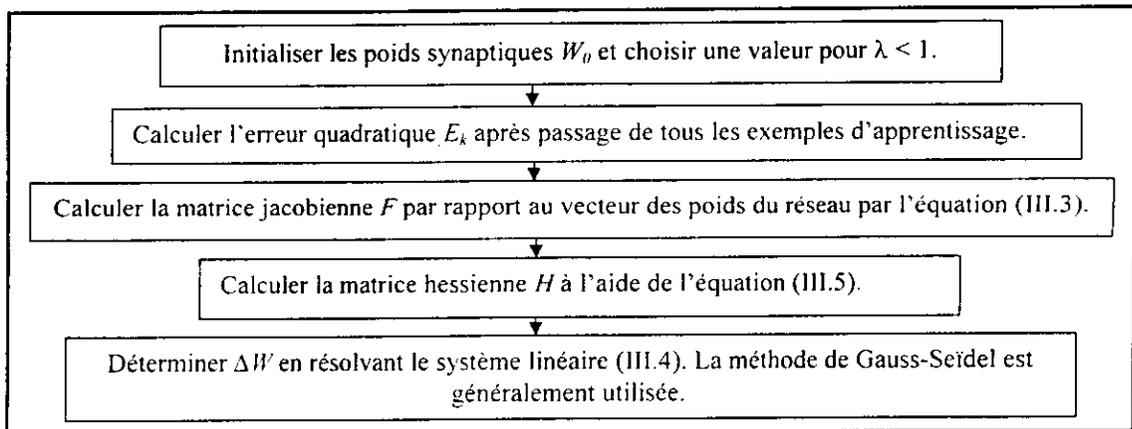


Fig.3.7 Algorithme de la méthode de Levenberg-Marquardt

Il existe d'autres variantes de cet algorithme, visant à le rendre plus efficace. Celles ci jouent, généralement, sur le choix optimal de λ , permettant de conduire vers un minimum global. Elles peuvent aussi agir sur le contrôle du gradient qui, en diminuant, ne permet plus de poursuivre la prospection dans d'autres directions, à la recherche d'un minimum.

Résultats du réseau Levenberg-Marquardt

Pour le réseau L.M, nous avons utilisé trois réseaux à une, deux et trois couches cachées et pour chacun de ces réseaux, nous avons changé le nombre de neurones. Dans les trois tableaux présentés ci-dessous (Tab.3.15, Tab.3.16 et Tab.3.17) les résultats sont parfaits pour les trois cas (1, 2 et trois couches).

D'après les résultats des tableaux 3.15, 3.16, 3.17 et d'après les figures 3.8, et 3.9, la méthode de Levenberg-Marquardt a une très bonne capacité de généralisation et une très bonne précision (10^{-15}) et sa convergence est rapide en comparant ses résultats avec ceux de la méthode du momentum [Mok5-02, Mok7-02, Mok13-02, Mok14-02, Mok15-03, Mok4-02]. Mais, par rapport à la méthode de la régularisation Bayésienne dont nous allons présenter les résultats, la convergence du réseau L.M est un peu lente.

Echantillon	1	2	3	4	5	1	2	3	4	5
Y_{G1}	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
Y_{G2}	0.000	0.000	1.000	0.006	1.000	0.000	0.000	1.000	0.000	1.000
Y_{G3}	0.000	0.000	0.000	0.002	0.001	0.000	0.000	0.000	0.000	0.000
Y_{G4}	0.000	1.000	0.000	1.000	0.001	0.000	1.000	0.000	1.000	0.000

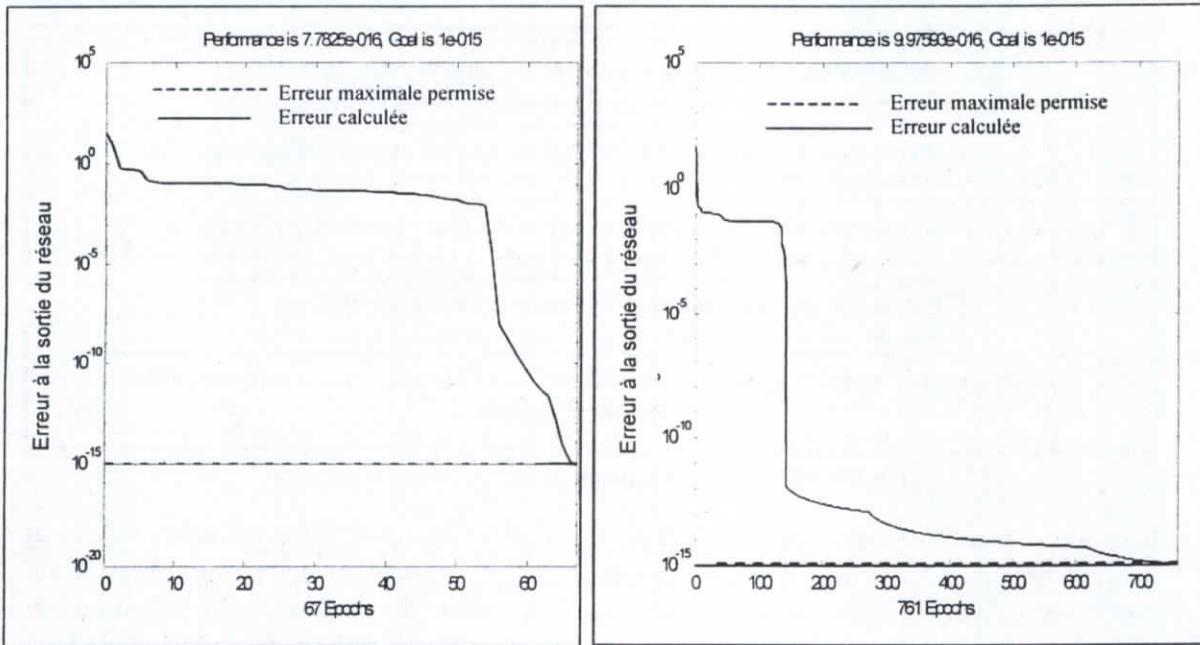
Tab.3.15 Diagnostic de la BORAK22 utilisant la méthode L.M à 3 couches cachées : 5, 4 et 3 neurones respectivement dans les couches 1, 2 et 3

Tab.3.16 Diagnostic de la BORAK22 utilisant la méthode L.M à 2 couches cachées : 5, 4 neurones respectivement dans les couches 1 et 2.

Echantillon	1	2	3	4	5
Y_{G1}	1.000	0.000	0.000	0.000	0.000
Y_{G2}	0.000	0.000	1.000	0.000	1.000
Y_{G3}	0.000	0.000	0.000	0.000	0.000
Y_{G4}	0.000	1.000	0.000	1.000	0.000

Tab.3.17 Diagnostic de la BORAK22 utilisant la méthode L.M à 1 couche cachée avec 6 neurones

L'influence du nombre de couches est surtout notable pour la rapidité de la convergence du réseau. L'utilisation de deux couches (Fig.3.8-(a)) est le choix optimal. En utilisant trois couches (Fig.3.8-(b)), le réseau se sature et devient lourd mais il donne toujours une très bonne performance (Fig.3.8). L'utilisation d'une couche (Fig.3.9) donne des résultats intermédiaires entre ceux de deux et trois couches.



(a) 2 couches

(b) 3 couches

Fig.3.8 Erreur de l'apprentissage en utilisant la méthode L.M à 2 et 3 couches

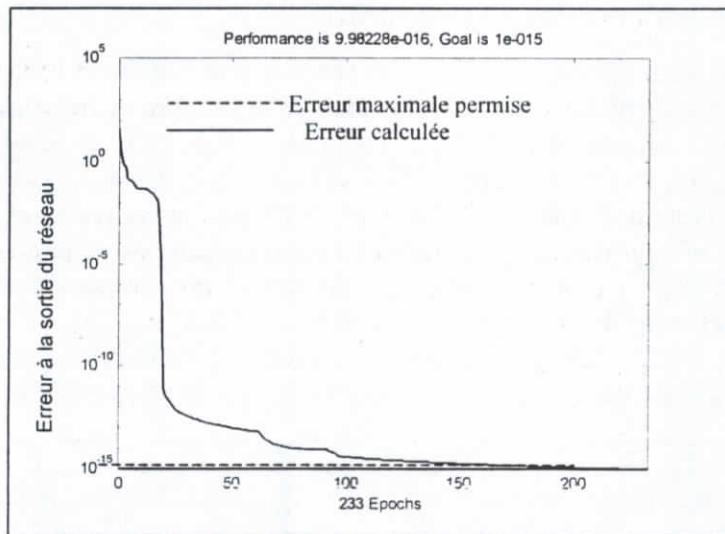


Fig.3.9 Erreur de l'apprentissage utilisant la méthode L.M à 1 couche

3.6.2 Régularisation bayésienne (RB)

Le réseau bayésien suppose que les poids et les biais sont des variables aléatoires avec les distributions spécifiques. Les paramètres de régularisation sont liés à des variances inconnues liées à ces distributions. Nous pouvons estimer alors ces paramètres par les techniques statistiques, qui ne dépendent pas de l'apprentissage du réseau [Hol-97].

Employer un réseau bayésien de régularisation, implique à employer une fonction d'apprentissage qui actualise les poids et les biais selon l'optimisation de Levenberg-Marquardt. Elle réduit au minimum une combinaison des erreurs et des poids et puis détermine la combinaison correcte afin de produire un réseau qui généralise bien. Le processus s'appelle la régularisation bayésienne. Cet algorithme peut entraîner n'importe quel réseau tant que ses poids, entrées, et fonctions de transfert sont des fonctions dérivables [Bis2-95]. L'algorithme d'optimisation des paramètres de régularisation qui est l'étape la plus importante dans cette méthode est présentée à la figure 3.10.

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

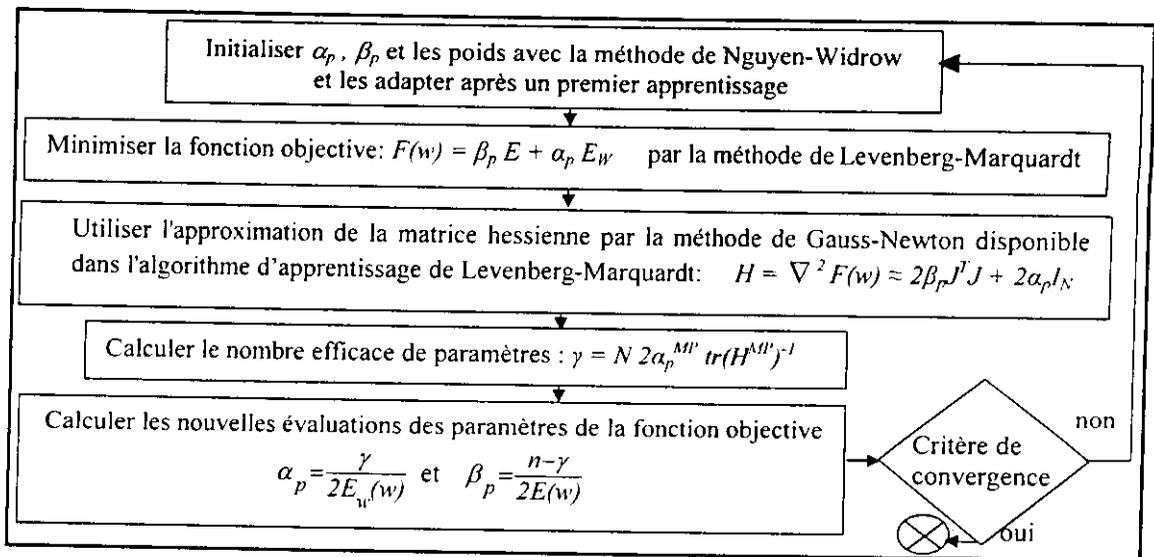


Fig.3.10 Optimisation bayésienne des paramètres de régularisation

Les paramètres donnés à la figure 3.10 sont définis comme suit:

J est la matrice jacobienne de l'ensemble des erreurs d'apprentissage et I_N est la matrice identité. γ est une mesure du nombre de paramètres dans le réseau de neurones qui sont efficacement employés en réduisant la fonction erreur. MP indique que le paramètre considéré est donné pour le point minimum, E_w représente les poids quadratiques.

La régularisation bayésienne minimise la combinaison linéaire des erreurs quadratiques et des poids. Elle modifie également la combinaison linéaire de sorte qu'à la fin de l'entraînement le réseau résultant ait de bonnes qualités de généralisation.

Cette régularisation prend place dans l'algorithme de Levenberg-Marquardt donné à la figure 3.7. La back-propagation du gradient est employée pour calculer le jacobien jX des performances, par rapport aux poids et biais que l'on note avec une variable globale (X). Chaque variable est ajustée comme suit :

$$dX = - \frac{jj + I\mu}{je} \quad (\text{III.6})$$

$$jj = jX \cdot jX \quad (\text{III.7})$$

$$je = jX \cdot E \quad (\text{III.8})$$

E représente l'erreur à la sortie du réseau, et I est la matrice identité et μ est une constante qu'on augmente jusqu'à ce que l'ajustement montré ci-dessus donne une valeur réduite de l'erreur de performance [Mus-02].

Résultat du réseau de la régularisation bayésienne

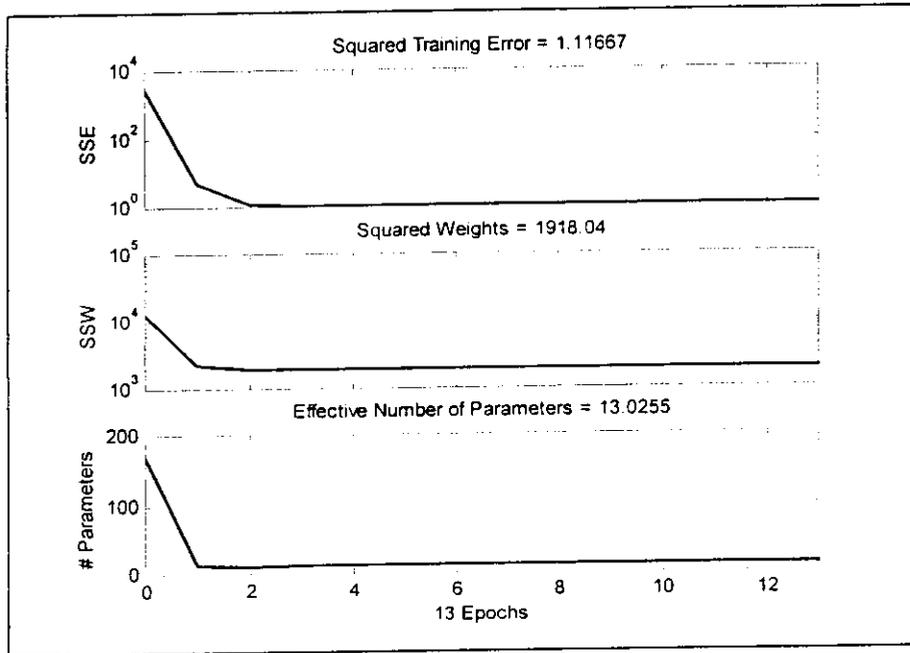
Pour ce réseau aussi, nous avons utilisé trois réseaux à une, deux et trois couches cachées et pour chacun de ces réseaux, nous avons changé le nombre de neurones. L'erreur pour les trois réseaux à une, deux et trois couches est respectivement 10^{-10} , 10^{-12} et 10^{-10} .

Nous donnons aux figures 3.11 a, b, c de haut en bas chacune, l'erreur quadratique (SSE), les poids quadratiques (SSW) et le nombre efficace de paramètres d'apprentissage (# Parameters) (voir chapitre 1 : 1.6.3.1 et 1.6.3.2) en fonction du nombre d'itérations (Epochs).

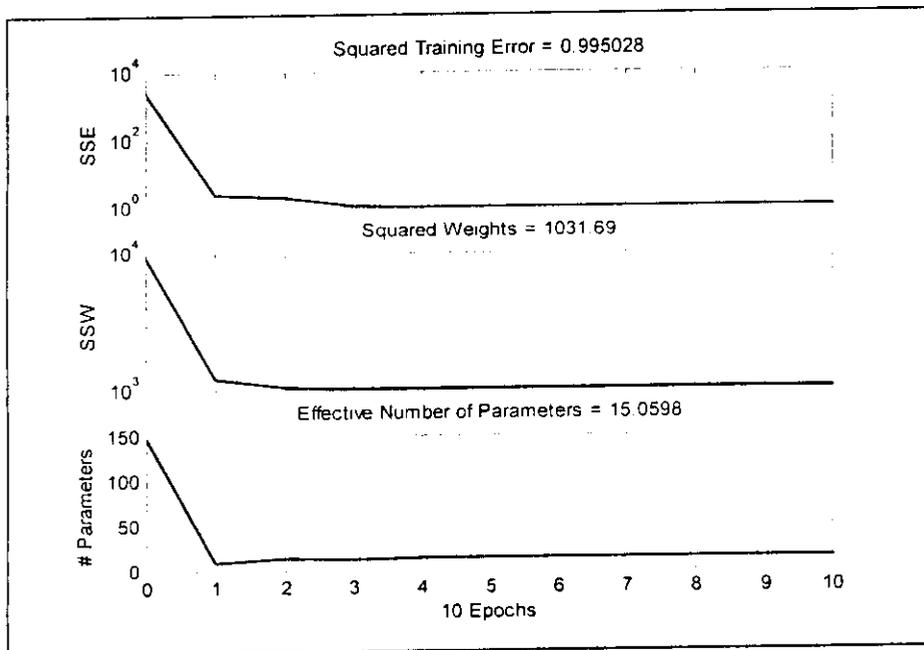
Nous présentons à la troisième courbe sur la figure 3.11-(c) à une couche cachée par exemple, la courbe d'optimisation des connexions dans le réseau RB où le nombre

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

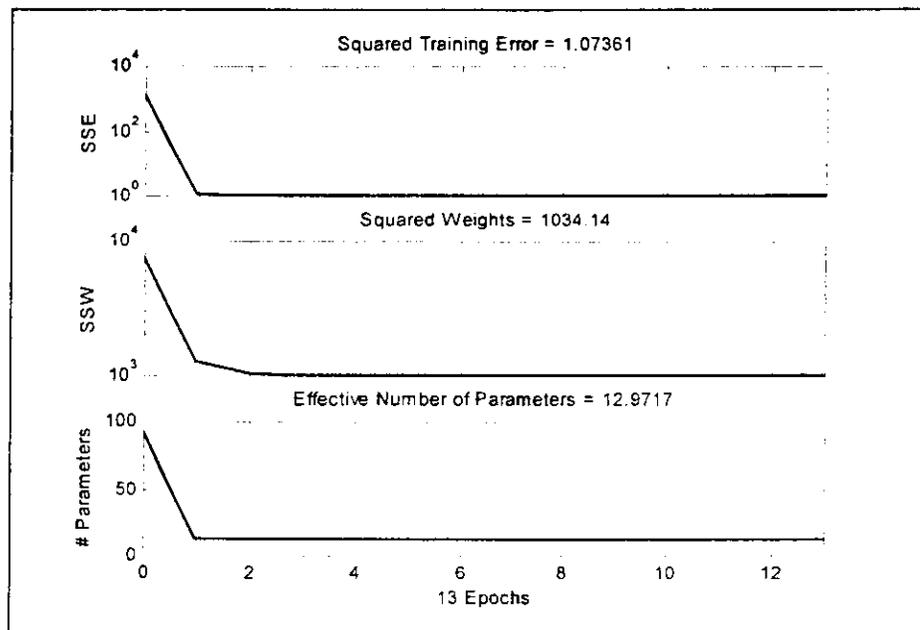
optimum des connexions représentatives dans le réseau est de **13** après un nombre initial de connexions de **92**. Le nombre optimum des connexions représentatives dans le réseau à deux couches par exemple (Fig.3.11-(b)) est de **15** après un nombre initial de connexions de **147**.



(a)



(b)



(c)

- Fig.3.11** Erreur de l'apprentissage utilisant la méthode de RB à 1, 2 et 3 couches
 (a) Erreur de l'apprentissage utilisant la méthode de RB à 3 couches 7, 8 et 4 neurones respectivement
 (b) Erreur de l'apprentissage utilisant la méthode de RB à 2 couches à 7 et 8 neurones respectivement
 (c) Erreur de l'apprentissage en utilisant la méthode de RB à 1 couche à 8 neurones

Nous remarquons que le nombre de couches cachées a une influence sur le nombre efficace de paramètres d'apprentissage.

D'après les résultats donnés dans les tableaux 3.18 à 3.20, le réseau à deux couches présente le meilleur apprentissage du point de vue précision.

Echantillon	1	2	3	4	5	1	2	3	4	5
Y_{G1}	0.826	0.040	0.044	0.042	0.040	1.108	0.030	0.089	0.030	0.062
Y_{G2}	0.110	0.032	0.697	0.032	0.679	0.106	0.120	0.902	0.000	0.895
Y_{G3}	0.0122	0.162	0.432	0.162	0.432	0.061	0.200	0.223	0.000	0.230
Y_{G4}	0.091	0.980	0.007	0.980	0.073	0.063	0.975	0.085	0.973	0.084

Tab.3.18 Diagnostic de la BORAK22 utilisant la méthode RB à 3 couches cachées : 7, 8 et 4 neurones respectivement dans les couches 1, 2 et 3

Tab.3.19 Diagnostic de la BORAK22 utilisant la méthode RB à 2 couches cachées : 7 et 8 neurones respectivement dans les couches 1 et 2.

Echantillon	1	2	3	4	5
Y_{G1}	0.998	0.001	0.001	0.001	0.001
Y_{G2}	0.016	0.108	0.805	0.108	0.804
Y_{G3}	0.013	0.193	0.343	0.190	0.340
Y_{G4}	0.005	0.918	0.150	0.918	0.150

Tab.3.20 Diagnostic de la BORAK22 utilisant la méthode RB à 1 couche cachée avec 8 neurones

Les réseaux bayésiens ont présenté la meilleure rapidité de convergence (10 à 13 itérations). D'ailleurs, en utilisant d'autres échantillons d'apprentissage et de généralisation, nous avons obtenu des résultats que nous présentons après dans le tableau 3.34 où la convergence est obtenue après une seule itération [Mok19-03, Mok21-03]. Ces réseaux généralisent bien.

3.6.3 PARAMETRES INFLUANT SUR L'APPRENTISSAGE DES RESEAUX DU DIAGNOSTIC

Les paramètres choisis ayant une influence sur l'apprentissage sont : le nombre de couches cachées, le nombre de neurones dans chaque couche, le taux d'apprentissage, la somme quadratique de l'erreur et les fonctions d'activation.

Nous avons fixé la fonction d'activation à la sortie du réseau et dans toutes les couches cachées ; à la sortie c'est la fonction linéaire 'Pureline' et dans les couches cachées ce sont des log-sigmoïde 'Logsig' [Mok21-03, Mok15-03]. Un paramètre seulement parmi cinq est modifié à la fois et les quatre autres sont gardés constants comme donné dans le tableau 3.21. Notons que quand nous utilisons deux couches cachées, le nombre de neurones dans chacune d'elles est respectivement 5 et 4 et pour trois couches, le nombre de neurones est respectivement 5, 4 et 3.

Nombre de couches cachées (NL)	1
Nombre de neurones dans la couche cachée (NbN)	6
Taux d'apprentissage (TR)	0.025
Seuil de l'erreur (Goal)	10^{-12}
Fonction d'activation (AF)	Logsig

Tab.3.21 Les paramètres influençant l'apprentissage

Certains des réseaux proposés sont stables dans leurs réponses de sortie indépendamment du changement de ses paramètres d'apprentissage. Pour les autres réseaux, influencés par le changement des paramètres d'apprentissage, un paramètre dominant a un effet important sur chaque réponse de réseau. Cette influence est donnée ci-dessous avec tous les algorithmes utilisés.

3.6.3.1 Méthode du momentum

Nous regroupons les différentes erreurs à la sortie de chaque RNA correspondant à chaque changement de paramètre d'apprentissage (Tab.3.22). Pour la méthode du momentum, la convergence est obtenue à partir d'un seuil d'erreur $\leq 10^{-2}$, donc la valeur fixe de l'erreur quand les autres paramètres changent est 10^{-2} (Tab.3.22) [Mok4-02].

	1	2	3
Nombre de couches cachées (NL)	1	2	3
Erreur à la sortie en fonction du changement de NL (%)	35	30	42
Nombre de neurones dans la couche cachée (NbN)	6	4	10
Erreur à la sortie en fonction du changement de NbN (%)	35	40	35
Taux d'apprentissage (TR)	0.025	0.15	0.5
Erreur à la sortie en fonction du changement de TR (%)	30	32	40
Erreur maximale permise	10^{-12}	10^{-2}	5.10^{-1}
Erreur à la sortie en fonction du seuil d'erreur permis (%)	Pas de convergence	35	30
Fonction d'activation (AF)	Logsig	Tansig	Purelin
Erreur à la sortie en fonction du changement de AF à l'entrée(%)	35	42	45

Tab.3.22 L'erreur à la sortie du réseau en fonction du changement de chaque paramètre d'apprentissage cité utilisant la méthode du momentum

En utilisant la méthode du momentum, nous avons noté que le paramètre d'apprentissage dominant est surtout le seuil d'erreur, puis vient le taux d'apprentissage. Le nombre de couches cachées n'a pas un grand effet sur la sortie du réseau (Tab.3.22) mais il ne peut être négligeable. En utilisant cette méthode, le réseau est instable et chaque changement de paramètres a son influence.

3.6.3.2 Méthode MADPTA

Pour cette méthode, la convergence est obtenue à partir d'une tolérance d'erreur $\leq 10^{-1}$, donc la valeur fixe de l'erreur maximale permise quand les autres paramètres changent est 10^{-1} (Tab.3.23). Le paramètre d'apprentissage dominant est le nombre de couches cachées. Le taux d'apprentissage est un paramètre influant aussi. Dans cette méthode, deux autres paramètres non cités dans le tableau 3.21 sont très importants : les deux rapports permettant d'augmenter et de diminuer le taux d'apprentissage. La fonction d'activation à l'entrée a le même effet que dans les autres réseaux.

Erreur à la sortie en fonction du changement de NL (%)	28	20	24
Erreur à la sortie en fonction du changement de NbN (%)	28	22	30
Erreur à la sortie en fonction du changement de TR (%)	28	25	38
Erreur à la sortie en fonction du changement du seuil de l'erreur permise (%)	28	30	35
Erreur à la sortie en fonction du changement de la fonction d'activation à l'entrée (%)	Pas de convergence	25	Pas de convergence

Tab.3.23 L'erreur à la sortie du réseau en fonction du changement de chaque paramètre d'apprentissage cité utilisant la MADPTA

3.6.3.3 Méthode L.M

Nous notons qu'avec la L.M, le paramètre dominant est le nombre de couches cachées. La fonction d'activation à l'entrée et le taux d'apprentissage influent considérablement sur les performances du réseau (Tab. 3.24).

Erreur à la sortie en fonction du changement de NL (%)	1	0	1.5
Erreur à la sortie en fonction du changement de NbN (%)	1	2	2.5
Erreur à la sortie en fonction du changement de TR (%)	1	1.8	4.5
Erreur à la sortie en fonction du changement du seuil de l'erreur permise (%)	1	1	0.8
Erreur à la sortie en fonction du changement de la fonction d'activation à l'entrée (%)	1	2.2	4.5

Tab.3.24 L'erreur à la sortie du réseau en fonction du changement de chaque paramètre d'apprentissage cité utilisant la LM

3.6.3.4 Méthode de régularisation Bayésienne (RB)

En utilisant la méthode RB, nous avons noté que le paramètre d'apprentissage dominant est le nombre de couches cachées comme dans la méthode L.M. Nous avons noté aussi que le réseau bayésien n'est pas sensible au changement du taux d'apprentissage et du seuil de l'erreur (Tab.3.25). La fonction d'activation à l'entrée a toujours son influence.

Erreur à la sortie en fonction du changement de NL (%)	1.5	1.	1.8
Erreur à la sortie en fonction du changement de NbN (%)	1.5	1.5	2.2
Erreur à la sortie en fonction du changement de TR (%)	1.5	2.2	3.5
Erreur à la sortie en fonction du changement du seuil de l'erreur permise (%)	1.5	2.8	3
Erreur à la sortie en fonction du changement de la fonction d'activation à l'entrée (%)	1.5	2.5	1

Tab. 3.25 L'erreur à la sortie du réseau en fonction du changement de chaque paramètre d'apprentissage cité en utilisant le réseau Bayésien

En conclusion des deux sections passées, le critère d'optimisation d'un réseau ne se base pas seulement sur la rapidité de la convergence du réseau et de l'erreur à sa sortie. Ces critères sont nécessaires mais pas suffisants pour définir le réseau optimal. L'exactitude des résultats est le critère que nous ajoutons pour définir le réseau optimal.

3.6.4 INFLUENCE DE LA BASE DE DONNEES SUR L'APPRENTISSAGE

Dans cette partie du travail, nous avons utilisé une base d'apprentissage différente de celle utilisée précédemment pour voir son influence lors de la phase de généralisation (test). Cette base de données est présentée dans le tableau 3.26. Nous soulignons que la base de tests ou généralisation est gardée la même.

3.6.4.1 Phase d'apprentissage

Nous avons noté que la méthode LM et la MADPTA (Tab.3.27, 3.28) répondent bien aux vecteurs d'entrée dans la phase d'apprentissage avec la même précision. Le réseau bayésien aussi présente de bonnes capacités d'apprentissage durant cette phase (Tab.3.29). Au début de l'apprentissage, la génération des poids a été faite par une fonction aléatoire. L'apprentissage était moins bon.

Après la première adaptation de poids, ces derniers sont ré-injectés dans le réseau et la phase est répétée une deuxième fois avec le nouvel ensemble de poids adaptés. De cette façon, l'apprentissage est bien amélioré pour la L.M et la MADPTA. Pour la méthode du momentum, le réseau n'a pas pu donner un diagnostic correct dans la phase d'apprentissage (Tab.3.26) [Mok4-02]. Y_L est la sortie du réseau durant la phase d'apprentissage.

Échantillon N°	1	2	3	4	5
P_1	1	3	2	4	4
P_2 (est)	12.54	11	12.3	11	10
P_3 (mgKOH/g)	0.046	0.106	0.072	0.09	0.3
P_4 (kV)	56	50	52	42	30
P_5	0.015	0.025	0.02	0.082	0.6
P_6 (ppm)	31	35	32	37	42
Y_{L1}	0.320	0.351	0.441	0.536	0.359
Y_{L2}	0.042	0.365	0.288	0.332	0.894
Y_{L3}	0.302	1.304	0.401	1.290	0.319
Y_{L4}	0.810	0.433	1.201	0.557	0.120

Tab.3.26 Diagnostic dans la phase d'apprentissage utilisant la BP avec momentum pour une nouvelle base d'apprentissage

Échantillon N°	1	2	3	4	5
Y_{L1}	0.0005	-0.0004	-0.0052	0.0009	1.0003
Y_{L2}	0.0004	1.0001	0.0006	0.0000	0.0005
Y_{L3}	0.0031	0.0810	0.0001	-0.9990	0.0075
Y_{L4}	1.0005	0.0013	1.0002	0.0015	-0.0001

Tab.3.27 Diagnostic dans la phase d'apprentissage utilisant la MADPTA avec une nouvelle base d'apprentissage

Échantillon N°	1	2	3	4	5
Y_{L1}	0.0002	-0.0004	-0.0002	0.0009	1.0002
Y_{L2}	0.0004	1.0006	0.0006	0.0000	0.0005
Y_{L3}	0.0001	0.0008	0.0001	0.9991	0.0005
Y_{L4}	1.0005	0.0003	1.0006	0.0004	0.0001

Tab.3.28 Diagnostic dans la phase d'apprentissage utilisant la L.M avec une nouvelle base d'apprentissage

Échantillon N°	1	2	3	4	5
Y_{L1}	0.0002	-0.0004	-0.0002	0.0009	1.0901
Y_{L2}	0.0004	0.7651	0.0006	0.0000	0.0005
Y_{L3}	0.0001	0.0008	0.0001	0.7562	0.0005
Y_{L4}	0.9120	0.0003	0.8610	0.0004	0.0001

Tab.3.29 Diagnostic dans la phase d'apprentissage utilisant le RB avec une nouvelle base d'apprentissage

3.6.4.2 Phase de généralisation

Afin d'examiner l'efficacité des réseaux présentés dans cette section, nous avons utilisé les mêmes échantillons d'huile employés dans les sections précédentes (3.6.1, 3.6.2) dans la phase de généralisation. En utilisant la méthode du momentum, (Tab.3.30), le réseau n'a pas complètement convergé mais les résultats de la phase de test ont été clairement améliorés. Une décision donnée par la sortie du réseau entraîné par le momentum n'est pas correcte. Le système a fait un diagnostic correct dans les cinq échantillons en utilisant les méthodes MADPTA, L.M et RB (Tab.3.31, 3.32, 3.33).

Nous avons fixé l'erreur maximale permise pour les réseaux du momentum, MADPTA L.M et RB respectivement à 10^{-2} , 10^{-5} , 10^{-13} et 10^{-8} . Nous notons que la méthode L.M est la meilleure pour cette application dans le diagnostic, quoique la qualité des résultats s'est dégradée comparée avec les résultats obtenus avec la première base d'apprentissage donnée dans le tableau 3.8. [Mok21-03, Mok15-03, Mok26-03, Mok4-02, Mok7-02].

Échantillon N°	1	2	3	4	5
P ₁	4	1	3	2	3
P ₂ (cst)	11	12	10	12.5	11
P ₃ (mgKOH/g)	0.4	0.04	0.08	0.02	0.1
P ₄ (kV)	30	60	47	56	42
P ₅	0.5	0.04	0.02	0.02	0.1
P ₆ (ppm)	40	33	35	32	37
Y _{G1}	0.919	0.005	0.404	0.046	0.109
Y _{G2}	0.325	-0.195	0.428	-0.182	0.489
Y _{G3}	0.011	-0.157	0.541	0.018	0.350
Y _{G4}	-0.082	1.118	0.037	0.625	0.415

Tab.3.30 Diagnostic dans la phase de généralisation utilisant la méthode du momentum

Échantillon N°	1	2	3	4	5
Y _{G1}	1.061	0.120	0.140	0.0018	0.8010
Y _{G2}	0.2001	0.3041	1.200	-0.0041	0.8685
Y _{G3}	0.1007	0.4236	0.333	0.1202	0.4001
Y _{G4}	0.0100	0.980	0.001	1.1165	-0.1126

Tab.3.31 Diagnostic dans la phase de généralisation utilisant la méthode MADPTA

Echantillon N°	1	2	3	4	5
Y _{G1}	1.001	-0.400	-0.002	0.000	0.4300
Y _{G2}	0.0001	0.2001	1.1005	-0.0001	0.86185
Y _{G3}	0.0007	0.006	0.3003	0.0002	0.4201
Y _{G4}	0.0000	1.005	0.000	1.0005	-0.1123

Tab.3.32 Diagnostic dans la phase de généralisation utilisant la méthode L.M

Échantillon N°	1	2	3	4	5
Y _{G1}	1.0000	-0.0002	-0.0020	0.0001	-0.0002
Y _{G2}	-0.0001	0.0001	1.0005	-0.0001	1.0033
Y _{G3}	0.0001	0.0001	0.0020	0.0002	0.0003
Y _{G4}	-0.0002	1.0055	0.0070	1.0005	-0.0003

Tab.3.33 Diagnostic dans la phase de généralisation utilisant le réseau Bayésien

Le réseau MADPTA a donné cinq décisions correctes, mais le problème est au niveau de l'échantillon N°5. Les sorties Y_{G1} et Y_{G2} sont à 0.8 et 0.86 respectivement. Donc, les deux sorties sont activées à la fois : changer et traiter. Le réseau est indécis pour cet échantillon.

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

Pour la méthode Bayésienne, elle donne des résultats exacts avec les meilleurs paramètres d'apprentissage (Tab.3.33) : nombre d'itérations (épochs) = 1 et l'erreur tolérée = 10^{-14} . La caractéristique principale de cette méthode est sa capacité de généraliser l'apprentissage sur d'autres exemples. Comparant les résultats des autres réseaux avec ceux de ce réseau, nous voyons bien que ce dernier présente les meilleures performances (Tab.3.34) [Mok19-03].

Nous présentons dans le tableau 3.34 les paramètres les plus importants dans un apprentissage (la vitesse et la précision des réseaux) pour les 4 réseaux utilisés. Ces résultats correspondent à la nouvelle base d'apprentissage que nous avons utilisée dans l'étude présentée dans cette section.

	Réseau L.M	Réseau du momentum	Réseau MADPTA	Réseau Bayésien
Nombre d'itérations	93	1000	1000	1
Erreur à la sortie	10^{-13}	10^{-2}	10^{-5}	10^{-14}

Tab.3.34 Paramètres donnant la précision et la vitesse de convergence pour les réseaux de diagnostic

En conclusion, nous confirmons la grande influence de la base d'apprentissage sur la qualité de l'apprentissage du réseau et ses capacités de généralisation. Nous confirmons aussi l'indépendance des réseaux bayésiens du modèle à traiter.

3.7 ARCHITECTURES ET APPRENTISSAGES DES RNA UTILISES DANS LA CLASSIFICATION DE LA BORAK22

Ces réseaux peuvent classer un ensemble d'exemples de la base d'apprentissage en un nombre spécifique de catégories sans savoir par avance comment les organiser. Ce sont les données elles-mêmes qui servent de réponses car les poids du réseau sont modifiés en fonction de leurs composantes. Dans ce type de réseaux, l'information utile qui participe dans l'apprentissage, se trouve donc uniquement dans les données. Les réseaux utilisés pour la classification dans les prochaines sections sont les réseaux compétitifs, les réseaux auto-organisés de Kohonen et les réseaux bayésiens.

3.7.1 Les Réseaux Compétitifs

Pour la classification de l'huile de transformateur, nous gardons la même base d'apprentissage et de généralisation utilisée dans les quatre réseaux du diagnostic. En utilisant la fonction d'apprentissage 'By-weight and bias, 1-vector-at-a-time', le réseau compétitif trouve l'organisation de ses neurones et donne un neurone gagnant pour chaque vecteur d'entrée.

Dans l'application que nous présentons ci-dessous, le réseau partage les entrées en trois classes. Le réseau utilisé contient 10 neurones [Mok16-02, Mok19-03, Mok25-03].

Dans le premier échantillon par exemple, le neurone gagnant N_W est le 2^{ème} neurone dont la sortie $Y_G = 1$, il représente une classe de données. En utilisant les mêmes échantillons donnés dans les méthodes supervisées, les résultats sont donnés par l'état de la sortie Y_G dans le tableau 3.35. Les chiffres 1, 2, 3, 4, 5 dans la parenthèse indique le N° de l'échantillon que nous notons Nech.

Notons que pour le deuxième et quatrième vecteur d'entrée, le neurone gagnant est le même (le 6^{ème}) ce qui signifie qu'il représente la même classe de données qui est différente de celle représentée par le neurone gagnant N° 2. En utilisant les mêmes échantillons donnés dans les méthodes supervisées durant la phase de tests, les résultats sont donnés par l'état de la sortie Y_G dans le tableau 3.35. Ceci est en concordance avec les résultats donnés par les réseaux de diagnostic qui présentent aussi trois classes. D'autres résultats sont donnés pour la base d'apprentissage présentée dans le tableau 3.26 [Mok21-03, Mok19-03] et la sortie Y_G donne quatre classes dans le tableau 3.36

$(N_w, Nech)$	Y_G	$(N_w, Nech)$	Y_G
(2,1)	1	(2,1)	1
(6,2)	1	(5,2)	1
(7,3)	1	(7,3)	1
(6,4)	1	(5,4)	1
(7,5)	1	(8,5)	1

Tab.3.35 Classification de la BORAK22 utilisant un réseau compétitif

Tab3.36 Classification d'un autre échantillon de la BORAK22 utilisant un réseau compétitif

La distribution des 10 neurones utilisés dans la couche compétitive est donnée par la figure 3.12. Dans ce réseau, nous avons réparti les neurones à une dimension.

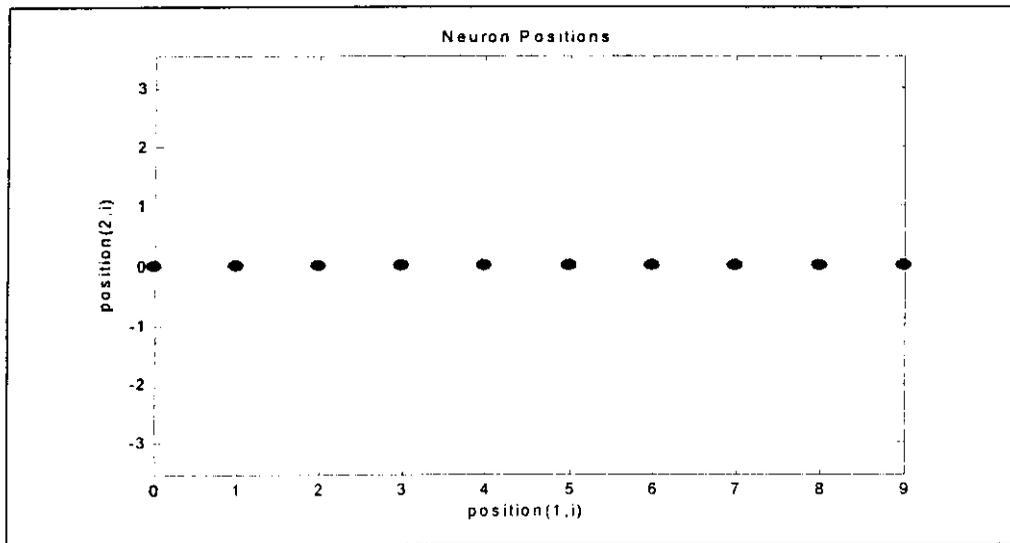


Fig.3.12 Distribution des neurones de la couche compétitive utilisée pour la classification

3.7.2 Les cartes auto-organisées de Kohonen (Self-Organization Maps -SOM)

Nous avons utilisé différentes distributions et topologies de la carte de Kohonen : le rayon du voisinage ($D=3$ et 1 , Fig.3.13), et la topologie des neurones (Hexagonal et aléatoire) (Fig.3.14, Fig.3.15).

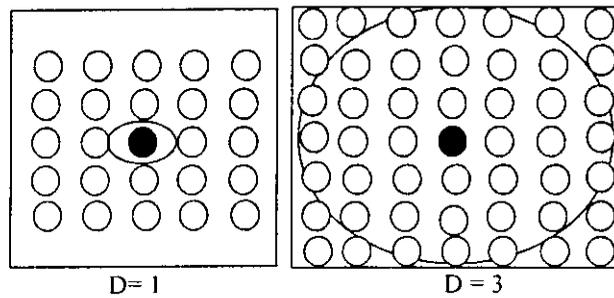


Fig3.13 Différents rayons de voisinage autour du neurone gagnant

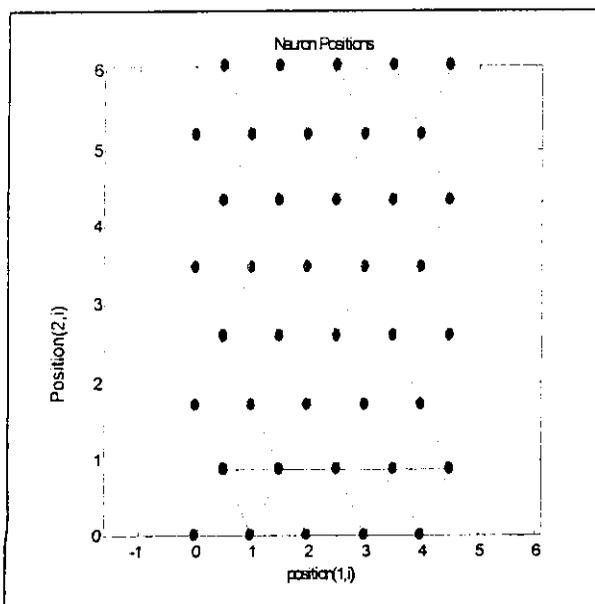


Fig. 3.14 Distribution et positions des neurones dans la carte de Kohonen en topologie hexagonale

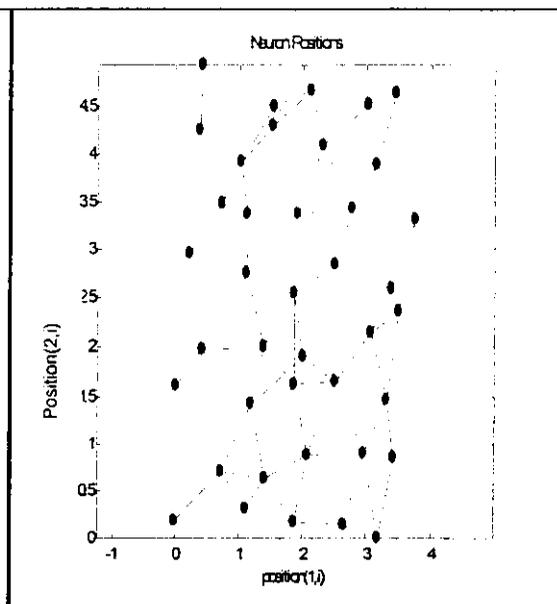


Fig. 3.15 Distribution et positions des neurones dans la carte de Kohonen en topologie aléatoire

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

Pendant l'apprentissage, le SOM trouve le neurone de sortie qui a la plus petite distance à partir des données d'apprentissage. Le SOM change les poids des neurones pour augmenter leur(s) similitude(s) avec les données, et il influe sur les poids des neurones voisins qui ont des relations aléatoires avec les données d'apprentissage.

Les neurones de sortie sont déplacés aux positions qui tracent la distribution des données de l'apprentissage. Après l'entraînement, les poids de chaque neurone modélisent les caractéristiques d'une classe dans les données [Kar-96, Mok19-03].

Le nombre de neurones utilisé est 40. Les résultats donnés dans les deux SOM utilisant $D=1$ et $D=3$ et ayant une topologie hexagonale des neurones, sont donnés dans les tableaux 3.37 et 3.38 respectivement. Nous remarquons que dans les deux cas, les deux réseaux donnent trois classes représentées respectivement par le 1^{er} échantillon, le 2^{ème} et 4^{ème} et enfin le 3^{ème} échantillon et le 5^{ème}.

(N _w , Nech)	Y _G	(N _w , Nech)	Y _G
(1,1)	1	(40,1)	1
(40,2)	1	(1,2)	1
(3,3)	1	(3,3)	1
(40,4)	1	(1,4)	1
(3,5)	1	(3,5)	1

Tab.3.37 Classification de la BORAK22 utilisant un SOM Hexagonal $D = 1$

Tab.3.38 Classification de la BORAK22 utilisant un SOM Hexagonal $D = 3$

En utilisant la topologie aléatoire (Fig.3.15) les résultats de classification sont donnés dans les tableaux 3.39 et 3.40. Les classes trouvées sont les mêmes données dans les tableaux 37 et 38 [Mok19-03, Mok16-02].

(N _w , Nech)	Y _G	(N _w , Nech)	Y _G
(2,1)	1	(11,1)	1
(38,2)	1	(1,2)	1
(1,3)	1	(2,3)	1
(38,4)	1	(1,4)	1
(1,5)	1	(2,5)	1

Tab.3.39 Classification de la BORAK22 utilisant un SOM Aléatoire $D = 1$

Tab.3.40 Classification de la BORAK22 utilisant un SOM Aléatoire $D = 3$

3.7.3 Les réseaux bayesiens

Sachant que les classes de l'huile sont définies par le vecteur de sortie suivant:

[mauvaise à jeter, polluée et nécessite un traitement, polluée et il faut régénérer, bonne à garder] ; ces sorties prennent des valeurs comprises dans les intervalles suivants respectivement : $[\{0-0.154\}, \{0.155-0.254\}, \{0.255-0.354\}, \{0.355-0.45\}]$.

Pour les mêmes échantillons d'huile présentés le long de ce travail, la sortie du réseau bayésien a classé les données dans le tableau 3.41 comme suit [Mok19-03]:

Nech	1	2	3	4	5
Y _G	0.4153	0.0976	0.2746	0.0953	0.2945

Tab.3.41 Classification de la BORAK22 utilisant un réseau bayésien (RB)

Rappelons que le paramètre Nech donne le numéro de l'échantillon d'huile.

Les trois RNA ont de bonnes capacités dans la classification de données. La classification de l'huile peut être utilisée comme une phase d'aide à la décision.

Nous remarquons que la sortie du réseau bayésien a donné des résultats qui concordent avec ceux donnés par les réseaux du diagnostic. Comme le RB est un réseau supervisé, les sorties sont données selon un modèle présenté par le superviseur (cible), donc les classes sont indiquées explicitement. L'allure de l'erreur est donnée à la figure 3.16.

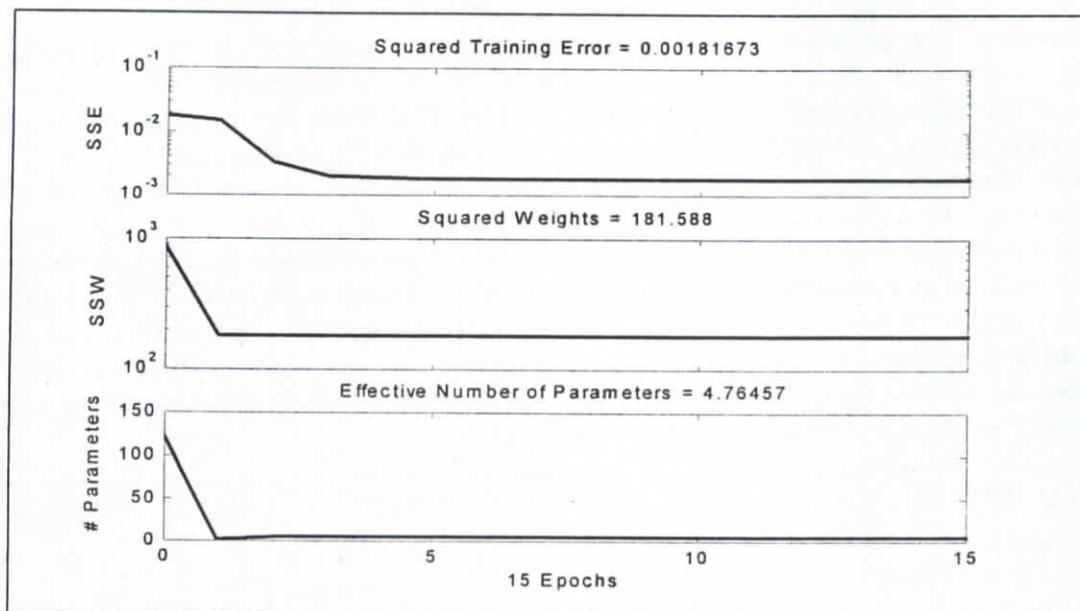


Fig.3.16 Erreur de l'apprentissage pour la classification de la BORAK22 utilisant la méthode RB à 2 couches avec 7, 8 neurones respectivement

Les réseaux non supervisés sont limités dans leurs applications par rapport aux réseaux supervisés ; ils sont utilisés en général dans les problèmes de classification et de reconnaissance où les données de sortie durant la phase d'apprentissage ne sont pas connues a priori. Si ces réseaux sont utilisés dans d'autres applications, c'est qu'ils sont en général couplés avec des réseaux supervisés.

Les réseaux non supervisés ont montré une grande souplesse d'apprendre par rapport aux réseaux supervisés. Ils sont plus stables vis-à-vis des paramètres d'apprentissage (architecture du réseau et coefficients intervenant dans l'apprentissage).

3.8 DISCUSSIONS

Le problème des capacités de généralisation des réseaux utilisés est lié directement à la taille de la base de données obtenue de SONELGAZ. Dans les archives du laboratoire de Contrôles, Mesures et Essais, les tests ne sont jamais tous réalisés. Ils sont au maximum trois pour le même échantillon: la couleur, la rigidité diélectrique et l'acidité. La base de données que nous avons pu obtenir est composée de valeurs de test de 15 échantillons d'huile. L'obtention d'une base de données plus large va mettre à l'épreuve les capacités de généralisation des réseaux utilisés.

L'utilisation des réseaux bayesiens est très intéressante pour leur robustesse d'une part et d'autre part pour leur caractéristique d'optimiser le nombre de connexions du réseau. Après l'optimisation, le réseau peut être remplacé par ces nouvelles connexions efficaces.

La méthode de Levenberg-Marquardt présente aussi de bonnes performances ; en plus, l'apprentissage se fait d'une façon assez simple en remplaçant la détermination de la matrice hessienne par un calcul simple.

L'optimisation d'un réseau est un but important dans l'amélioration de la qualité de l'apprentissage des RN. Des chercheurs [Din-95, Zha-97] ont défini un RN optimal par une convergence rapide et une petite erreur. Mais la convergence rapide reflète une

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

régularité (monotonie) dans la base d'apprentissage qui peut ne pas couvrir tous les cas pouvant exister dans l'état de l'huile surtout si cette base d'apprentissage n'est pas assez large pour être représentative. L'exactitude des résultats est le seul paramètre qui peut confirmer l'optimisation du réseau utilisé et ceci exige la connaissance des sorties (décisions) de la base de généralisation (test) aussi. Pour cela, une large base de données est nécessaire.

Chacun des paramètres d'apprentissage en liaison avec l'architecture du réseau (nombre de couches cachées, nombre de neurones et leurs distribution,...) ou avec les valeurs numériques des paramètres d'apprentissage (taux d'apprentissage, seuil d'erreur, les fonctions d'activation, ...) a une influence sur le déroulement de l'apprentissage et sa qualité. Leur optimisation n'est pas une tâche facile parfois.

L'optimisation d'un réseau de neurone n'est pas obtenue seulement par une convergence rapide et par l'erreur la plus petite. Ce sont des paramètres nécessaires mais pas suffisants pour la définition d'un réseau optimal.

Quand le vecteur d'entrée est volumineux, la classification servira à séparer les données selon leurs états dans des classes. A partir de là, la phase de diagnostic devient plus facile à faire même sans l'aide du système des RNA. Il suffit de connaître les quatre cas de diagnostic donnés dans quatre vecteurs d'entrée et de les insérer dans la base de données. Leur classification donnera le diagnostic pour les autres données.

3.9 CONCLUSION

Les essais les plus répandus pour faire le diagnostic de l'état des transformateurs sont les essais d'huile. L'utilisation d'autres essais (off-line et on-line) s'accroît mais reste limitée par un certain nombre de facteurs, dont:

- Coût : le coût élevé des essais et de la surveillance peut rendre difficile la justification des essais.
- Interprétation de données : l'interprétation des essais exige souvent un personnel expert. L'interprétation incorrecte des données peut mener à des conclusions fausses au sujet de l'état du transformateur.

Beaucoup de travaux de recherche développent des techniques d'analyse standard qui permettraient au personnel de terrain l'utilisation des résultats d'essai plus facilement et réduiront le besoin d'interprétation par des experts. On s'attend à ce que des logiciels qui combinent les résultats de différents essais et qui donnent une évaluation globale des conditions trouvent une utilisation croissante. Les réseaux de neurones en font partie et leurs applications ont prouvé leur capacité d'apprendre et d'évaluer l'état de l'huile de transformateurs.

- Les essais peuvent et devraient être employés comme référence pour le système de diagnostic.
- Une surveillance continue des paramètres appropriés est nécessaire pour un diagnostic fiable.
- La connaissance du matériau isolant, la technique de mesure, les appareils et le système sont très importants pour le diagnostic.
- Les réseaux de neurones supervisés et non supervisés sont de bons systèmes pour le diagnostic et la classification pour les nouveaux ingénieurs analystes non-expérimentés pour prendre les bonnes décisions concernant l'état de l'huile et en un temps réduit.

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

- L'expérience des anciens experts analystes peut être enseignée aux RNA et bien conservée.
- Les réseaux de Levenberg-Marquardt et les réseaux bayesiens présentent une robustesse pour le diagnostic.
- L'architecture du RNA (nombre de couches cachées, nombre de neurones) est un paramètre important décidant de la qualité de l'apprentissage plus que les paramètres d'apprentissage (taux d'apprentissage, ...).
- La taille et la qualité de la base de données ont une influence majeure sur l'apprentissage des RNA.
- La classification de l'huile peut être utilisée comme une phase d'aide à la décision.
- Les réseaux non supervisés ont montré une grande souplesse dans la classification et l'apprentissage de ces réseaux et ne présente pas beaucoup de complexité comme dans certains réseaux supervisés.

3.10 BIBLIOGRAPHIE

- [Aln-96] B. Al-Najjar, 'Total quality maintenance an approach for continuous reduction in costs of quality products', *Journal of Quality in Maintenance Engineering*, Vol. 2 No. 3, pp. 4-20. MCB University Press, 1996.
- [Ara-02] V.G. Ara-02elian, 'Efective diagnostics for oil-filled equipment', *IEEE-Electrical Insulation Magazine*, Vol. 18, N° 6, pp: 12-25, Nov-Dec. 2002.
- [Bel1-99] M. Belanger, 'Transformer diagnosis: Part 1: A statistical justification for preventative maintenance', *Electricity today*, Vol. 11, N° 6, pp: 5-8, 1999.
- [Bel2-99] M. Belanger, 'Transformer diagnosis: Part 2: A look at the reference data for interpreting test results', *Electricity today*, Vol. 11, N° 7, pp: 19-26, 1999.
- [Bis-95] C.M. Bishop, 'Neural networks for pattern recognition', Oxford University Press-1995.
- [Bre-92] G.Breen, 'Essential requirements to maintain transformers in service', *Int. Council on large electric systems (CIGRÉ)*, Paris, France, 1992.
- [Cig1-83] CIGRÉ Working Group 05, 'An international survey on failures large power transformers in service', *Electra*, N° 88, May 1983.
- [Cig2-93] CIGRÉ Working Group 09 of study committee 12, 'Lifetime evaluation of transformers', *Electra*, N° 150, pp: 39-51, 1993.
- [Cor-99] A.J.M. Corso, L.M.R. Oliveira, 'Condition monitoring and diagnostics of power transformers', *Int. J. COMADEM*, Sunderland, UK, pp: 5-11, 1999.
- [Dar3-98] M. Darveniza, et al., 'Investigations into effective methods for assessing the condition of insulation in aged power transformers', *IEEE Trans. Power Delivery*, Vol. 13, N° 4, pp: 1214-1223, 1998.
- [Din-95] X. Ding, E. Yao, Y. Liu, P. J. Griffin, 'ANN Based Transformer fault Diagnosis Using Gas-in-oil Analysis', *Proceedings of the 57th American Power Conference*, Chicago IL, Apr. 1995.
- [Dor-01] R. Dor et al., 'On-line transformer monitoring : Detection of partial discharges from HF measurements using FFT and time domain filters', *12th International Symposium on High Voltage Eng., ISH'2001*, Volume 5, pp: 1230-1233, Banagalore, India 2001.
- [Duk-98] G. Duk, 'Predictive maintenance a case study in infrared thermography', *Electrical Maintenance*, pp: 11-12, 1998.

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

- [Fal-75] B. Fallou, 'Detection of and research for the characteristics of an incipient fault from analysis of dissolved gases in the oil of an insulation', *Electra*, N° 42, pp: 31-52, 1975.
- [Goc-99] E. Gocenbach, 'Testing and monitoring as basis of the dielectric diagnostic', *ISH'99*, IEE Conference publishing, paper 5.1.S10, London, UK, 22-27 Aug 1999.
- [Gre-96] O.N. Grechko, I. Kalacheva, 'Current trends in the development of in-service monitoring and diagnostic systems for 110-750 kV power transformers', *Applied Energy, Russian Journal of fuel, power and heat systems*, Vol.34, N° 5, pp: 84-97, 1996.
- [Ham-97] A. Hamimi, A.Boubakeur, M.Lebid, 'Influence de la température sur certaines propriétés de l'huile de transformateur utilisée par la SONELGAZ', 2^{ème} CNHT pp : 102-104, Tizi-Ouzou, Nov. 1997.
- [Har-97] J.W. Harey, 'CIGRÉ Working Group 12.18 TF02, 'Survey on diagnostics & monitoring techniques transformer subsystems', *CIGRÉ SC 12*, Sydney Colloquium, Australia, 1997.
- [Hoc-88] B.Hocard, 'Transformateur de puissance', Editions Eyrolles, 1988.
- [Hol-97] A. Holst, 'The use of Bayesian neural network model for classification tasks', PHD thesis, Dept. of numerical analysis and computer science, Royal institute of technology, Stockholm, Sweden, TRITA-NA, 1997.
- [Ida-03] voir Internet IDA200, GE Energy Services, Programma Products, 2003.
- [IEC-92] 'Guide for the sampling of gases and oil from oil-filled electrical equipment and for the analysis of free and dissolved gases', IEC Publication 567, 1992.
- [IEEE-91] IEEE guide for the interpretation of gases generated in oil-immersed transformers, IEEE Std. C57.104-1991, 1991.
- [Jar-98] P.N. Jarman, J.A. Lapworth, A. Wolson, 'Life assessment of 275 and 400 kV transmission transformers', *Int.Council on large electric systems (CIGRÉ)*, Paris, France, 1998.
- [Jer-93] T. Jervis, 'Connectionist adaptive control', PHD thesis at the university of Cambridge, 1993.
- [Kac-97] A.J. Kacler, 'Diagnostic and monitoring-technology for large power transformer', *CIGRE SC 12 colloquium*, Sidney, Australia, 1997.
- [Kar-96] S.V.Karalopoulos, 'Understanding neural networks and fuzzy logic- basic concepts and applications', IEEE Press, IEEE order number : PP5591, 1996.
- [Kaz-98] M. Kazierski, R. Sobocki, W. Olech, 'Selected elements of life management of large power transformers – A Polish experience', *Int.Council on large electric systems (CIGRÉ)*, Paris, France, 1998.
- [Kog-88] V.I. Kogan et al., 'Failure analysis of EHV transformers', *IEEE Trans. Power Delivery*, vol.3, N°2, pp: 672-683, 1988.
- [Kum-00] U. Kumar, J.P. Liyanage, 'Design and development of maintenance measurement system for the offshore industry', 3rd International conference on Quality, Reliability and Maintenance (QRM), Professional Publishers IMecE, pp 227-230, Oxford, UK, Apr. 2000.
- [Lap-95] J.A.Lapworth, P.N. Jarman, I.L. Funnel, 'Condition assessment techniques for large power transformers', 2nd International Conf. Reliability of transmission and distribution equipment, , Conf. Publ. N° 406, pp: 85-90, Coventry, UK, 1995.
- [Mal-00] R. Malewski, M. Kaz-98ierski, 'Diagnostic techniques for power transformers', *Int.Council on large electric systems (CIGRÉ)*, Paris, France, 2000.

- [Mye-98] C. Myers, 'Transformers – Condition Monitoring by Oil Analysis, Large or Small; Contentment or Catastrophe', Proceedings of the 1st IEE/IMEChE Int. Conf. on Power Station Maintenance – Profitability through Reliability”, pp.53-58, 1998.
- [Mok4-02] L.Mokhnache, A.Boubakeur, N.Nait Said, 'Application of Neural Networks Paradigms in the Diagnosis and Thermal Ageing Prediction of Transformer Oil', Int. Conf. on Dielectric Liquids-ICDL'2002 proceedings, IEEE-DEIS, Paper N° 048, pp: 258-261, Graz, Austria, Juil. 2002.
- [Mok5-02] L. Mokhnache, A. Boubakeur, N.Nait Said, 'Application Of Neural Networks In The Diagnosis Of Transformer Oil Used By Sonelgaz', 4^{ème} CNHT Série B-Numéro special AJOT, pp. 51-54, Ghardaia, Algérie, Jan. 2002.
- [Mok7-02] L.Mokhnache, A.Boubakeur, 'Comparison of Different Neural Networks Algorithms Used in the Diagnosis of Transformer Oil', IEEE-DEIS proceedings CEIDP'2002, pp: 244-247, Cancun, Quintana Roo, Mexico, Oct 2002.
- [Mok13-02] L. Mokhnache, A. Boubakeur, N.Nait Said, 'Diagnosis Of Transformer Oil Using Some Neural Networks Algorithms', 4th QRM, Publishers IMecE, pp.191-194, Oxford, UK, 2002.
- [Mok14-02] L.Mokhnache, A.Boubakeur, 'The use of some back-propagation paradigms in the classification of the insulating transformer oil', Proceeding of the Int.. Federation of Operational Research Societies-IFORS'2002, session mb22, pp: 22, Edenburgh, UK, Juil. 2002.
- [Mok15-03] L.Mokhnache, A.Boubakeur, N. Nait Said, 'Some Neural networks Paradigms In Insulating Transformer Oil Maintenance decision', 2nd Int.. Conf. of Electrical and Power Engineering EPE, Iasi, Romania, Nov. 2002.
- [Mok16-02] L.Mokhnache , A.Boubakeur, N.Nait Said, 'Comparison of Different Neural Networks Algorithms Used in the Diagnosis and thermal ageing prediction of Transformer Oil', Int. Conf. on Systems, Man & Cybernetics, IEEE-SMC CD-ROM, paper ID: SMC02-REG-485, Hammamat, Tunisia Oct.,2002.
- [Mok19-03] L.Mokhnache, A.Boubakeur, 'Classification Of Transformer Oil Using Self-Organizing Networks And Bayesian Neural Networks, Int. Conf. on Energy Efficiency ICEE CD-ROM, Algiers-Algeria, May 2003.
- [Mok21-03] L.Mokhnache, A.Boubakeur, N.Nait Said, 'Influencing training parameters in some supervised neural networks for transformer oil diagnosis', Int.Conf. of COMADEM'03, Växjö, Sweden, Aug. 2003.
- [Mok25-03] L.Mokhnache, A.Boubakeur, 'Application of Many Neural Networks Paradigms in the High Voltage Insulation, Archives of Electrical Engineering (AEE), Polish Academy of Science, vol.3, Oct. 2003.
- Mok26-03] L.Mokhnache, A.Boubakeur, N.Nait Said, Ali Feliachi, R. Kattan, 'Supervised and Unsupervised Neural Networks Used in the classification and Diagnosis of Transformer Oil', submitted to IEEE-Energy Conversion Transactions 2003.
- [Mus-02] T. Musicka and all, 'Steel processing using a Bayesian adaptive network', 4th QRM, published by IMecE, pp.191-194, Oxford, UK, 2002.
- [Pet1-90] L. Pettersson, 'Estimation of the remaining service life of power transformer and their insulation', Electra, N° 133, pp: 65-71, 1990.
- [Pet2-98] L. Pettersson, N.L. Fantana, U. Sundermann, 'Life assessment ranking of power transformers using condition based evaluation – Anew approach', Int.Council on large electric systems (CIGRÉ), Paris, France, 1998.

Surveillance et diagnostic de l'huile de transformateur et applications des RNA

- [Sah1-97] T.K. Saha, et al., 'Electrical and chemical diagnostics of transformers insulation – Part A: Aged transformers samples', IEEE Trans. Power Delivery, Vol. 12, N°4, pp:1547-1554, 1997.
- [Sah2-97] T.K. Saha, et al., 'Electrical and chemical diagnostics of transformers insulation – Part B: Accelerated aged insulation samples', IEEE Trans. Power Delivery, Vol. 12, N°4, pp: 1555-1561, 1997.
- [Tan-98] p. Tantin, P.V. Goosen, P. Christensen, 'CIGRÉ SC 12 Power transformers special report', Int.Council on large electric systems (CIGRÉ), Paris, France, 1998.
- [Ten-00] S. Tenbohlen, et al., 'Enhanced diagnosis of power transformers using on- and off-line methods: results, examples and future trends', Int.Council on large electric systems (CIGRÉ), Paris, France, 2000.
- [Tom-93] K. Tomovic, M. Tapper, T. Ingvarsson, 'Performance evaluation of a transformer condition monitoring expert system', Int.Council on large electric systems (CIGRÉ), Berlin, Germany, 1993.
- [Wan2-00] Z. Wang, Y.Liu, P.J. Griffin, 'Neural net and expert system diagnose transformer faults', IEEE Computer Applications in Power, Vol.13, N° 1, pp: 50-55, 2000.
- [Wam-02] M.Wang, A.J. Vandermaar, K.D. Srivastava, 'Review of condition assessment of power transformer in service', IEEE-Electrical Insulation Magazine, Vol. 18, N° 6, pp: 12-25, Nov-Dec. 2002.
- [Wan3-00] Z. Wang, 'Artificial Intelligence Applications in the Diagnosis of Power Transformer Incipient Faults', Dissertation submitted to the Faculty of the Virginia In partial fulfilment of the requirement for the degree of Polytechnic Institute and State University Doctor of Philosophy In Electrical Engineering, Aug. 8, 2000.
- [Zha-96] Y. Zhang, 'An Artificial Neural Network Approach to Transformer Fault Diagnosis', PHD thesis, Virginia Tech, 1996.
- [Zwi-95] G. Zwielsstein, 'Diagnostic des défaillances théorie et pratique pour les systèmes industriels', édition Hermès, 1995.

Chapitre IV

Application des RNA à la prédiction

- 4.1. Introduction
- 4.2. Application des réseaux supervisés dans la prédiction
 - 4.2.1 Entraînement du réseau RBFG
 - 4.2.1.1 Phase d'apprentissage
 - 4.2.1.2 Phase de prédiction
 - 4.2.2 Réseau RBFG entraîné par la BP
 - 4.2.3 Réseaux RBFG entraînés par la ROM
 - 4.2.3.1 Prédiction du vieillissement thermique de la BORAK22 et du PVC
 - a) En utilisant la technique du FFN pattern
 - b) En utilisant la technique du Batch learning
 - 4.2.3.2 Prédiction de la tenue diélectrique des intervalles d'air pointe-barrière-plan
 - 4.2.4 RBFG entraîné par la ROM couplée avec les K-Means clustering
 - 4.2.4.1 Taux de recouvrement
 - 4.2.4.2 Méthode du classement K-means (K-means clustering)
 - 4.2.4.3 Application du RBFG entraîné par la ROM couplée à la méthode des K-means clustering pour la prédiction du vieillissement thermique
- 4.3 Application des réseaux non-supervisés dans la prédiction
 - 4.3.1 Réseau de neurones SOM et règle d'apprentissage
 - 4.3.2 Phase d'apprentissage du réseau de prédiction de Kohonen
 - 4.3.3 Prédiction du vieillissement thermique du PVC et de la BORAK22
 - 4.3.4 Prédiction de la tenue diélectrique des intervalles d'air pointe-barrière-plan
- 4.4 Discussions
- 4.5 Conclusion
- 4.6 Bibliographie

4.1 INTRODUCTION

Les réseaux de neurones (RN) utilisés dans ce chapitre sont appliqués pour prédire le vieillissement thermique de l'huile de transformateur et du PVC utilisé dans les câbles. Les mêmes RN sont utilisés pour prédire la rigidité diélectrique des intervalles d'air pointe-barrière-plan.

Les expériences du vieillissement thermique du PVC ont été réalisées par M. Nedjar au laboratoire de haute tension de CABEL, la Compagnie Nationale des Câbles (ex. ENICAB) [Nej-91]. Les expériences sur éprouvette ont été effectuées sous différentes températures s'étendant de 80°C à 140°C avec un temps de vieillissement maximum de 5000 heures.

Des essais de vieillissement thermique d'huile de transformateurs utilisée par SONELGAZ et appelée BORAK22, ont été aussi réalisés par A. Hamimi et S. Abdi respectivement pour un temps de vieillissement de 2000 et 5000 heures [Ham-97, Abd-02]. Les essais consistaient à vieillir le matériau et à mesurer les propriétés électriques et mécaniques à des intervalles de temps réguliers.

Les câbles utilisés sont des câbles à moyenne tension (10kV). Les tests ont été faits sur éprouvette de 7.5 cm de diamètre et 2mm d'épaisseur selon les recommandations de la CEI 540. Les mesures du facteur de pertes diélectriques de la BORAK22 ont été effectuées à l'aide d'un pont de Schering selon la norme CEI 250 sous une tension de 2 kV et une distance inter électrodes de 5mm.

Pour le PVC, les tests mécaniques ont été réalisés à l'aide d'un dynamomètre ayant une vitesse de 300mm/min.

Le but de cette partie du travail est de prédire la contrainte du vieillissement thermique pendant de plus longues périodes en utilisant les réseaux de neurones artificiels. Cela va réduire la période expérimentale du vieillissement.

Ayant la base de données de tout l'intervalle du vieillissement thermique, nous avons entraîné nos réseaux pour prédire des caractéristiques d'isolation à de plus longs intervalles de temps de vieillissement. Nous avons entraîné les réseaux dans un intervalle de temps plus petit que l'intervalle expérimental et après nous avons prédit des caractéristiques d'isolation correspondant à tout l'intervalle d'expérimentation. Ainsi, nous pouvons valider nos résultats de prédiction.

Deux catégories de réseaux sont utilisées dans la prédiction : réseaux supervisés et réseaux non-supervisés. Pour chacun des deux réseaux, la prédiction est faite pour les trois cas d'isolations : vieillissement thermique de la BORAK22 et du PVC et la rigidité diélectrique des intervalles d'air.

La prédiction du vieillissement thermique des huiles des transformateurs et des PVC utilisés dans les câbles à moyenne tension est faite pour toutes les températures de vieillissement utilisées.

La prédiction de la tenue diélectrique des intervalles d'air pointe-barrière-plan est faite pour différents paramètres de l'arrangement : la distance pointe-plan, la distance pointe-barrière, la largeur de la barrière et sous différentes tensions: choc de foudre, choc de manœuvres et à fréquence industrielle.

4.2. APPLICATION DES RESEAUX SUPERVISES DANS LA PREDICTION

Les RNA supervisés que nous avons utilisé dans la prédiction sont des réseaux à base radiale de fonction gaussienne (Radial Basis Function Gaussian RBF) entraînés par trois algorithmes : la back-propagation (BP), la méthode d'optimisation aléatoire (ROM) et la ROM couplée avec la méthode des K-means clustering. Deux techniques d'adaptation de poids sont utilisées dans les RBF entraînés par la ROM et la ROM couplée aux K-means clustering : la technique FFN et la technique Batch Learning.

L'avantage principal des réseaux RBF entraînés par la ROM est leur indépendance du modèle mathématique. En outre, l'erreur diminue, évitant une divergence possible. Le réseau est optimisé dans le but d'avoir le même nombre de données d'apprentissage [Bou1-00, Mok2-01, Mok3-01, Mok10-01, Mok11-01, Mok12-01, Mok17-02, Mok23-03, Mok24-03].

Après de nombreux apprentissages, nous avons trouvé que nous pouvons employer un temps d'apprentissage de moins de 2000 heures pour le PVC et de moins de 1000 heures pour la BORAK22 pour prédire des caractéristiques non linéaires obtenues en expériences, à 2000h pour la BORAK22 et à 5000h pour le PVC.

Nous avons aussi entraîné un réseau RBF avec l'algorithme de la Back-propagation (BP) que nous présentons dans la section 4.2.2, mais les difficultés rencontrées dans cet apprentissage nous ont poussé à opter pour l'entraînement avec la ROM.

Les réseaux RBF constituent des modèles efficaces pour la généralisation. L'utilisation de la fonction gaussienne laisse tirer bénéfice de sa caractéristique locale pour faciliter l'apprentissage et pour améliorer la généralisation. L'avantage principal de ces réseaux est qu'il ont toujours une topologie fixe avec une couche cachée et satisfont naturellement aux contraintes suivantes [Khe-94]:

- Capacités universelles d'approximation.
- Capacités de couverture (recouvrement).
- Capacité de généralisation.

Le réseau RBF que nous avons utilisé se compose de trois couches ordonnées : une couche d'entrée avec n unités, une couche cachée avec m unités et une sortie ayant un seul neurone. Chaque neurone dans la couche cachée a une réponse symétrique radiale autour d'un vecteur de paramètres du neurone appelé centre. La couche de sortie est un combinatoire linéaire avec les poids synaptiques de raccordement [Khe-94, Pol-94]. Donnant un ensemble de données d'entrée x_i , $i=1,2,\dots,n$, l'état de l'unité cachée j (ou fonction d'activation) sera dénoté par :

$$\rho_j(x_i) = \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - c_{ij})^2}{\sigma_{ij}^2}\right) \quad (\text{IV.1})$$

c_{ij} : $i = 1 \dots n$ et $j = 1 \dots m$ sont les centres du réseau RBF et σ_{ij} définissent les largeurs des gaussiennes. Ces centres sont des vecteurs de dimension n obtenus à partir des données d'apprentissage.

Dans l'apprentissage avec la ROM, nous avons utilisé deux techniques pour la distribution des centres. La première technique utilisée consiste en un arrangement des centres dans un treillis régulier afin de couvrir uniformément l'espace des données d'entrée. La deuxième technique est une idée originale qui combine entre la distribution des centres à deux dimensions et la méthode des K-means [Mok18-03]. Nous présentons en détail cette dernière technique dans le paragraphe 4.2.4.

Le taux de recouvrement des données est donné par :

$$\tau = \exp \left(- \frac{1}{2} \frac{\left(\frac{\delta_d}{2} \right)^2}{\sigma^2} \right) \quad (IV.2)$$

δ_d est la distance entre deux centres voisins et σ est l'écart type de chaque gaussienne. La réponse normalisée y^* du réseau correspondant aux entrées x_i est donnée par :

$$y^* = \frac{\sum_{j=1}^n w_j \rho_j(x_i)}{\sum_{j=1}^n \rho_j(x_i)} \quad (IV.3)$$

avec w_j , les poids liant les entrées à la couche cachée.

Ce réseau est différent de celui non-normalisé dans lequel le neurone reçoit seulement des données locales pour le calcul de son activation. Mais ici, chaque neurone a ses propres données et des informations sur les autres neurones.

4.2.1 Entraînement du réseau RBFG

4.2.1.1 Phase d'apprentissage

Deux réseaux sont examinés ; le premier est entraîné par un ensemble d'échantillons (y_i, y_{i+1}) , y_i est une valeur d'une propriété correspondant au temps de vieillissement t_i et y_{i+1} est la valeur prédite au temps inconnu t_{i+1} . Cette technique est appelée technique d'apprentissage adaptatif des données (FFN pattern).

Le deuxième réseau est entraîné par des échantillons de la forme (t_i, y_i) . Dans ce deuxième réseau, la technique d'apprentissage adaptatif en bloc 'Batch Learning' est appliquée uniquement pour la prédiction du vieillissement thermique de l'huile de transformateur. Dans ce cas, l'adaptation de poids est faite après le passage de tous les échantillons de l'apprentissage.

En utilisant cette technique, le réseau n'est pas sensible à chaque exemple, contrairement à la technique d'apprentissage adaptatif des données 'FFN Pattern' utilisée dans le premier réseau appliqué au vieillissement thermique de la BORAK22 et du PVC, qui est influencé par le bruit que les entrées peuvent contenir.

4.2.1.2 Phase de prédiction

Pour prédire une valeur future y_{n+1} d'un ensemble de données mesurées y_t , $t=1\dots,n$, l'algorithme sera entraîné sur un ensemble d'échantillons ayant la forme (y_i, y_{i+1}) , $i=1\dots, n-1$. Après cet apprentissage, les poids du réseau sont mis à jour de telle façon à ce que, quand le réseau reçoit la valeur y_i , sa réponse sera y_{i+1} . Maintenant, un nouvel ensemble de données d'entrée est obtenu avec $n+1$ valeurs $(y_t, t=1\dots, n+1)$. L'apprentissage est repris du début afin de prédire la valeur y_{n+2} et le nouvel ensemble de données contiendra $n+2$ valeurs $(y_t, t=1\dots, n+2)$. Ce procédé est répété jusqu'à ce que la prédiction de toutes les valeurs soit obtenue.

Afin d'améliorer la qualité de la prédiction, après chaque prédiction d'une valeur, la première valeur du vecteur de données est omise de l'ensemble des données. Par exemple, la valeur y_1 est omise après la prédiction de y_{n+1} , et y_2 est omise après la prédiction de y_{n+2} etc. De cette façon, le même nombre de valeurs forme le réseau durant tout l'apprentissage.

Applications des RNA dans la prédiction

La même philosophie est appliquée pour prédire des valeurs entraînées sur un ensemble d'échantillons de la forme (t_i, y_i) . L'organigramme global du RBFG entraîné par la ROM est donné à la figure 4.1. Nous présentons par la suite séparément les cas du réseau RBFG entraîné par la BP, la ROM et la ROM couplée aux K-means clustering.

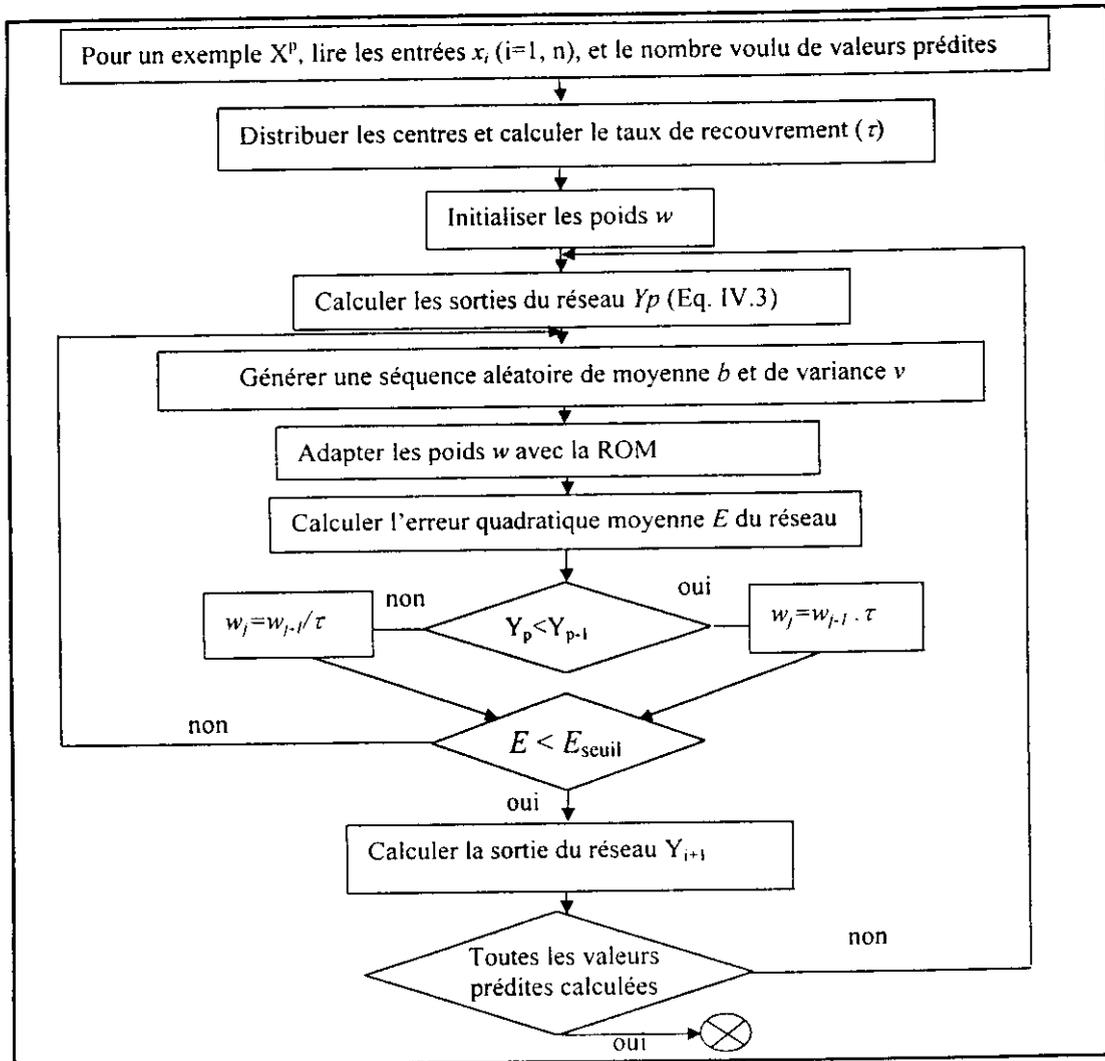


Fig.4.1 Organigramme général de prédiction en utilisant un RBFG

4.2.2 Réseau RBFG entraîné par la BP

Au début de ce travail, un réseau RBFG entraîné par la BP utilisant la méthode de descente du gradient a été employée. Ce réseau exige un choix des taux d'apprentissage pour les ajustements des poids w_j , des centres gaussiens c_{jk} et des rayons des gaussiennes σ_{jk} . Ces ajustements sont l'une des causes des minima locaux en raison de leur dépendance du gradient de l'erreur. L'adaptation de w_j , c_{jk} et σ_{jk} est donnée par :

$$\Delta c_{jk} = -\eta_c \frac{\partial J}{\partial c_{jk}}, \Delta w_j = -\eta_w \frac{\partial J}{\partial w_j} \text{ et } \Delta \sigma_{jk} = -\eta_\sigma \frac{\partial J}{\partial \sigma_{jk}} \quad (\text{IV.4})$$

$J = \frac{1}{2} (y - y^*)^2$ où y est la cible et y^* est la sortie du réseau et η_w , η_c et η_σ sont les taux d'apprentissage pour ajuster respectivement les poids, les centres et les rayons gaussiens. En utilisant la méthode de descente du gradient pour l'ajustement des

paramètres de l'apprentissage, la convergence devient lente et parfois impossible en plus de la difficulté d'optimiser les trois paramètres en même temps. Toutes ces difficultés favorisent l'utilisation de la ROM, qui ne nécessite pas la détermination du gradient. Elle échappe aux minimums locaux [Sol-81].

4.2.2.1 Résultats et discussion

Les résultats du réseau RBFG entraîné par la BP qui est appliquée au vieillissement thermique du PVC sont montrés à la figure 4.2. Le réseau prédit l'allongement à la rupture avec une erreur acceptable mais avec un temps d'exécution plus grand que celui écoulé en exécutant le RBFG entraîné par la ROM (que nous présentons dans la section suivante) et avec un choix problématique des taux d'apprentissage utilisés dans l'ajustement des poids, des centres et des rayons gaussiens. Le calcul du gradient de l'erreur dans cette méthode rend le temps d'entraînement long. Si l'ensemble des entrées est grand, le temps d'entraînement sera problématique aussi. Les résultats présentés à la figure 4.2, obtenus pour la température de vieillissement de 120°C, illustrent bien les performances du RBFG entraîné par la BP [Mok23-03, Mok11-01, Bou1-00, Mok25-03].

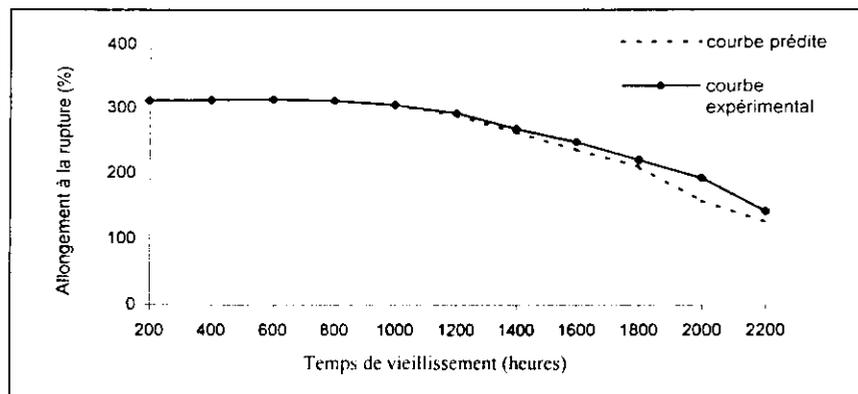


Fig.4.2 Prédiction de l'allongement à la rupture du PVC utilisant la BP
 $\eta_w=0.18$, $\eta_c=0.08$, $\eta_\sigma=0.05$.

4.2.3 Réseaux RBFG entraînés par la ROM

La méthode aléatoire d'optimisation ROM est basée sur la technique de recherche aléatoire. Elle a été employée avec succès dans plusieurs problèmes d'optimisation. Elle est plus appropriée dans certains cas où la fonction erreur a plusieurs minimums. La ROM converge à un minimum global avec une probabilité égale à 1 dans un ensemble compact. L'idée de cet algorithme est d'attribuer aux poids du réseau un ensemble de séquences de bruit blanc et puis de calculer la sortie du réseau. Si l'erreur actuelle est moins que l'erreur précédente, les nouveaux poids sont gardés ; sinon, les poids précédents sont gardés jusqu'à ce que l'erreur désirée soit obtenue.

L'apprentissage de ce réseau est identique à celui donné dans le RBFG entraîné par la BP. La différence est dans l'ajustement des paramètres d'apprentissage. En outre, l'algorithme de la ROM n'a pas besoin d'ajuster les centres c_{jk} et les rayons des gaussiennes σ_{jk} parce qu'il est indépendant du critère à minimiser.

La ROM développée par Solis et Wetts [Sol-81] dépend de la variation des vecteurs gaussiens, c.-à-d. une séquence avec une grande variation génère un grand changement de l'erreur, mais sans aucune oscillation. Cependant, une grande variation donne parfois une stagnation de la surface d'erreur [Pol-94]. Dans ce travail, l'apprentissage est lancé

par une grande valeur de la variance qui est réduite progressivement. Cette méthode est basée sur la génération des séquences aléatoires, afin de déplacer les valeurs des poids dans l'espace à la recherche de l'optimum et elle permet d'échapper aux minima locaux [Sol-81]. Son organigramme est donné à la figure 4.3.

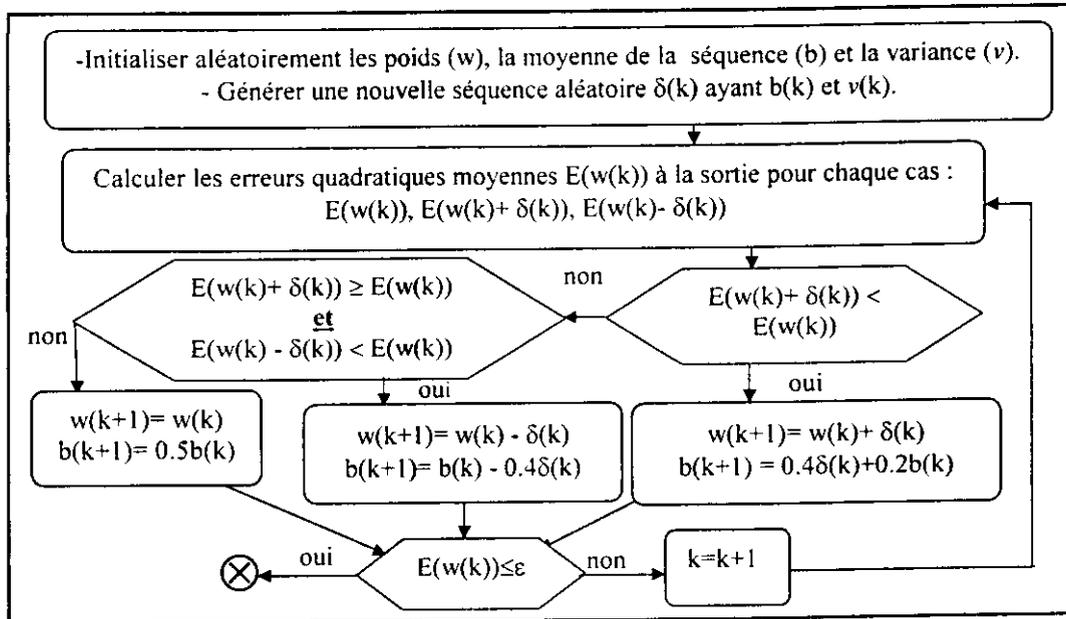


Fig.4.3 Algorithme de la ROM

4.2.3.1 Prédiction du vieillissement thermique de la BORAK22 et du PVC

Pour toutes les figures présentées dans ce chapitre, les lignes discontinues et continues représentent respectivement les courbes prédites et les courbes expérimentales.

a) En utilisant la technique du FFN pattern

Nous présentons aux figures 4.4 et 4.5 respectivement les valeurs mesurées et prédites de l'allongement à la rupture et de la résistance à la traction en fonction du temps de vieillissement du PVC sous différentes températures. Différents paramètres d'apprentissage de chaque propriété du PVC sont donnés dans les tableaux 4.1 et 4.2.

Pour la température de vieillissement $T = 80^{\circ}\text{C}$, la prédiction des deux propriétés est faite au-delà de 7000 heures. Comme cette température est la plus basse, le matériau pourrait être vieilli sur une plus longue période. Quand l'intervalle expérimental est grand, la période de prédiction augmente.

Pour les deux autres températures (100°C et 120°C), le matériau vieillit rapidement et de ce fait, les intervalles de mesures sont réduits. Il en est aussi ainsi pour les intervalles prédits. Les résultats présentés aux figures 4.4 et 4.5 sont obtenus en deux phases : phase d'apprentissage et phase de généralisation. Dans la première phase, le temps d'apprentissage pour la température de vieillissement de 100°C par exemple est de 1600h. En utilisant cet intervalle, qui est plus petit que l'intervalle expérimental, le réseau a pu prédire d'une façon satisfaisante la propriété jusqu'à la totalité de l'intervalle expérimental. Ceci nous permettra de valider notre prédiction.

Dans la deuxième phase, nous entraînons le réseau sur tout l'intervalle expérimental pour prédire les propriétés à des temps plus longs que les intervalles expérimentaux.

T (C°)	Nombre de valeurs mesurées	Pas d'apprentissage (h)	Période de mesure (h)	Nombre de valeurs prédites	Période de prédiction (h)
80	12	400	5000	7	2800
100	9	200	2600	8	1600
120	5	100	1300	8	800

Tab.4.1 Paramètres d'apprentissage pour l'allongement à la rupture

T (C°)	Nombre de valeurs mesurées	Pas d'apprentissage (h)	Période de mesure (h)	Nombre de valeurs prédites	Période de prédiction (h)
80	12	400	4800	7	2800
100	10	200	2600	8	1600
120	5	100	1300	8	800

Tab.4.2 Paramètres d'apprentissage pour la résistance à la traction

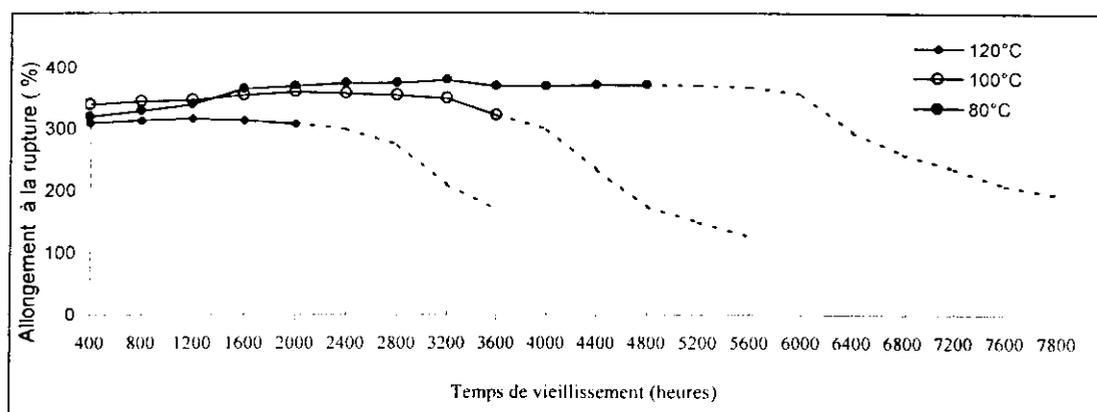


Fig.4.4 Prédiction de l'allongement à la rupture du PVC en fonction du temps de vieillissement

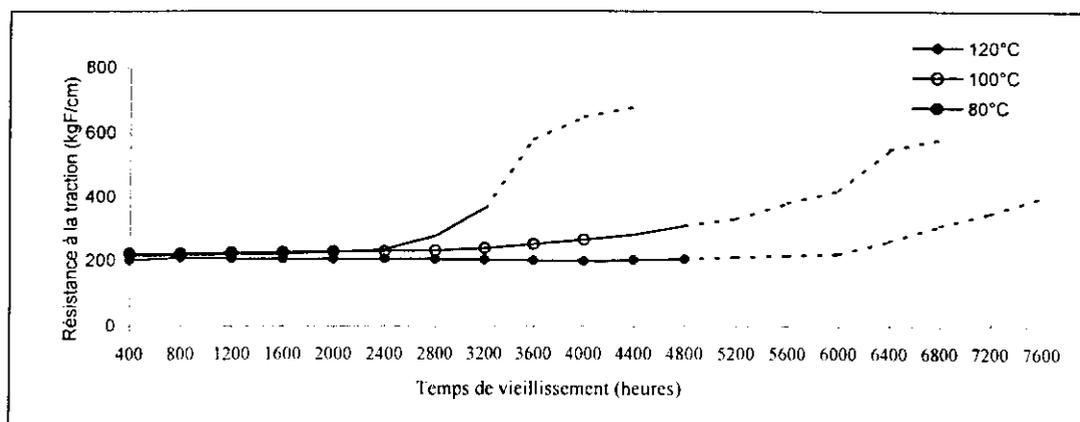


Fig.4.5 Prédiction de la résistance à la traction du PVC en fonction du temps de vieillissement

A la figure 4.6, le réseau de prédiction est appliqué pour atteindre le critère de la durée de vie pour une réduction de 50% de la propriété de l'allongement à la rupture. Nous donnons la prédiction de la durée de vie à 20.000 heures dans le but d'obtenir l'indice de température du polymère. Les valeurs mesurées et prédites de la perte de masse sur la surface du PVC en fonction du temps de vieillissement sont données à la figure 4.7 sous une température de 100°C [Bou1-00].

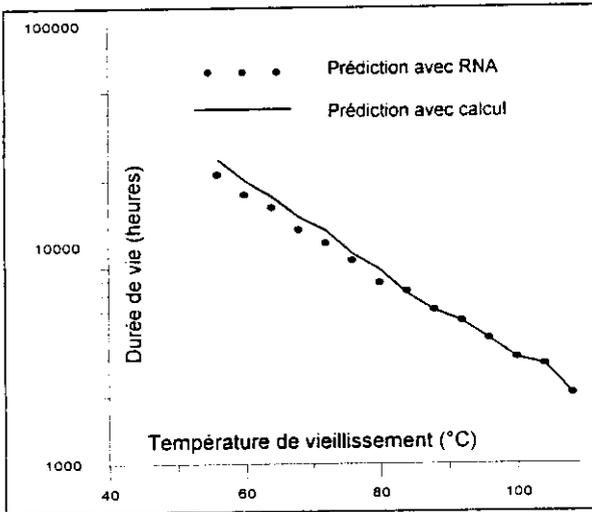


Fig.4.6 Endurance Thermique du PVC correspondant à l'allongement à la rupture

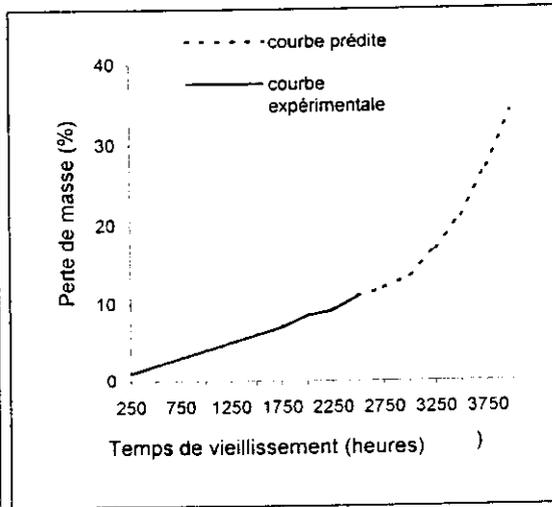
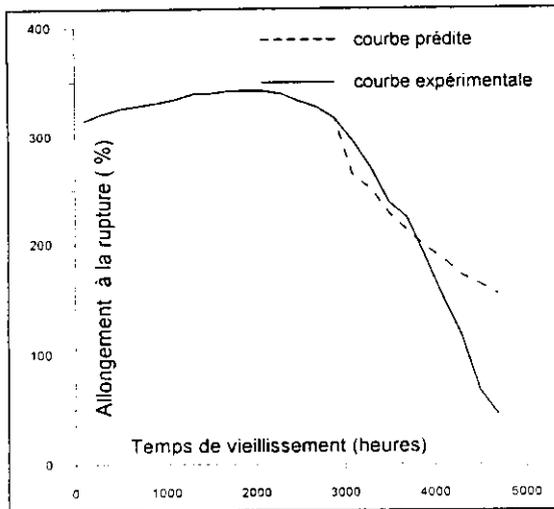
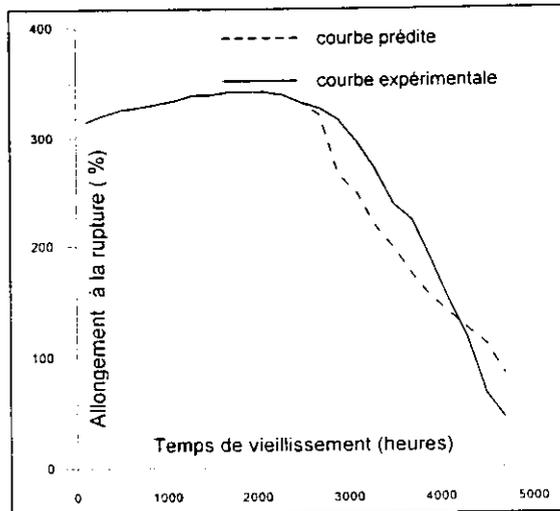


Fig.4.7 Perte de masse sur la surface du PVC à T= 100°C.

Nous avons employé deux autres temps d'apprentissage pour prédire la caractéristique de l'allongement à la rupture du PVC (2700h, 3200 à T= 100°C) pour le même pas d'apprentissage utilisé à la figure 4.4 et nous avons remarqué que le temps d'apprentissage a une grande influence sur l'apprentissage du réseau (Fig.4.8 (a-b)).



(a) Temps d'apprentissage = 3200h



(b) Temps d'apprentissage = 2700h

Fig.4.8 Variation de l'allongement à la rupture du PVC en fonction du temps de vieillissement à T = 100°C

A la figure 4.9, nous donnons la variation du facteur de pertes diélectriques de la BORAK22 sous une température de vieillissement de 120°C, en fonction du temps de vieillissement en utilisant le réseau FFN. Nous avons utilisé un pas de 150h et deux temps d'apprentissage (750h et 900h) pour les figures 4.9 (a-b) et un pas de 200h et un temps d'apprentissage de 1000h pour la figure 4.9(c).

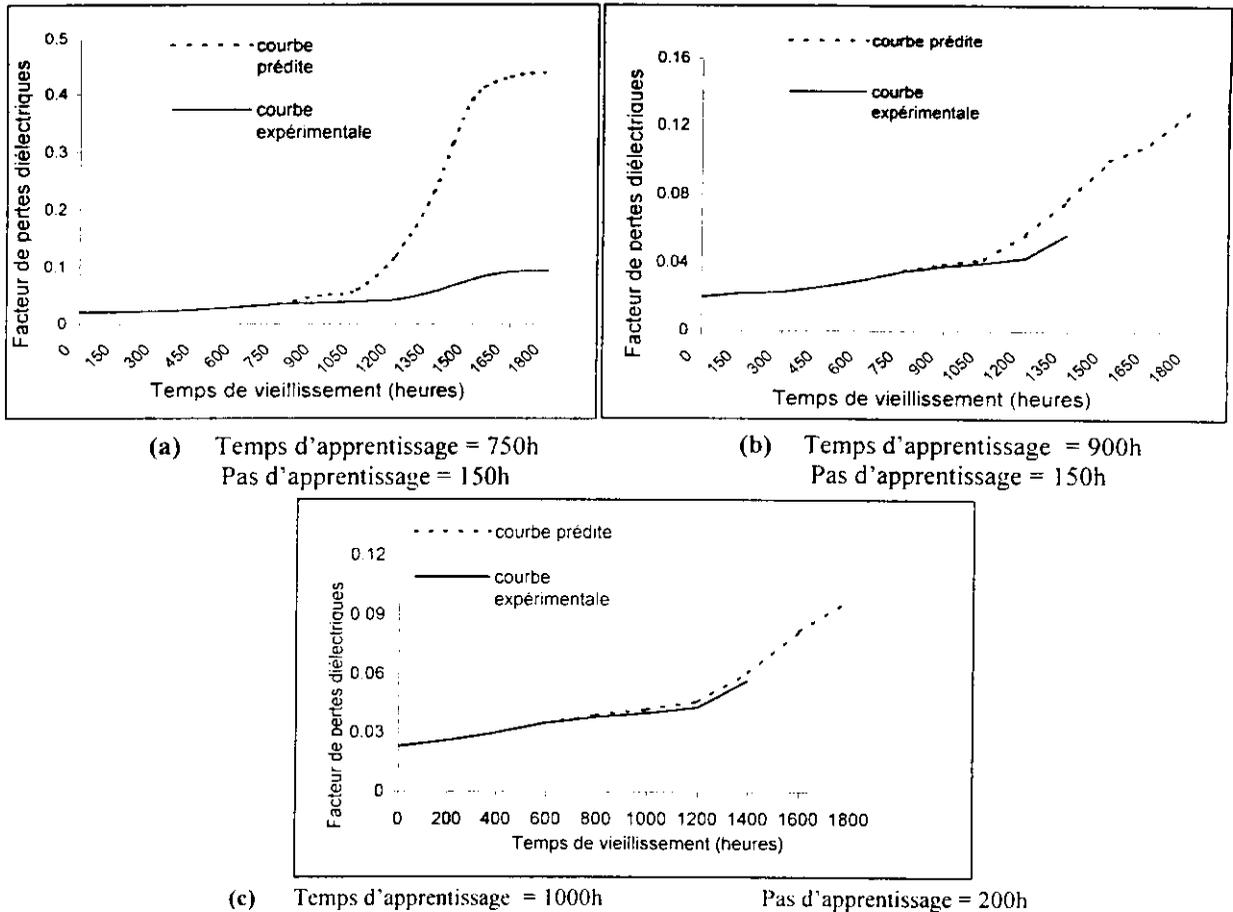


Fig.4.9 Prédiction du facteur de pertes diélectriques de la BORAK22 pour $T = 120^{\circ}\text{C}$ utilisant la technique FFN pattern

Utilisant un temps d'apprentissage de 900h (Fig.4.9-b), la prédiction est meilleure que dans le cas de 750h (Fig.4.9-a) et la prédiction pour un temps d'apprentissage de 1000h et un pas d'apprentissage de 200h est meilleure que celle donnée pour le temps d'apprentissage de 900h.

D'après les résultats de la figure 4.9, il paraît que la prédiction est meilleure pour les temps d'apprentissage les plus longs puisque dans ce cas, le réseau a plus d'informations sur le comportement des entrées. Mais le dilemme de la précision-généralisation se posera quand le vecteur entrée devient volumineux.

b) En utilisant la technique du Batch learning

La technique du Batch learning (BL) est testée uniquement pour la prédiction du vieillissement thermique de la BORAK22 sous une température de 120°C . A la figure 4.10, nous présentons la prédiction du facteur de pertes diélectriques en utilisant deux temps d'apprentissage de 750h (Fig.4.10-a) et de 900h (Fig.4.10-b) avec un pas de 150h.

Une comparaison entre les résultats des figures 4.9 et 4.10 conduit à conclure que la technique BL a amélioré clairement les capacités d'apprentissage du réseau. Avec cette technique, l'apprentissage n'est pas sensible au bruit que peut contenir la base de données car l'ajustement des poids est fait après le passage de tous les exemples d'apprentissage, et ainsi la réadaptation est 'prudente' [Mok23-03]. Contrairement à la technique FFN pattern, le réseau adapte ses poids avec le passage de chaque exemple.

Applications des RNA dans la prédiction

Nous avons employé deux autres pas d'apprentissage de 200h et 250h [Mok4-02] (Fig.4.11 (a-b)) et nous avons remarqué que le temps et le pas d'apprentissage ont une grande influence sur l'apprentissage du réseau, mais il n'y a aucune dépendance linéaire entre eux. Nous avons noté qu'il y a un temps d'apprentissage optimal et un pas optimal donnant les meilleurs résultats mais pas tous les deux ensemble dans le même apprentissage.

Il s'avère qu'un temps d'apprentissage plus grand donne la meilleure prédiction mais d'après les résultats présentés dans cette section, nous avons noté que l'erreur a des minima locaux en fonction du pas d'apprentissage (Fig.4.12) qui dépend à son tour du temps d'apprentissage.

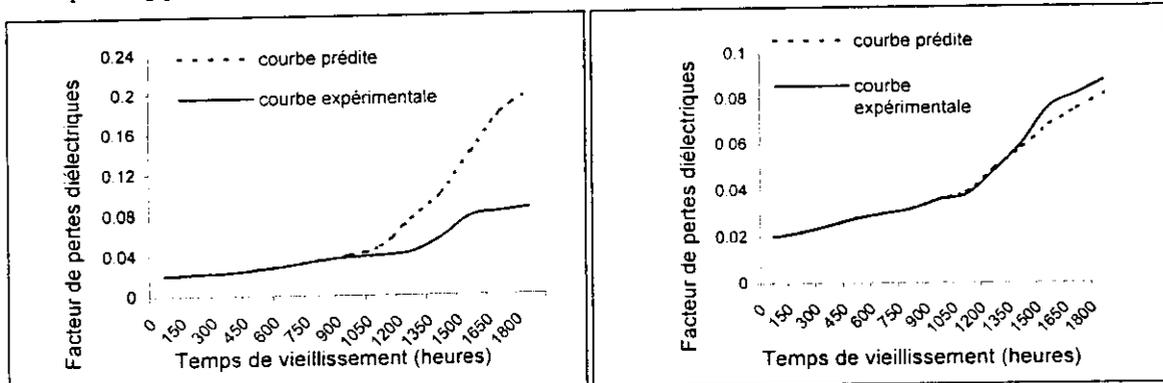


Fig.4.10 Prédiction du facteur de pertes diélectriques de la BORAK22 utilisant le batch learning $T = 120^{\circ}\text{C}$

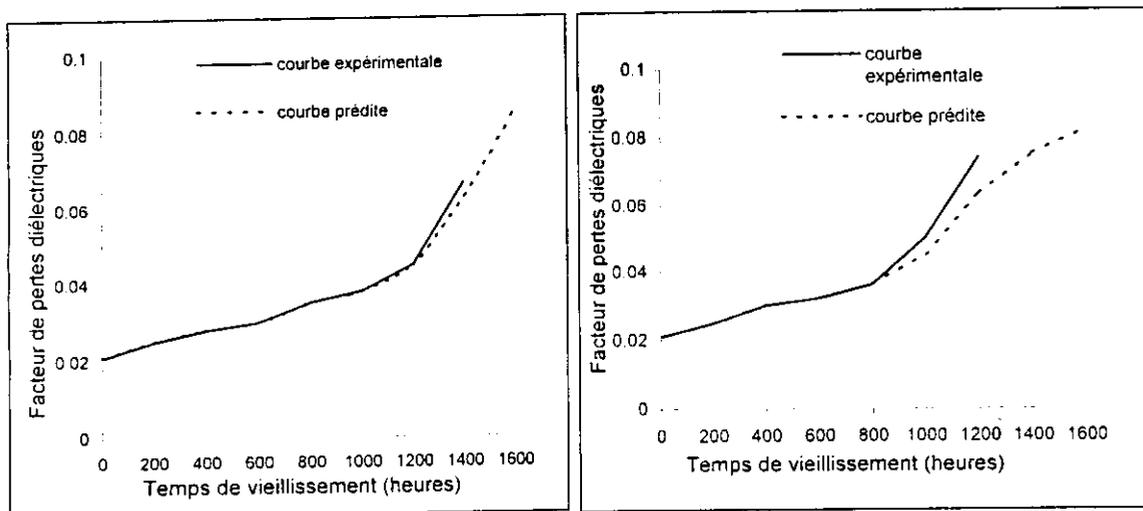


Fig.4.11 Prédiction du facteur de pertes diélectriques de la BORAK22 à $T = 120^{\circ}\text{C}$

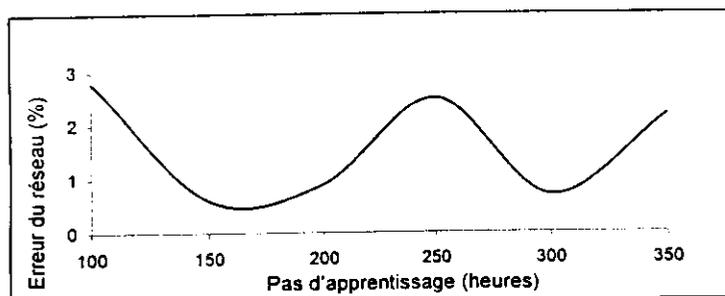


Fig.4.12 Influence du pas d'apprentissage sur l'erreur du réseau pour un temps d'apprentissage de 1000h

Nous avons aussi déterminé la prédiction du facteur de pertes diélectriques de la BORAK22 jusqu'à 3000h [Mok3-01, Mok11-01] pour un temps d'apprentissage de 1700h et un pas d'apprentissage de 170h (Fig.4.13). Cette prédiction est faite pour une huile vieillie jusqu'à 2000h. Dans l'apprentissage, nous avons utilisé un temps d'apprentissage de 1700h de vieillissement et nous avons prédit la propriété pour les 1300h qui suivent. Les valeurs expérimentales sont en bonne corrélation avec les valeurs prédites. Les réseaux entraînés par la ROM ayant utilisé les techniques FFN et Batch learning présentent respectivement des erreurs de 5% et 3%.

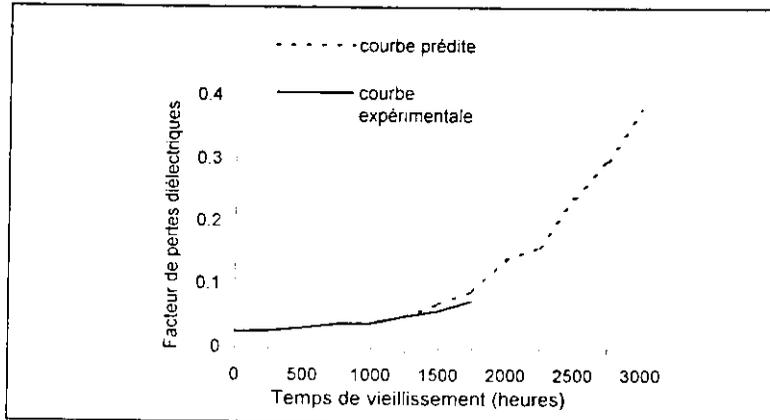


Fig.4.13 Prédiction jusqu'à 3000h du facteur de pertes diélectriques de la BORAK22 à $T = 120^{\circ}\text{C}$
Temps d'apprentissage = 1700h Pas d'apprentissage = 170h.

Comparaison entre les réseaux RBFG entraînés par la BP et par la ROM

Nous avons comparé les résultats expérimentaux et ceux de la prédiction concernant l'allongement à la rupture du PVC vieilli à 120°C , utilisant la BP et la ROM avec les deux méthodes FFN et BL. Le temps d'apprentissage utilisé est de 900h et le pas d'apprentissage est de 250h pour les deux méthodes (Fig.4.14).

Nous déduisons que la BP et la ROM sont tous les deux des réseaux de neurones bons pour la prédiction et présentent tous les deux des erreurs acceptables. Cependant, la ROM présente la meilleure prédiction. En plus, l'utilisation du réseau BP présente beaucoup de difficultés durant la phase d'apprentissage et prend un temps plus long pour ajuster les paramètres d'apprentissage du réseau [Mok6-02, Mok25-03].

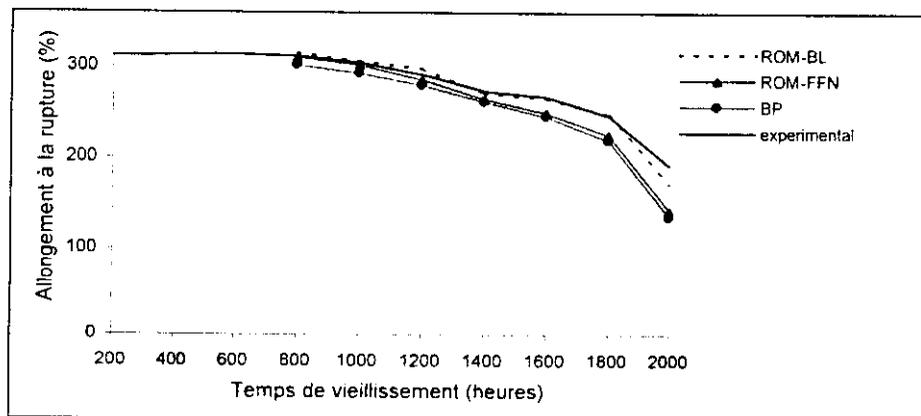


Fig.4.14 Comparaison entre les réseaux RBFG entraînés par la BP et par la ROM pour la prédiction de l'allongement à la rupture du PVC à $T = 120^{\circ}\text{C}$.
Temps d'apprentissage = 900h, Pas d'apprentissage = 200h

4.2.3.2 Prédiction de la tenue diélectrique des intervalles d'air pointe-barrière-plan

Nous présentons les résultats de prédiction de la tension disruptive à 50% d'un intervalle d'air utilisant un réseau de neurones RBFN entraîné par la ROM en fonction de la longueur de l'intervalle d'air pour plusieurs positions de la barrière sous la tension de choc de foudre et de choc de manœuvres.

Pour cette application et après avoir testé les deux méthodes FFN et BL, nous avons choisi d'utiliser la technique FFN grâce aux meilleurs résultats qu'elle présente pour cette application. L'apprentissage se fait dans cette application de la même manière présentée pour la prédiction du vieillissement thermique du PVC et de la BORAK22.

Les données expérimentales utilisées dans cette partie de notre travail sont obtenues par A.Boubakeur [Bou8-79] pour différentes positions de la barrière ($a/d = 0.2, 0.4, 0.6$ et 0.8 avec a la distance pointe-barrière) et pour différentes formes de tension. La barrière utilisée a une permittivité relative égale à 5.8 et une largeur de $1m$.

Nous donnons à la figure 4.15 la prédiction de la tension de rupture de l'intervalle d'air pointe-plan (sans barrière) pour les longues distances et aux figures 4.16 et 4.17, la prédiction de la tension de rupture à 50% respectivement au choc de foudre et au choc de manœuvres pour un intervalle d'air pointe-barrière-plan avec différentes distances pointe-plan (d). La distance d'apprentissage pour les trois figures est respectivement $200cm, 150cm$ et $150cm$.

Le RNA utilisé a un nombre d'entrées dépendant du nombre de valeurs des distances pointe-plan (d_i) utilisées et a une sortie Y donnant la valeur prédite correspondant à la valeur d_{i+1} . Nous obtenons une bonne concordance entre les caractéristiques prédites et les caractéristiques expérimentales de la tension disruptive.

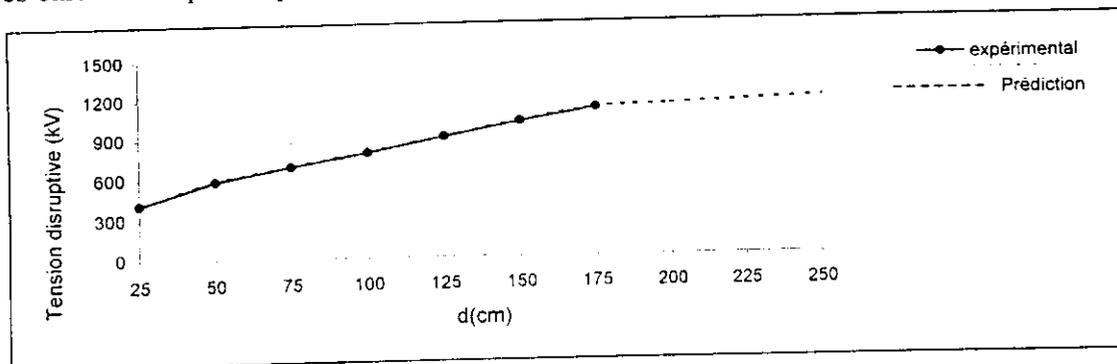


Fig.4.15 Prédiction de la tension disruptive au choc de manœuvres ($170/2500\mu s$) d'un intervalle d'air pointe-plan (sans barrière) en fonction de la longueur de l'intervalle d'air en utilisant la FFN.

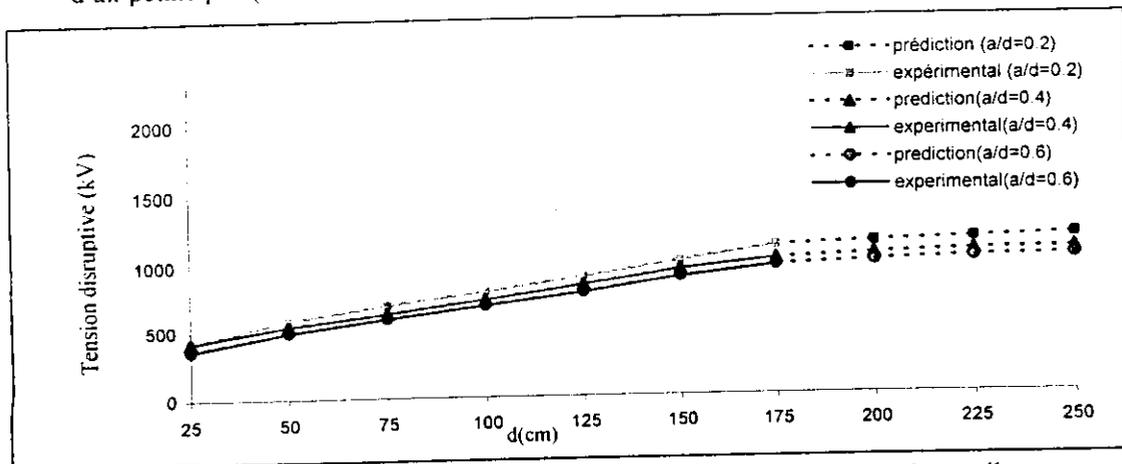


Fig.4.16 Prédiction de la tension disruptive au choc de foudre ($1,2/50\mu s$) d'un intervalle d'air pointe-barrière-plan pour différentes positions de la barrière en utilisant la FFN

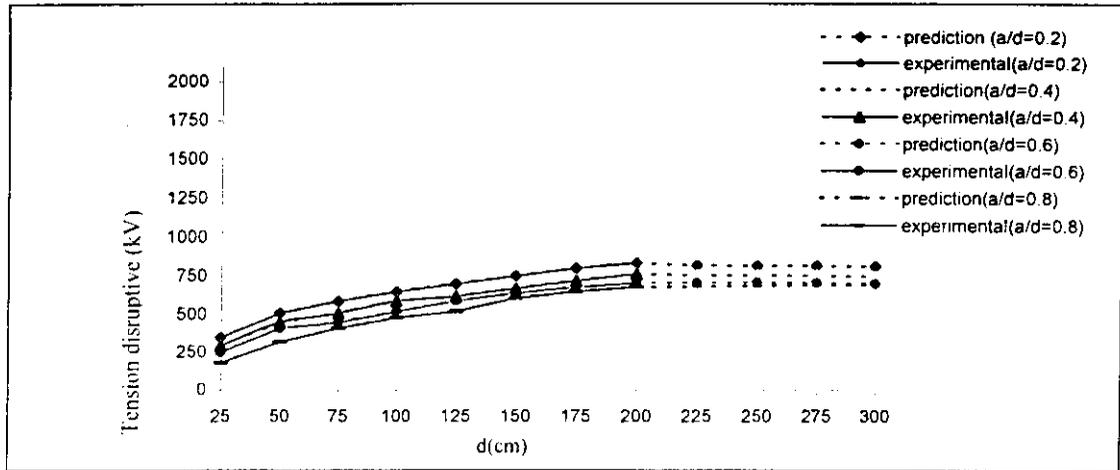


Fig.4.17 Prédiction de la tension disruptive aux chocs de manœuvre (170/2500 μ s) d'un intervalle d'air pointe-barrière-plan pour différentes positions de la barrière en utilisant la FFN

La prédiction de la tension de rupture à 50% pour plusieurs largeurs de la barrière en fonction de la longueur de l'intervalle d'air est présentée à la figure 4.18. La distance d'apprentissage est de 60cm. Les valeurs expérimentales et prédites sont comparées sur la même figure.

Les résultats présentés aux figures 4-15 à 4-18 sont confirmés aussi par ceux obtenus dans le travail effectué sur la modélisation avec la méthode des éléments finis des systèmes pointe-barrière-plan sous une tension positive [Mok1-01, Mok26-97, Mok27-97, Mok28-98, Bou10-98, Bou11-99, Mok29-00, Mok30-00, Mok31-03].

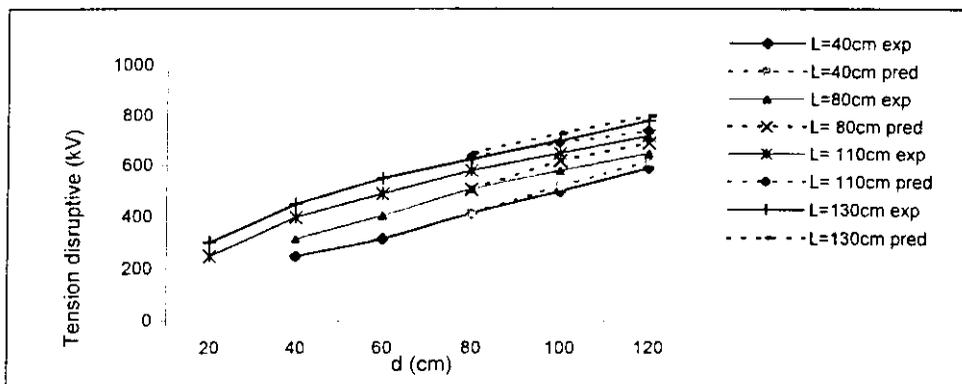


Fig.4.18 Prédiction de la tension disruptive à fréquence industrielle d'un intervalle d'air pointe-barrière-plan pour différentes largeurs de la barrière en utilisant la FFN

4.2.4 RBFG entraîné par la ROM couplée avec les K-Means clustering

Cette méthode est une idée originale. Son implémentation dans l'algorithme RBFG entraîné par la ROM se fait seulement dans la phase de répartition des centres dans l'espace des données d'entrée.

Dans ce réseau, deux idées principales sont ajoutées au réseau RBFG entraîné par la ROM présentée dans la section précédente. La première idée est que les centres partagent les données d'apprentissage en classes, en minimisant la distance euclidienne quadratique entre les points d'apprentissage et le centre d'une classe donnée. Cette partie de l'algorithme représente un apprentissage non supervisé pour répartir les centres. La deuxième idée est que la distribution des centres se fait en deux dimensions.

4.1.3.1 Taux de recouvrement

L'idée principale d'un réseau RBF était introduite par Powell [Pol-94], où il déclare, comme nous l'avons présenté au chapitre 1, que n'importe quelle fonction $f(x)$ peut être approximée par une interpolation composée de la somme de p fonctions noyaux par :

$$f(x) = \sum_{i=1}^p w_i \Phi(|x - c_i|) \quad (IV.5)$$

c_i sont les nœuds d'interpolation $i = 1-n$, appelés centres.

w_j sont les poids synaptiques liant les neurones à la sortie et Φ est une fonction noyau.

Si le réseau contient deux entrées, la couverture du treillis donnée par l'équation (IV.2) assure les mêmes rayons pour toutes les gaussiennes. Mais le nombre de neurones cachés est égal au carré du cardinal de C , où C est l'ensemble des centres gaussiens (Fig.4.19) [Khe-94]. La figure 4.20 est donnée pour une entrée (1 dimension).

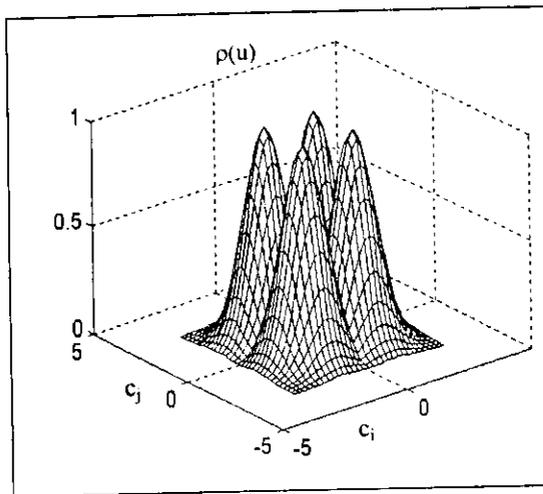


Fig.4.19 Couverture de l'espace d'entrées en deux dimensions

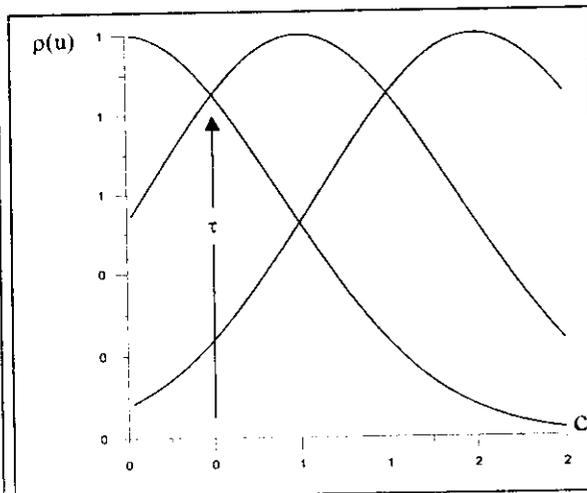


Fig.4.20 Couverture de l'espace d'entrées en une dimension

4.2.4.2 Méthode du classement K-means (K-means clustering)

Cette méthode définit un ensemble de centres groupés (classés) et partage des données d'apprentissage dans des sous-ensembles. Chaque centre d'un groupe est associé à l'un des neurones cachés (M). L'algorithme du K-means clustering trouve un minimum local dans les distances euclidiennes quadratiques totales $E_{K\text{-means}}$ entre les points d'apprentissage assignés à chaque classe et le centre c_m de cette classe :

$$E_{K\text{-means}} = \sum_{m=1}^M \sum_{k=1}^K B_{mk} \|x_k - c_m\|^2 \quad (IV.6)$$

B_{mk} est le facteur d'appartenance de classe et K est le nombre de classes.

L'implémentation de cette méthode dans l'algorithme RBFG entraîné par la ROM se fait seulement dans la partie de répartition des centres dans l'espace des données d'entrée comme suit :

- Initialiser avec des valeurs aléatoires les centres c_i ($1 \leq i \leq M$).
- Calculer les distances euclidiennes et trouver la distance minimale.
- Réadapter les centres comme suit :

$$\left. \begin{aligned} c_k(n) &= c_k(n-1) + E_{K\text{-means}} && \text{pour } i \neq k \\ c_k(n) &= c_k(n-1) && \text{pour } i = k \end{aligned} \right\} \quad (IV.7)$$

Dans la méthode de distribution de centres proposée dans cette section, l'équation (IV.2) est appliquée pour un calcul de la distance entre deux centres voisins δ_d en deux dimensions (y, t) (y est la caractéristique à prédire et t est le temps de vieillissement) multiplié par le facteur B_{mk} ($0 < B_{mk} < 1$). Ce dernier change d'un ensemble d'entrées à un autre si le nouvel ensemble d'entrées représente une autre classe d'entrées. Les largeurs des gaussiennes σ sont calculées par conséquent en deux dimensions.

De l'équation (IV.5), nous constatons que toutes les fonctions RBF indiquent un changement des contributions dans l'approximation pendant qu'elles s'éloignent radialement du centre.

Le réseau RBF construit efficacement un espace de fonctions linéaires, qui dépend de la position des points de données connus selon une mesure arbitraire de distance. Alternativement, il peut être vu comme une méthode pour performer une reconstruction d'hypersurface. Cette approche a été largement acceptée pour mener à une bonne approximation multivariable pour laquelle d'autres techniques échouent [Khe-94].

La taille du RBF indiqué dans l'équation (IV.5) dépend du nombre de points de données. Dans beaucoup de situations pratiques, ceci peut ne pas être faisable pour une grande base de données. D'ailleurs, ces données sont souvent perturbées par un bruit. Ainsi, l'entraînement d'un réseau pour interpoler strictement, mènera à une situation où ces points produiront un complexe, sinon un réseau erroné.

À cet effet, une approximation étroite est proposée par Broomhead et Lowe [Broo-88] aux points d'interpolation. Ceci donne naissance à un réseau RBF avec des points de données ($x_i, i=1, n$) et des centres ($c_j, j=1, m$) où $m \ll n$.

Le choix des centres est crucial pour les performances du réseau RBF. Il doit être choisi pour échantillonner le domaine d'entrées.

4.2.4.3 Application du RBFG entraîné par la ROM couplée à la méthode des K-means clustering pour la prédiction du vieillissement thermique

Dans cette section, nous avons utilisé la méthode du Batch learning. Les résultats que nous présentons aux figures 4.21 et 4.22 donnent la prédiction de l'allongement à la rupture du PVC respectivement pour deux températures de vieillissement de 80°C et 100°C en utilisant un temps d'apprentissage de 1600h et un pas d'apprentissage de 150h.

Le principe d'apprentissage de ce réseau RBFG reste le même comme pour le cas d'entraînement avec la BP et la ROM à une différence dans la distribution des centres gaussiens qui se fait selon les classes représentant les entrées et à deux dimensions au lieu d'une dimension.

Dans cette méthode de distribution des centres à deux dimensions, une entrée est employée (Y) mais dans le calcul de la distance entre les centres voisins δ_d , le temps est associé directement avec la propriété de l'isolant comme si deux entrées sont introduites. Le coefficient B_{mk} change quand le vecteur d'entrée représente une autre classe de données d'entrée. C'est la cause pour laquelle nous l'avons appelé le facteur d'appartenance à une classe.

La disposition des centres à deux dimensions (représentée sur les figures par RBFG2dim) a nettement amélioré la prédiction en comparant avec les résultats d'un

Applications des RNA dans la prédiction

treillis régulier à une dimension. La prédiction de l'allongement à la rupture est faite jusqu'à 3500h pour un temps d'apprentissage de 1200h et un pas de 200h.

Pour la prédiction de la propriété de la BORAK22, nous avons utilisé dans cette section une base de données qui correspond à 5000h de vieillissement [Abd-02].

Nous présentons alors la prédiction du facteur de pertes diélectriques de la BORAK22 en fonction du temps de vieillissement utilisant un pas d'apprentissage de 500h et un temps d'apprentissage de 4000h à la figure 4.23. La prédiction est faite pour 6000h [Mok18-03].

Nous avons noté qu'avec une disposition des centres à 1 dimension, la prédiction suit les valeurs expérimentales réelles au début de la prédiction et s'éloigne en avançant dans la prédiction. Le deuxième réseau présente la meilleure prédiction [Mok18-03].

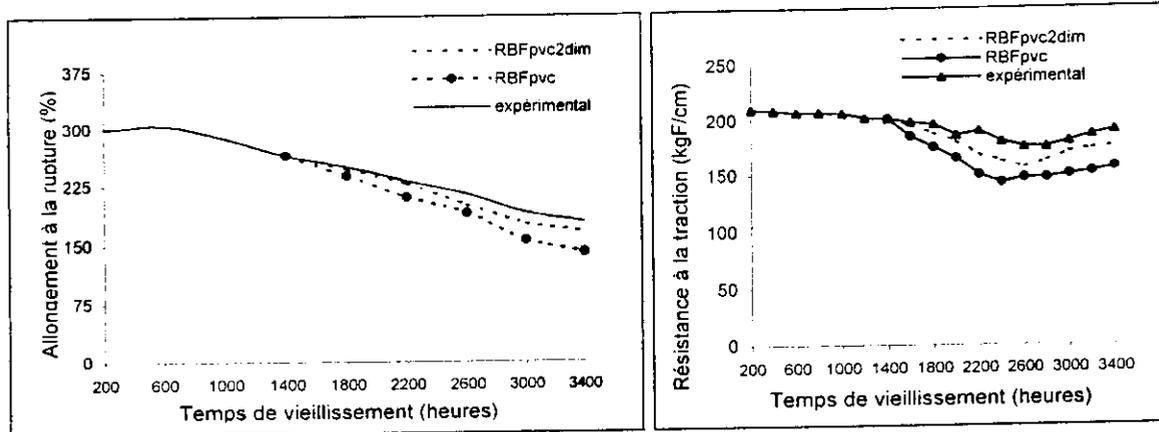


Fig.4.21 Prédiction de l'allongement à la rupture du PVC à $T = 80^{\circ}\text{C}$ utilisant un RBF entrainé par la ROM couplée avec les K-means

Fig.4.22 Prédiction de la résistance à la traction du PVC à $T = 100^{\circ}\text{C}$ utilisant un RBF entrainé par la ROM couplée avec les K-means

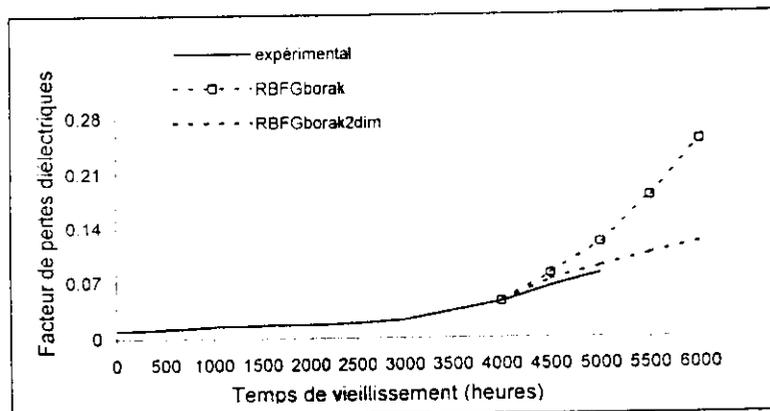


Fig.4.23 Prédiction du facteur de pertes diélectriques de la BORAK22 à $T = 120^{\circ}\text{C}$ utilisant un RBF entrainé par la ROM couplée avec les K-means

4.3 APPLICATION DES RESEAUX NON-SUPERVISES DANS LA PREDICTION

4.3.1 Réseau de neurones SOM et règle d'apprentissage

Le réseau de Kohonen (Self-Organization Maps -SOM), souvent appelé la carte de Kohonen, se compose d'une seule couche de neurones fonctionnant en mode compétitif auto-organisé [Koh3-95]. Chaque neurone est alimenté par les n composantes du vecteur d'entrée, par les coefficients de pondération, formant un vecteur w_k à dimension N .

L'algorithme de l'auto-organisation (self-organization) est constitué par les séquences des opérations suivantes :

- Présenter le vecteur d'entrée e au réseau.
- Trouver la région dans le réseau où un neurone spécifique répond le plus fortement au vecteur e précédemment présenté. L'unité gagnante N_w est celle dont le vecteur de poids est le plus proche dans le sens d'une mesure de distance par rapport au vecteur d'entrée.
- Adapter les poids des neurones choisis de cette région. Répétant ces séquences plusieurs fois (selon les données d'entrée, la taille du réseau et les facteurs d'adaptation) mène le réseau à un état organisé, dans lequel chaque neurone représente une classe séparée des données. Dans l'algorithme généralisé de Kohonen, l'adaptation des poids des neurones trouvés dans le voisinage autour du neurone gagnant N_w , se fait selon la règle suivante :

$$w(i+1)=w(i)+\alpha_i G(i,e)[e-w_i] \quad (IV.8)$$

e_i est le vecteur des poids du $i^{\text{ème}}$ neurone trouvé dans le voisinage du neurone gagnant et α_i est le taux d'apprentissage qui diminue avec le temps. Habituellement, c'est la diminution linéaire qui est utilisée, à partir d'une certaine valeur initiale et jusqu'à $\alpha_i = 0$.

Le voisinage du neurone gagnant diminue aussi avec le temps. L'une des méthodes les plus utilisées dans l'ajustement du voisinage est l'application de la fonction gaussienne définie autour du gagnant et décrite par la relation :

$$G(i,e)=\exp\left(-\frac{d_i^2(i,w)}{2.\lambda^2}\right) \quad (IV.9)$$

$d_i(i,w)$ est la distance euclidienne entre le gagnant et le $i^{\text{ème}}$ neurone et λ est une constante.

Dans cette application, nous avons employé la règle d'apprentissage de Kohonen, pour l'ajustement des poids en utilisant la règle citée ci-dessus dans l'équation (IV.8).

Supposons que le $i^{\text{ème}}$ neurone gagne, la $i^{\text{ème}}$ rangée de la matrice de poids d'entrée est ajustée comme suit :

$$w(i)=w(i-1)+\alpha_i.(e(i)-w(i-1)) \quad (IV.10)$$

$e(i)$ est le vecteur d'entrée correspondant.

La règle de Kohonen permet aux poids d'un neurone d'apprendre à reconnaître un vecteur d'entrée. C'est pour cette raison qu'il est utile dans la classification [Oso-97].

Le neurone dont le vecteur de poids était le plus proche du vecteur d'entrée est adapté pour être encore plus proche. Le résultat est que le neurone gagnant va encore gagner la compétition la prochaine fois qu'un vecteur semblable est présenté alors que la probabilité est moindre pour lui de gagner quand un vecteur très différent est présenté.

Par la suite, s'il y a assez de neurones, chaque classe de vecteurs d'entrée semblables aura un neurone qui donne une sortie égale à "1" quand un vecteur dans cette classe est

présenté, tout en produisant un "0" à toutes autres fois. Ainsi, le réseau compétitif apprend à classer les vecteurs d'entrée qu'il voit [Mok16-02]. L'algorithme général des SOM pour la classification est donné à la figure 4.24.

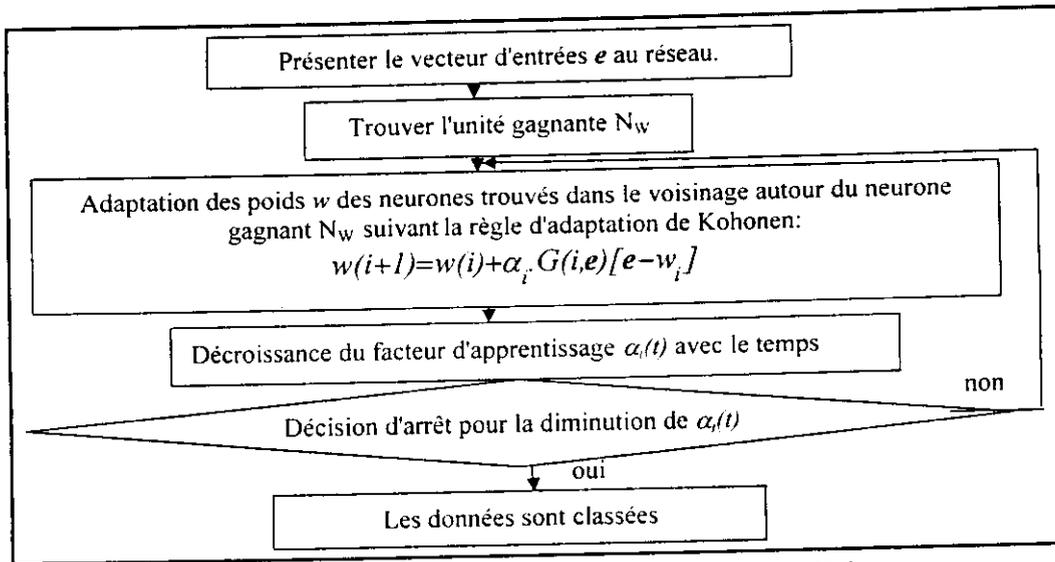


Fig.4.24 Algorithme du classement des données utilisant la carte de Kohonen

4.3.2 Phase d'apprentissage du réseau de prédiction de Kohonen

Dans cette méthode, pour prédire une caractéristique, nous utilisons dans l'apprentissage deux autres caractéristiques présentées avec celle à prédire. Pour rendre la prédiction indépendante de la nature de la caractéristique d'isolation à prédire, nous transformons la base de données d'entrée en faisant soustraire la valeur moyenne $P_m(t)$ de la propriété considérée de la valeur réelle $P(t)$ et en divisant le résultat par la variance $v(t)$ des données. Au lieu d'utiliser directement les vraies données, nous employons la méthode des prétendus 'profils' définis par [Oso-97]:

$$pr(t) = \frac{P(t) - P_m(t)}{v(t)} \quad (IV.11)$$

Pour toutes les caractéristiques d'isolation participant à l'apprentissage, l'ensemble des profils forme les données d'apprentissage du réseau. Une fois que le réseau est entraîné, chaque neurone représente les données les plus proches de son vecteur de poids dans le sens métrique choisi.

En mode d'apprentissage, quand nous voulons faire une prédiction pour une caractéristique, nous fournissons ses données expérimentales réelles et le SOM trouve les neurones gagnants correspondants. Le profil prédit $pr_d(t)$, est alors la moyenne pondérée de tous les neurones gagnants trouvés :

$$pr_d(t) = \frac{\sum_{i=1}^{n_g} k_i w_i}{\sum_{i=1}^{n_g} k_i} \quad (IV.12)$$

n_g est le nombre de neurones gagnants.

k_i est le nombre d'apparitions du $i^{\text{ème}}$ neurone parmi les gagnants pour la caractéristique choisie et n est le nombre de neurones gagnants qui est égal au nombre de vecteurs

d'entrée utilisés. w_i est le facteur de pondération attribué à chaque neurone gagnant i correspondant à l'une des trois caractéristiques utilisées.

La prédiction de la caractéristique réelle est alors calculée à partir de l'équation (IV.11) par:

$$P_d(t) = v_d(t) \cdot pr_d(t) + P_{md}(t) \quad (IV.13)$$

Sur la base du profil prédit, la prédiction de la caractéristique réelle est calculée à partir de l'équation (IV.13). Le point le plus important après la détermination du profil prédit $pr_d(t)$, est de prédire la valeur moyenne $P_{md}(t)$ et la variance $v_d(t)$ au temps (t). Pour cette tâche nous avons appliqué un modèle linéaire de prédiction, tenant compte des valeurs appropriées ($P_m(t)$ et $v(t)$) des autres caractéristiques (2 pour le cas de la prédiction du vieillissement thermique) [Mok8-03, Mok9-03, Mok20-03, Mok22-03]. La prédiction linéaire de la valeur moyenne de la caractéristique $P_m(t)$ est donnée par S. Osowski [Oso-97] :

$$P_m(t) = a i + b \quad (IV.14)$$

i est une constante

Nous dénotons par w_k les valeurs de pondération de chaque caractéristique d'apprentissage. a et b sont des solutions du problème dont la fonction coût est définie par:

$$E = \sum_{j=1}^n [(aj + b) - P_j] w_j \quad (IV.15)$$

$$b = \frac{\sum_{k=1}^n k w_k \frac{\sum_{k=1}^n k P_k w_k}{k^2 w_k} - \sum_{k=1}^n P_k w_k}{\sum_{k=1}^n k w_k \frac{\sum_{k=1}^n k w_k}{k^2 w_k} - \sum_{k=1}^n w_k} \quad (IV.16)$$

$$a = \frac{\sum_{k=1}^n k P_k w_k - b \sum_{k=1}^n k w_k}{\sum_{k=1}^n k^2 w_k} \quad (IV.17)$$

P_k représente la valeur moyenne réelle du $k^{\text{ème}}$ vecteur d'entrée pour la prédiction de la valeur moyenne et représente la variance du $k^{\text{ème}}$ vecteur pour la prédiction de la variance et n le nombre de caractéristiques utilisées (3 pour le cas de la prédiction du vieillissement thermique).

Dans la méthode de calcul moyen pondéré donné par (IV.12), le neurone gagnant de la caractéristique à prédire est pondéré par la valeur la plus élevée proche de 1 et les autres caractéristiques sont pondérées avec des valeurs décroissantes. Ce mode de pondération est employé quand l'allure de la caractéristique à prédire est croissante. L'opposé est appliqué pour les caractéristiques décroissantes.

Différentes simulations, pour la méthode de pondération adoptée ci-dessus, ont été faites pour trouver les meilleures valeurs pour les coefficients de pondération w_k . Comme résultat de ces simulations et pour une courbe croissante, nous avons posé $w_1 = 0.98$ pour les données de la caractéristique à prédire, $w_2 = 0.65$ et $w_3 = 0.25$ pour les

autres caractéristiques d'apprentissage. Pour une courbe descendante, $w_1 = 0.25$, $w_2 = 0.52$ et $w_3 = 0.9$.

Nous avons pensé aussi à une autre méthode pour le choix des w_k , où ces poids représentent les pentes moyennes de chaque caractéristique d'apprentissage.

Cette méthode de prédiction de caractéristiques d'isolation que nous avons développée permet de prédire pour plusieurs pas du temps de vieillissement jusqu'à atteindre le temps de prédiction voulu.

Si nous voulons prédire pour une période plus longue que celle de l'expérience, nous entraînons le réseau ayant n valeurs dans le vecteur d'entrée. Après la première prédiction, le vecteur de sortie contiendra $n+1$ valeurs. Avec ce vecteur contenant $n+1$ valeurs, l'apprentissage sera lancé encore pour prédire une valeur $n+2$ et ainsi de suite [Mok8-03, Mok9-03, Mok20-03, Mok22-03].

Nous devons noter ici que la prédiction n'est pas totalement faite par le réseau de Kohonen SOM, mais elle l'utilise dans la phase de prédiction des profils.

4.3.3 Prédiction du vieillissement thermique du PVC et de la BORAK22

Les données expérimentales du PVC et de la BORAK22 ont été d'abord converties en 'profils', qui sont employés après, dans la phase d'apprentissage. Le réseau de Kohonen utilisé contient 40 neurones distribués sur des hexagones et le rayon de voisinage est donné par $D = 1$.

Après l'apprentissage, les neurones sont groupés dans des classes séparées correspondant au type de courbes présentées : ascendante, descendante, et la troisième catégorie contient les deux types de courbes citées. Cette phase est nécessaire pour prédire le profil. Nous notons que toutes les courbes expérimentales utilisées sont non linéaires.

Nous présentons à la figure 4.24 la prédiction de la résistance à la traction du PVC vieilli à une température de vieillissement de 120°C utilisant la carte de Kohonen SOM. La prédiction est faite jusqu'à 3500h pour un temps d'apprentissage de 1500h [Mok8-03, Mok9-03]. Les poids w_k utilisés pour cette courbe sont ceux donnés pour une courbe croissante.

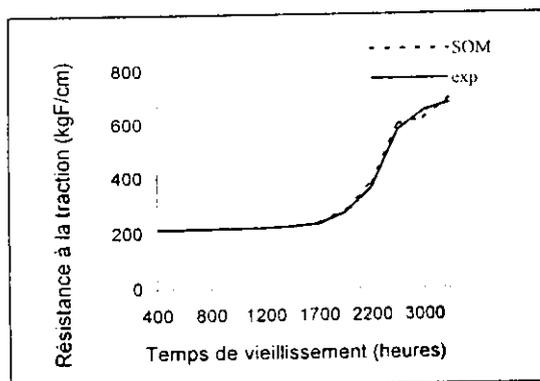


Fig.4.24 Prédiction de la résistance à la traction pour le PVC utilisant la carte de Kohonen $T = 120^\circ\text{C}$

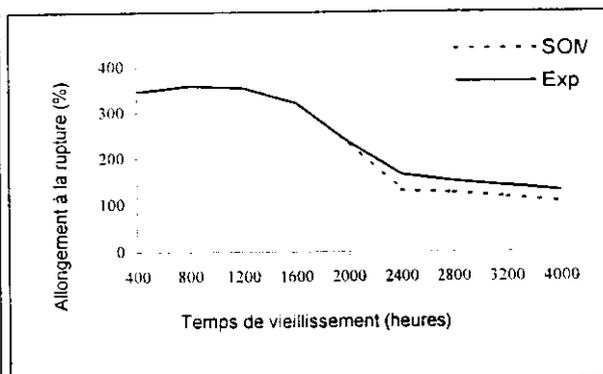


Fig.4.25 Prédiction de l'allongement à la rupture pour le PVC utilisant la carte de Kohonen $T = 100^\circ\text{C}$

Pour les résultats donnés à la figure 4.25, les poids utilisés dans (IV.16) et (IV.17) sont ceux appropriés à la classe de courbe descendantes. Les résultats présentés à cette figure donnent la prédiction de l'allongement à la rupture du PVC jusqu'à 4000h pour un temps d'apprentissage de 2500h à $T = 100^\circ\text{C}$. Les résultats de la prédiction sont en bon accord avec les valeurs expérimentales.

A la figure 4.26, nous donnons la prédiction du facteur de pertes diélectriques pour la BORAK22 vieillie à 120°C. La prédiction est faite jusqu'à 1850h pour un temps d'apprentissage de 1200h et un pas de 150h.

Pour tous les résultats numériques, nous avons noté que le réseau de Kohonen utilisé est un bon réseau pour la prédiction.

Pour notre application, l'exactitude de la prédiction SOM excède les réseaux multicouches présentés en utilisant les réseaux RBFG. La capacité du SOM de classer ses entrées et puis prédire les profils lui donne une meilleure capacité dans la prédiction [Mok8-03]. L'allure de la courbe expérimentale est un paramètre qui contrôle l'apprentissage et la courbe qui contient un point d'inflexion reste difficile à apprendre.

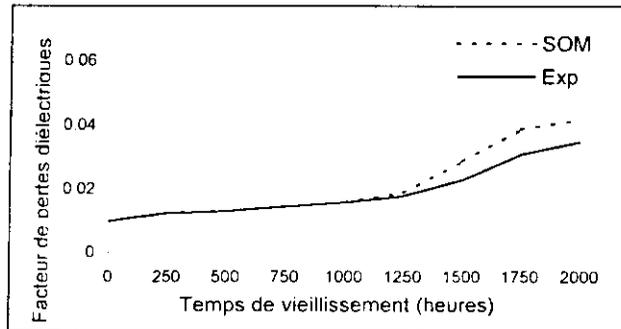


Fig. 4.26 Prédiction du facteur de pertes diélectriques utilisant la carte de Kohonen T= 120°C

Dans une autre méthode, les pondérations w_k sont calculés à partir des pentes partielles de chaque courbe expérimentale de l'allongement à la rupture (Fig.27). Les résultats sont bons mais moins précis que ceux donnés à la figure 4.25.

Utilisant la méthode de détermination des poids w_k à partir des pentes moyennes de chaque courbe expérimentale, trois exemples sont donnés pour prédire l'allongement à la rupture (Fig. 4.27, Fig.4.29) et la résistance à la traction (Fig.4.28). Aux figures 4.27 et 4.28, les pentes moyennes utilisées sont calculées à partir de pentes partielles, mais pour la figure 4.29, les pentes utilisées sont les pentes moyennes globales de chaque caractéristique utilisée dans l'apprentissage.

Les poids positifs indiquent que la courbe correspondante est croissante et les poids négatifs indiquent que la courbe est décroissante.

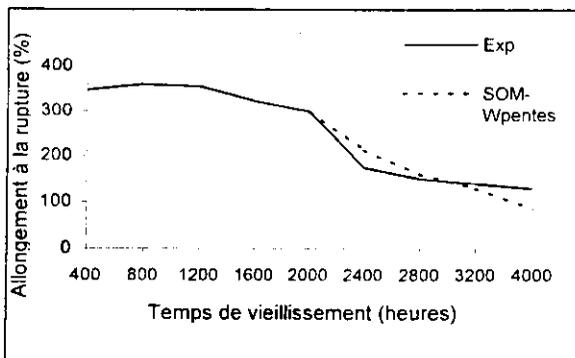


Fig. 4.27 Prédiction de l'allongement à la rupture du PVC utilisant les SOM T = 100°C
 $w_1 = -0.0652, w_2 = -0.35, w_3 = -0.62$

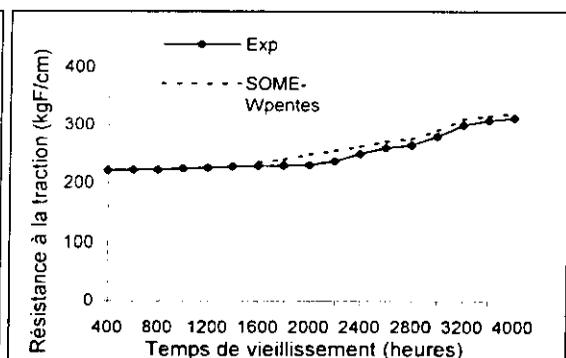


Fig. 4.28 Prédiction de la résistance à la traction du PVC utilisant les SOM T = 100°C
 $w_1 = 0.00215, w_2 = 0.000894, w_3 = 0.0216$

La légende contenant la lettre 'P' indique la courbe expérimentale et la lettre 'R' indique la courbe prédite à la figure 4.29. Les nombres 80, 100, 120 sont les températures de vieillissement du PVC. Pour cette figure toujours, nous avons utilisé par exemple, un

temps d'apprentissage de 1600h pour la caractéristique qui correspond à la température de vieillissement de 80°C et la prédiction est faite pour 5000h.

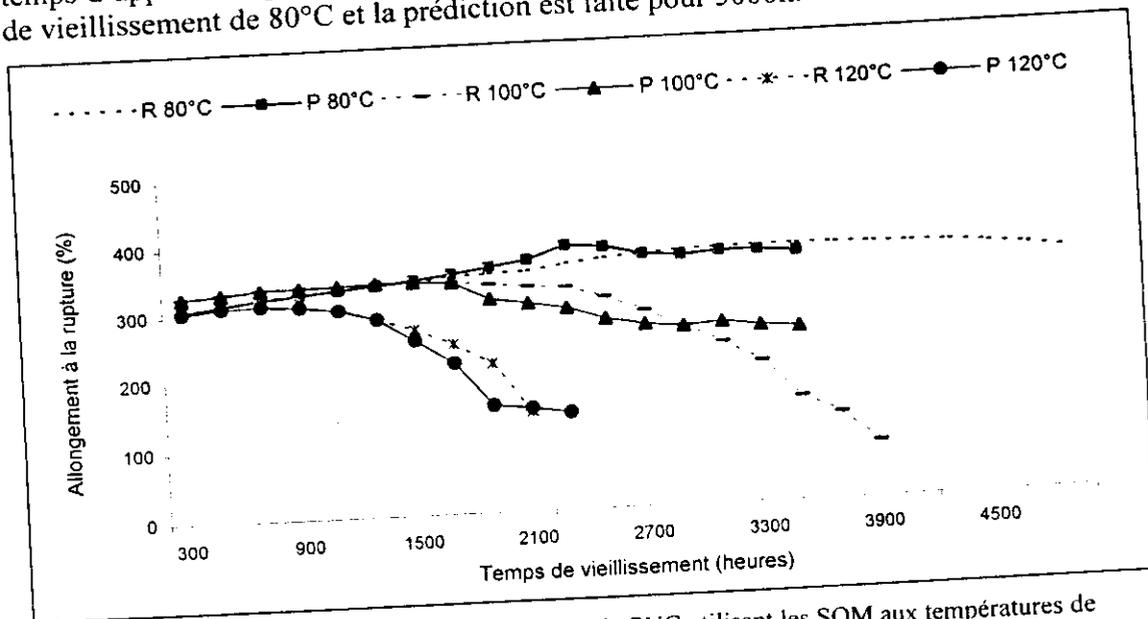


Fig.4.29 Prédiction de l'allongement à la rupture du PVC utilisant les SOM aux températures de vieillissement 80°C, 100°C, 120°C
 Pour T = 80°C : $w1=0.0180556$, $w2=-0.025$, $w3=0.006561106$.
 Pour T = 100°C : $w1=0.0048056$, $w2=0.0180556$, $w3=0.0016947011$

4.3.4 Prédiction de la tenue diélectrique des intervalles d'air pointe-barrière-plan

Le SOM est appliqué aussi pour prédire la tension disruptive à 50% d'un intervalle d'air. Nous présentons aux figures 4.30 et 4.31 les valeurs mesurées et prédites de la tension disruptive à 50% pour un intervalle d'air pointe-barrière-plan en fonction de la longueur de l'intervalle d'air pointe-plan (d) respectivement sous une tension de choc de manœuvres et une tension de choc de foudre [Mok20-03, Mok22-03]. La largeur de la barrière isolante est de 1m. Les résultats présentés aux deux figures montrent une bonne concordance avec les caractéristiques expérimentales de la tension disruptive. Une première prédiction est faite pour la distance 200cm. Cette valeur prédite y_{i+1} devient une valeur d'entrée pour prédire la valeur y_{i+2} dans le même vecteur d'entrée correspondant à 225cm et ainsi de suite. Une prédiction est faite jusqu'à une distance inter-électrodes de 300cm pour la tension de choc de manœuvres (Fig.4.30) et 250cm pour la tension de choc de foudre (Fig.4.31).

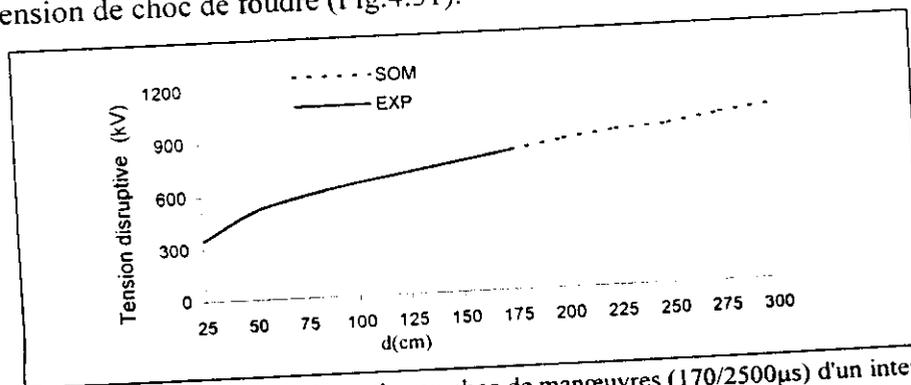


Fig.4.30 Prédiction de la tension disruptive au choc de manœuvres (170/2500µs) d'un intervalle d'air pointe-barrière-plan en fonction de la longueur de l'intervalle d'air utilisant les SOM.

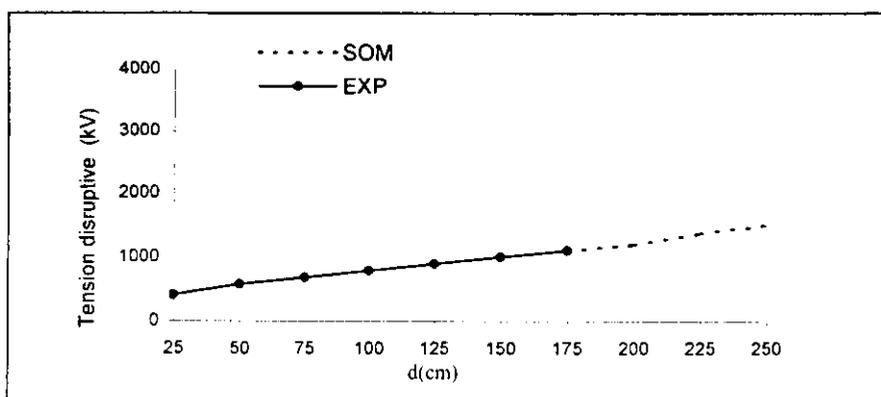


Fig.4.31 Prédiction de la tension disruptive au choc de foudre (1,2/50µs) d'un intervalle d'air pointe-barrière-plan en fonction de la longueur de l'intervalle d'air utilisant les SOM

A la figure 4.32, la prédiction de la tension disruptive est faite pour deux largeurs de la barrière (40cm et 110cm) en fonction de la longueur de l'intervalle d'air pointe-barrière-plan et sous une tension à fréquence industrielle

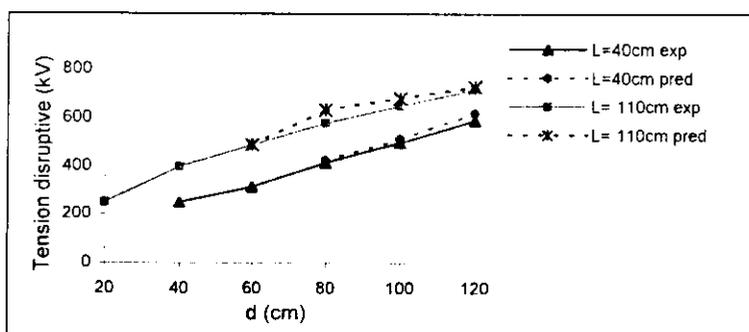


Fig.4.32 Prédiction de la tension disruptive à fréquence industrielle d'un intervalle d'air pointe-barrière-plan pour différentes largeurs de la barrière en utilisant les SOM

Dans la figure 4.33, nous donnons la prédiction de la tension disruptive à fréquence industrielle d'un intervalle d'air pointe-barrière-plan pour différents rapports distance pointe-barrière/distance pointe-plan (a/d), en utilisant les pentes globales des caractéristiques pour les valeurs de w_k . La qualité de la prédiction est bonne.

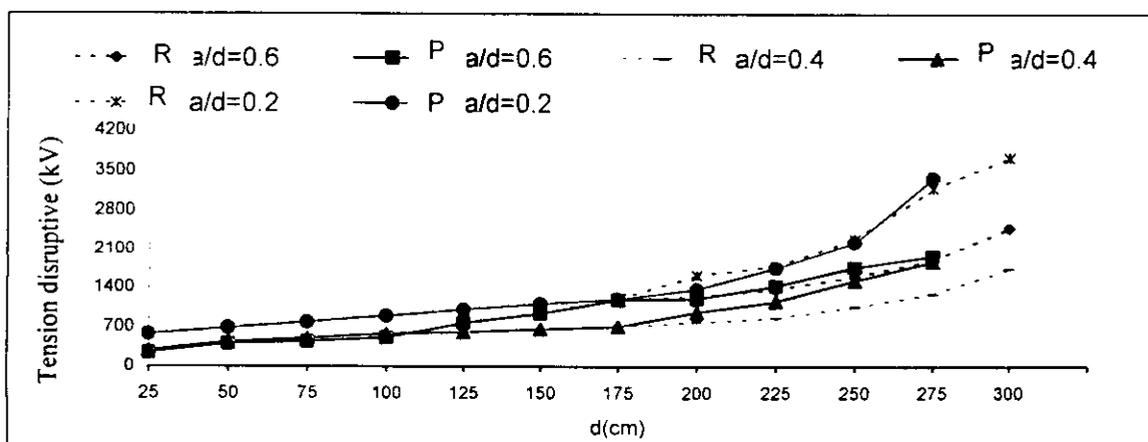


Fig.4.33 Prédiction de la tension disruptive au choc de foudre (1,2/50µs) d'un intervalle d'air pointe-barrière-plan pour différents rapports (a/d) utilisant les SOM.

Pour $a/d = 0.2$: $w_1=0.1732$, $w_2=0.1573$, $w_3=0.14108$.
 Pour $a/d = 0.4$: $w_1=0.02794$, $w_2=0.01732$, $w_3=0.017829821$
 Pour $a/d = 0.6$: $w_1=0.02794$, $w_2=0.01573$, $w_3=0.0216728$

4.4 DISCUSSIONS

Le processus de vieillissement des isolants peut être surveillé par plusieurs mesures de propriétés. Pour prédire le comportement diélectrique de l'isolation, on a besoin d'une longue période d'essai pour avoir une base de données assez complète afin de donner une bonne qualité de prédiction. L'utilisation des réseaux de neurones dans le vieillissement thermique des isolants de HT est économique. Elle permet de réduire le temps de vieillissement d'essai au laboratoire et laisse le réseau prédire les valeurs des propriétés pour des durées plus longues.

La prédiction du vieillissement thermique du PVC et de la BORAK22 peut aider considérablement à améliorer la qualité de leur maintenance.

Les mêmes RN sont développés pour la prédiction de la tension disruptive dans un intervalle d'air pointe-barrière-plan en fonction de la longueur de l'intervalle d'air pour plusieurs positions de la barrière dans l'intervalle et pour plusieurs largeurs de la barrière. Les trois types de tensions sont appliqués : choc de foudre, choc de manœuvres et à fréquence industrielle. Les résultats paraissent prometteurs pour l'usage dans l'industrie.

Dans un volet du travail présenté dans ce chapitre, l'application du réseau RBFG entraîné par la ROM a été présentée et offre une bonne qualité de prédiction. Le choix du réseau RBFG est basé sur le fait qu'il résout le problème du choix du nombre de couches cachées et qu'il soit indépendant du problème à résoudre.

L'optimisation d'un réseau n'est pas une tâche facile et sa convergence peut être longue comme il peut diverger complètement. La caractéristique locale du RBFG évite ce problème. Trois principaux réseaux RBFG ont été utilisés : le premier entraîné par la BP, le deuxième par la ROM et le troisième entraîné par la ROM couplée avec la méthode des K-means clustering. Utilisant les deux derniers réseaux, la technique Batch learning a présenté une meilleure performance en comparaison avec la technique FFN pour la prédiction du vieillissement thermique.

Le RBFG entraîné par la ROM présente les meilleurs résultats le comparant avec le RBFG entraîné par la BP car il présente les meilleures performances: sa capacité d'ajuster les poids sans déterminer aucun gradient, ses capacités de généralisation et sa précision.

Le réseau RBFG entraîné par la ROM couplée avec la méthode des K-means clustering, présente les meilleurs résultats le comparant à son tour au RBFG entraîné par la ROM. La distribution des centres selon les classes qu'ils représentent, affecte considérablement la qualité de la prédiction. La disposition des centres à deux dimensions a amélioré nettement la prédiction en comparant avec les résultats présentés avec un treillis régulier (une dimension).

Dans un deuxième volet de ce chapitre, les cartes auto-organisatrices de Kohonen ont été utilisées. Dans la méthode des 'profils', les réseaux SOM montrent une meilleure performance dans le classement de leurs entrées avant d'être utilisés dans la prédiction. Les SOM sont dans la pratique flexibles et faciles à implémenter. L'exactitude de la prédiction dépend de la qualité des données d'entrée.

Si nous voulons prédire pour plusieurs pas du temps de vieillissement, le classement du nouveau vecteur entrée est fait pour chaque prédiction, par conséquent le nombre de neurones augmentera.

Comparant les résultats donnés par les réseaux RBFG et SOM, les deux réseaux présentent de bons résultats. La différence est dans la facilité d'utilisation des SOM. L'application de ces derniers directement dans la prédiction des caractéristiques est un axe de recherche intéressant, sur lequel nous comptons travailler à l'avenir.

4.5 CONCLUSION

Il serait très économique d'employer les RNA dans les investigations sur l'isolation de HT. Nous pouvons réduire les essais au laboratoire et laisser le réseau prédire les propriétés d'isolation.

Le choix du réseau RBF est justifié par son indépendance du modèle mathématique qui régit le système à traiter.

Trois réseaux RBF ont été présentés : ceux entraînés par la ROM ont présenté leur capacité d'ajuster les poids sans déterminer aucun gradient. L'un de ces réseaux est entraîné par la ROM ayant une distribution des centres en treillis réguliers et l'autre réseau est une tentative originale pour combiner deux techniques connues : le RBF entraîné par la ROM dont les centres sont distribués en treillis régulier à deux dimensions et la méthode des K-means clustering. Cette dernière est une méthode non supervisée qui partage l'ensemble des données en classes et en attribue des centres. Son couplage avec la ROM a amélioré les performances de cette dernière méthode.

Le troisième RBF entraîné par la BP a donné aussi des résultats satisfaisants mais avec une difficulté du choix des taux d'apprentissage.

Le choix de la distribution des centres gaussiens est un choix crucial pour la qualité de l'apprentissage du réseau RBF.

Les résultats de simulation donnent pour la distribution des centres à deux dimensions les meilleures performances de prédiction pour les réseaux RBF.

Les cartes auto-organisatrices de Kohonen ont été aussi utilisées pour la prédiction du vieillissement thermique du PVC et de la BORAK22 et de la tension disruptive des intervalles d'air pointe-barrière-plan. L'utilisation des SOM dans la prédiction se limite seulement à la prédiction des 'profils' qui sont une autre représentation des données pour les rendre indépendantes des caractéristiques expérimentales utilisées.

Les SOM présentent de bons résultats et une facilité d'implémentation du réseau.

4.6 BIBLIOGRAPHIE

- [Abd-02] S.Abdi,'Influence du vieillissement thermique sur les propriétés de l'huile de transformateur', thèse de Magister, ENP, 2002.
- [Bou1-00] A. Boubakeur, L. Mokhnache, A.L. Nemmour, A. Khelifi, 'Neural network application in the field of HV cable insulation thermal ageing', 3rd Int. Conf. Quality, Reliability and Maintenance (QRM), Professional Publishers IMecE, pp.122-126, Oxford, UK, Apr 2000.
- [Bou8-79] A.Boubakeur,'Influence des barrières sur la tension de décharge disruptive des moyens intervalles d'air pointe-plan', thèse de Doctorat, E.P Varsovie, 1979.
- [Bou10-98] A. Boubakeur, L. Mokhnache, S. Boukhtache,' theoretical investigation on barrier effect in point- plane air gap using finite elements method in presence of space charge' Int. conf. POWERCON'98, IEEE-proceedings, PP 116-121,Beijing, China, Aug. 1998.
- [Bou11-99] A. Boubakeur, L. Mokhnache, S. Boukhtache,' Numerical Model of Electrostatic Barrier Effect in a Point-Plane Air Gap in Presence of Positive Space Charge', Archives of Electrical Eng., Polish Academy of Science (AEE), VOL.XLVIII N°189, 3./ 1999.
- [Bou12-03] A. Boubakeur, L. Mokhnache, A. Feliachi, 'Theoretical Investigation on Barrier Effect on Point- Plane Air Gaps Breakdown Voltage Based on Streamers Criterion' déposée pour publication à IEE Proceedings - Science, Measurement and Technology, Feb.2003

- [Broo-88] D.S. Broomhead, D. Lowe, 'Multivariable function interpolation and adaptive networks', Complex systems, Vol.2, 1988.
- [Khe-94] S. Khemaissia, A.S. Morris 'Review of Networks and choice of radial basis function Networks for systems identification', Technologies Avancées, NG. pp 55-85, 1994.
- [Koh3-95] T. Kohonen, 'Self-Organizing Maps', Springer Verlag, Berlin. 1995.
- [Mok1-01] L. Mokhnache, A. Boubakeur, 'Prediction of the Breakdown Voltage in a Point-Barrier-Plane Air gap using Neural Networks', IEEE Annual report Conf. on Electrical Insulation and Dielectric Phenomena (CEIDP), pp: 369-372, Ontario- Canada, Oct. 2001
- [Mok2-01]. L.Mokhnache, A.Boubakeur, 'Neural networks application to predict high voltage insulation thermal ageing', 12th Int. Symp. on High Voltage Eng., (ISH), Vol. 5, paper N° 8-8, Banagalore, India, 2001.
- [Mok3-01] L.Mokhnache, A.Boubakeur, B.Ould Noureddine, M.A.R.Bedja, 'Prediction Of Long Thermal Ageing Of Transformer Oil Using Neural Networks', Proceedings of the Int. Conf. on Advances in Processing, Testing and Application of Dielectric Materials (APDATM), N° Spécial du journal PRZEGLAD ELEKTROTECHNICZNY, pp: 277-280, Wroclaw, Poland, Sept. 2001.
- [Mok4-01] L.Mokhnache, A.Boubakeur, N.Nait Said, ' Application of Neural Networks Paradigms in the Diagnosis and Thermal Ageing Prediction of Transformer Oil', IEEE- Int. Conf. On Dielectric Liquids (ICDL), IEEE-DEIS proceedings, Paper N° 048, pp: 258-261, Graz, Austria, Juil. 2002.
- Mok6-02] L.Mokhnache, A.Boubakeur, ' The Use of some Paradigms of Neural Networks in Prediction of Dielectric properties for High Voltage liquid-solid and gas Insulations', Int. Sym. on Electrical Insulation (ISEI), IEEE-DEIS proceedings, pp; 306-309, Boston, Massachusetts, USA. Apr. 2002.
- [Mok8-03] L.Mokhnache, A.Boubakeur, 'Self-Organization neural networks in prediction of insulation thermal ageing' ,ISH2003, IEEE-DEIS proceedings, Delft-Netherlands, Aug. 2003.
- [Mok9-03] L.Mokhnache, A.Boubakeur, 'RBF and Self-Organization neural networks in prediction of insulation thermal ageing', CNHT2003, Oran-Algeria, Dec. 2003.
- [Mok10-01] L.Mokhnache, A.Boubakeur, A.Khelifi, A.L.Nemmour, 'Neural Networks Application in the Thermal Ageing Prediction of the PVC used in the High Voltage Cables', CD-ROM 1^{er} Symp. Int. de la Maintenance Industrielle, SIMI, session 7, Alger, Algérie, Jan. 2001.
- [Mok11-01] L. Mokhnache, A. Boubakeur, A. Feliachi and al., 'Application of neural networks in the thermal ageing prediction of transformer oil', IEEE-PES summer meeting, paper N°01sm059; Vancouver, Canada, Juil. 2001.
- [Mok12-01] L. Mokhnache, A. Boubakeur, 'The Use of Neural Networks in Prediction of Dielectric properties for High Voltage liquid-solid and gas Insulations' CD-ROM, 3^{ème} Sym. des Industries Électriques du Monde Arabe, Alger, Algérie, Nov. 2001.
- [Mok16-02] L.Mokhnache , A.Boubakeur, N. Nait Said, 'Comparison of Different Neural Networks Algorithms Used in the Diagnosis and thermal ageing prediction of Transformer Oil', Int. Conf. on Systems, Man & Cybernetics, IEEE-SMC CD-ROM, paper ID: SMC02-REG-485, Hammamat, Tunisia Oct.,2002.

- [Mok17-02] L.Mokhnache, A.Boubakeur, 'Insulation properties prediction applying artificial neural networks', 5^{ème} Conf. Nat. sur la Physique Appliquée (CNPA), PP: 261-262, Batna, Algeria, Oct. 2002.
- [Mok18-03]] L.Mokhnache , A.Boubakeur,'RBF Neural Networks for Insulation Thermal Ageing', Int. Conf. On Artificial Neural Networks and Genetic Algorithms, (ICANNGA) Proceedings, pp: 144-148, Springer-Verlag, Roanne-France, Apr.2003
- [Mok20-03] L.Mokhnache, A.Boubakeur, 'Self-Organization Neural Networks In Prediction of the breakdown voltage in a Point Barrier–Plane Air Gap using different barrier widths', GCC, IEEE proceedings, paper ID: 8, Bahrain, May 2003.
- [Mok22-03] L.Mokhnache, A.Boubakeur,'Self-Organization Neural Networks In Prediction of the breakdown voltage in a Point Barrier–Plane Air Gap', Trans. of 4th Int. Conf. on Neural Networks and Applications of World Scientific And Engineering Academy And Society (WSEAS) , ID: 454-187, Athena-Greece, Jun. 2003.
- [Mok23-03] L. Mokhnache, A. Boubakeur, A. Feliachi, 'Thermal Ageing Prediction of Transformer Oil and PVC of High Voltage Cables Using Neural Networks', IEE Proceedings - Science, Measurement and Technology – pp: 107-112, Vol. 150, Issue 03, May 2003.
- [Mok24-03] L. Mokhnache, A. Boubakeur,' Use of Neural Networks in the Monitoring of High Voltage Insulation Thermal Ageing', acceptation finale pour publication dans Journal of COMADEM, Oct. 2003.
- [Mok25-03] L.Mokhnache, A.Boubakeur, 'Application of Many Neural Networks Paradigms in the High Voltage Insulation, Archives of Electrical Engineering (AEE), Polish Academy of Science, Vol.3, Oct. 2003.
- [Mok26-97] L. Mokhnache, A. Boubakeur, S. Boukhtache' Modèle numérique du calcul du champ électrique dans un système pointe-barrière-plan en présence de la charge d'espace', 2^{ème} CNHT'97, Tizi Ouzou, Algérie, Nov.97.
- [Mok27-97] L. Mokhnache, A. Boubakeur, S. Boukhtache,'Calcul Numérique du champ et du potentiel dans un intervalle d'air pointe-barrière-plan en présence de la charge d'espace', 1st Int. Conf. Of Electronique, Electrotechnique & Automatique (IEEA), pp. 244-248, Batna-Algérie, Dec. 1997.
- [Mok28-98] L. Mokhnache, A. Boubakeur, S. Boukhtache,'Application of finite elements method to electric field calculation in point-barrier-plane arrangement in presence of space charge'.4th Int. Workshop on Electric and Magnetic Fields (WEMF), pp 111- 116, Marseille-France, 1998.
- [Mok29-00] L. Mokhnache, A. Boubakeur, S. Boukhtache 'Numerical Model Describing the Effect(s) of a Barrier and the Space Charge Fields on the Electrical Strength of a Point-Plane Air Gap using the finite elements', CEIDP, IEEE-DEIS Proceedings, pp 466-469, Victoria (BC)- Canada, Oct. 2000.
- [Mok30-00] L. Mokhnache, A. Boubakeur, S. Boukhtache 'Numerical Model Describing the Effect of a Barrier and the Space Charge Fields on the Electrical Strength of a Point-Plane Air Gap' Universities Power Engineering Conf. (UPEC), pp: 211, Belfast, UK Sept. 2000.
- [Mok31-03] L.Mokhnache,'Contribution à l'étude de l'influence des barrières dans les intervalles d'air pointe-plan par le calcul numérique du champ à l'aide de la méthode des éléments finis avec et sans charge d'espace', thèse de Magister, Université de Batna, Juil.1997.

- [Nej1-91] M.Nedjar,'Influence du vieillissement thermique sur les propriétés du polychlorure de Vinyle utilisé dans l'isolation des câbles de haute tension', Thèse de Magister Université de Tizi-Ouzou, (1991).
- [Oso-97] S. Osowski, K. Swek,'Kohonen neural network for load forecasting in power system', XX KKTOIUE, Kolobrzeg, 1997.
- [Pol-94] R. Polli, R.D. Jones 'A neural net model for prediction', Journal of the American Statistical Association, Vol. 89, N°. 425, pp. 117-121, March (1994).
- [Sol-81] F.J. Sollis, R.J.B. Wetts,'Minimization by random search techniques',Math.Oper.Res.6,(1),pp 19-29, 1981.

CONCLUSION GÉNÉRALE

CONCLUSIONS

Les réseaux de neurones artificiels sont appliqués dans ce travail dans le diagnostic et la classification de l'huile de transformateurs et dans la prédiction des propriétés d'isolants solides, liquides et gazeux. Les conclusions suivantes mettent en lumière les résultats obtenus.

- Les limitations des méthodes traditionnelles de diagnostic d'huile de transformateurs sont dues essentiellement à leur forte dépendance de l'expérience humaine, d'où les réseaux de neurones (RN) peuvent apporter de bonnes solutions.
- Les réseaux de neurones supervisés et non supervisés sont de bons systèmes pour le diagnostic et la classification pour les nouveaux ingénieurs analystes non expérimentés afin de prendre les bonnes décisions concernant l'état de l'huile et en un temps réduit.
- L'optimisation d'un réseau est une fin importante dans l'amélioration de la qualité de l'apprentissage des RN. Des chercheurs [Din-95, Zha-97] ont défini un RN optimal par une convergence rapide et une petite erreur. Mais la convergence rapide reflète une régularité (monotonie) dans la base d'apprentissage qui peut ne pas couvrir tous les cas pouvant exister dans l'état de l'huile, surtout si cette base d'apprentissage n'est pas assez large pour être représentative. L'exactitude des résultats est le seul paramètre qui peut confirmer l'optimisation du réseau utilisé et ceci exige la connaissance des sorties (décisions) de la base de généralisation (test) aussi. Pour cela une large base de données est nécessaire.
- En se basant sur la conclusion citée juste ci-dessus, les réseaux de Levenberg-Marquardt et les réseaux bayésiens présentent une bonne robustesse pour le diagnostic.
- L'architecture du RNA (nombre de couches cachées, nombre de neurones) est un facteur important décidant de la qualité de l'apprentissage plus que les paramètres d'apprentissage (taux d'apprentissage, ...).
- La taille et la qualité de la base de données ont une influence majeure sur l'apprentissage des RNA. C'est pour cela que le choix de cette base qui est parfois le seul paramètre décidant de la qualité de l'apprentissage est une étape prudente dans les applications des réseaux de neurones.
- En pratique, le système de diagnostic utilisé dans ce travail, ressemble à un système off-line d'où les RN utilisés ne demandent pas à être entraînés en temps réel. Par conséquent une convergence rapide des réseaux utilisés n'est pas obligatoire. Alors, les réseaux de la BP avec momentum et la BP avec momentum et à adaptation du taux d'apprentissage présentent aussi une qualité de diagnostic acceptable.
- La classification de l'huile peut être utilisée comme une phase d'aide à la décision, surtout si la base d'apprentissage est assez grande, que l'apprentissage risque de rencontrer des problèmes de stagnation.

- Les réseaux non supervisés ont montré une grande souplesse dans la classification et l'apprentissage de ces réseaux ne présente pas beaucoup de complexité comme dans le cas de certains réseaux supervisés.
- La prédiction du vieillissement thermique du PVC et de la BORAK22 peut aider considérablement à améliorer la qualité de leur maintenance. Cette prédiction réduit aussi le temps des essais de vieillissement qui sont coûteux.
- Le dimensionnement des intervalles d'air pointe-barrière-plan exige beaucoup d'essais tenant compte de tous les paramètres influant sur la rigidité du système. La prédiction des tensions disruptives des intervalles d'air permettra de prévoir la coordination de l'isolement des ouvrages de haute tension.
- D'un autre côté, les modèles décrivant la décharge dans un intervalle d'air n'ont pas pu être combinés ensemble dans un modèle général, capable de prévoir le comportement d'un espace donné sujet à une tension donnée. Les RN ont pu présenter une solution.
- L'application du réseau RBFG offre une bonne qualité de prédiction. L'utilisation de ce réseau résout le problème du choix de topologie.
- L'optimisation d'un réseau n'est pas une tâche facile et sa convergence peut être lente, comme il peut diverger complètement. La caractéristique locale du RBFG évite ce problème.
- Les résultats donnés par le réseau RBFG entraîné par la ROM couplée à la méthode des K-means clustering à deux dimensions (temps de vieillissement, caractéristique à prédire) présentent les meilleures performances du RBFG. Nous confirmons que la distribution des centres affecte considérablement la qualité de la prédiction.
- La combinaison d'un réseau supervisé et d'un autre non supervisé pour la prédiction a donné au réseau résultant ayant la précision et la souplesse qu'avait chacun d'eux séparés.
- Les performances présentées du RBFG entraîné par la ROM sont meilleures que celles du RBFG entraîné par la BP en plus de la simplicité de l'entraînement du premier.
- Les SOM présentent aussi de bons résultats dans la prédiction et une facilité d'implémentation du réseau. Ceci est obtenu grâce à la méthode des profils qui rend la prédiction indépendante de la nature de la caractéristique à prédire.

CONTRIBUTIONS

Les contributions majeures sont résumées comme suit:

- ◊ Un système de diagnostic et de classification d'huile de transformateur utilisée par SONELGAZ a été élaboré en appliquant les réseaux de neurones artificiels (RNA).
- ◊ Le diagnostic physico-chimique demande l'expérience humaine. L'application des réseaux de neurones est l'une des meilleures façons de conserver efficacement cette expérience. Elle servira surtout à aider les jeunes ingénieurs analystes à prendre les bonnes décisions en liaison avec la maintenance de l'huile de transformateurs.
- ◊ L'optimisation des RNA utilisés dans le diagnostic est réalisée en changeant surtout d'algorithmes d'entraînement et de topologie du réseau et le critère d'optimisation est basé sur l'exactitude des résultats.
- ◊ L'application des RNA dans la prédiction des caractéristiques du PVC, de l'huile de transformateurs et des intervalles d'air pointe-barrière-plan a pu réduire le coût et la durée des essais nécessaires et pourra améliorer la qualité de leur maintenance.

PERSPECTIVES

Ce travail est une partie de deux projets de recherche ; le premier au laboratoire de haute tension à l'ENP sur l'utilisation des RNA dans l'isolation de haute tension et le second au laboratoire L.S.P.I-E à l'université de Batna sur l'application des méthodes d'intelligence artificielles dans les équipements électriques.

Nous avons commencé à utiliser la logique floue dans le diagnostic et les premiers résultats sont très encourageants surtout que l'implémentation de cette méthode a présenté plus de souplesse qu'avec les RNA. Les résultats de ce travail seront communiqués à la prochaine conférence internationale sur la Qualité, Fiabilité et Maintenance (QRM) à Oxford (2004).

L'utilisation de systèmes hybrides (neuro-flous ou neuro-génétiques) donneront encore de meilleures performances aux systèmes de diagnostic.

Des mesures de gaz dissous dans l'huile de transformateurs, faites par des chercheurs de l'ENP nous permettront d'utiliser une autre méthode de diagnostic d'huile en utilisant les méthodes intelligentes suscitées.

Une coopération avec l'institut d'Engineering & Technology de Thapar de l'université de Deemed – Patiala (Inde) nous a permis d'avoir une base de données de mesures faites pour d'autres isolants solides. Ceci nous permettra de compléter notre système de diagnostic pour couvrir les isolants solides aussi.

L'utilisation des cartes topologiques de Kohonen (SOM) directement dans la prédiction est un axe majeur dans nos futurs travaux. Ceci a commencé par la collaboration du Professeur T.Kohonen, fondateur des réseaux de neurones non supervisés qui prennent son nom.

ANNEXE 1

Matrice hessienne

Définition de la matrice hessienne (le hessien)

La matrice hessienne est la matrice des dérivées secondes d'une fonction multivariées. C'est-à-dire, le gradient du gradient d'une fonction.

Le hessien d'une fonction scalaire d'un vecteur est la matrice des secondes dérivées partielles. Ainsi, la matrice hessienne d'une fonction $f: \mathbf{R}^n \rightarrow \mathbf{R}$ a la forme générale suivante:

$$H = \frac{\partial \mathbf{F}}{\partial \mathbf{x}} = \begin{matrix} & \mathbf{x}_1 & \mathbf{x}_2 & \cdot & \cdot & \mathbf{x}_n \\ \left[\begin{array}{cccc} \frac{\partial \bar{F}_0}{\partial x_0} & \frac{\partial \bar{F}_0}{\partial x_1} & \cdot & \cdot & \frac{\partial \bar{F}_0}{\partial x_n} \\ \frac{\partial \bar{F}_1}{\partial x_0} & \frac{\partial \bar{F}_1}{\partial x_1} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \frac{\partial \bar{F}_{n-1}}{\partial x_{n-1}} & \frac{\partial \bar{F}_{n-1}}{\partial x_n} \\ \frac{\partial \bar{F}_n}{\partial x_0} & \cdot & \cdot & \frac{\partial \bar{F}_n}{\partial x_{n-1}} & \frac{\partial \bar{F}_n}{\partial x_n} \end{array} \right] & \begin{matrix} \mathbf{F}_1 \\ \mathbf{F}_2 \\ \cdot \\ \mathbf{F}_{n-1} \\ \mathbf{F}_n \end{matrix} \end{matrix}$$

avec

$$\frac{\partial \bar{F}_{(i,j)}}{\partial \bar{x}_i} = \frac{\partial^2 E(\mathbf{x})}{\partial x_i \partial x_j}$$

$E(x)$ peut être l'erreur du réseau de neurones

La matrice hessienne est symétrique en raison de l'égalité des dérivées partielles mixtes.

L'utilisation de la hessienne dans les RNA

Le but de l'algorithme de la Back-Propagation, qui est très utilisé dans les RNA, est de trouver un minimum pour la fonction objective ; mais lors de la recherche de solution, on ne tient pas compte des propriétés locales.

Il y a beaucoup de méthodes pour améliorer cet algorithme. Elles consistent à minimiser la fonction objective en l'approchant par une forme quadratique. Par exemple, l'algorithme des régions de confiance (TR : Trust Region) a été développé pour améliorer l'algorithme d'apprentissage. Il consiste à approcher successivement la fonction objective par une forme quadratique et à réduire cette fonction dans la direction qui minimise cette forme quadratique. La taille de la région sur laquelle est approchée cette fonction est déterminée directement par la mesure de la qualité de l'approximation. Comme la forme quadratique est une approximation du second ordre qui dépend du gradient et du hessien de la fonction objective, le minimum trouvé possède aussi les bonnes propriétés du second ordre.

ANNEXE 2

Application et interprétation des essais

Propriété	Lieu d'essai S- site L- laboratoire	Catégorie de matériel	Fréquence des essais	Limites recommandées nécessitant une intervention	Intervention	Notes
Couleur et aspect	S ou L	O, A, B, C, D, E	Simultanément avec d'autres essais quantitatifs	Limpides sans contamination visible	Suivant les indications d'autres essais	
Tension de claquage	S ou L	O, A, B, C, D, E, F, G	<p>O, A, B Après remplissage ou re-remplissage et avant mise sous tension . Ensuite après 12 mois puis tous les deux mois.</p> <p>C, D, E Après remplissage ou re-remplissage et avant mise sous tension. Ensuite après 12 mois puis tous les six mois.</p> <p>F Après remplissage ou re-remplissage et avant mise sous tension . ensuite tous les quatre ans ou toutes les 70000 manœuvres si ce nombre est atteint avant ou suivant les instructions du constructeur.</p> <p>G Se référer aux spécifications du constructeur (voir paragraphe 8 2 2 note 3).</p>	<p>O, A, B > 50kV B, E > 40Kv C > 30kV G > 20Kv</p> <p>Rupteurs des changeurs de prise en charge en étoile ou en triangle des transformateurs. O, A, B > 40kV Transformateurs C > 30kV.</p>	Retraiter l'huile en variantes si c'est économique ou si des résultats d'autres essais le confirment, remplacer l'huile	

Propriété	Lieu d'essai S- site L - laboratoire	Catégorie de matériel	Fréquence des essais	Limites recommandées nécessitant une intervention	Intervention	Notes
Teneur en eau	L	O, A, B, C, D, E	<p>O, A Après remplissage ou re-remplissage et avant mise sous tension . Ensuite après 3 et 12 mois puis en même temps que l'analyse des gaz dissous.</p> <p>B, D, E Après remplissage ou re-remplissage et avant mise sous tension. Ensuite après 12 mois puis tous les six ans avec l'analyse des gaz dissous.</p> <p>C Pas un essai de routine seulement puisque la tension de claquage est voisine de la valeur limite.</p>	<p>O, A, D < 20 mg/kg B < 40 mg/kg E < 30 mg/kg C pas d'eau libre à température ambiante.</p>	Vérifier l'origine de l'humidité envisager de retraiter.	Les valeurs citées ne s'appliquent que lorsque l'acidité ne dépasse pas 0.1mgKOH/g d'huile CEI 296
Indice neutralisation	L	O, A, B, C, D, E, F, G	O, A, B, C tous les six ans. D, E, F, G pas un essai de routine	O, A, B, C, D, E max 0.5 mgKOH/g	Remplacer ou régénérer l'huile	Lorsque l'indice de neutralisation est > 0.3 mg KOH/g, effectuer les essais plus souvent.

Propriété	Lieu d'essai S- site L - laboratoire	Catégorie de matériel	Fréquence des essais	Limites recommandées nécessitant une intervention	Intervention	Notes
Sédiments et dépôts	L	O, A, B, C, D, E	Pas un essai de routine à effectuer suivant les résultats de l'examen visuel ou la valeur de l'indice de neutralisation.	Des sédiments ou des dépôts précipitables ne doivent en principe pas être détectés. Les résultats < 0.02% en masse peuvent être négligés	Lorsque des sédiments sont détectés, il faut retraiter l'huile en variante. Remplacer si ce là est plus économique ou si les résultats d'autres essais le confirment. Lorsque des dépôts précipitables sont détectés, il convient de remplacer ou de régénérer l'huile.	
Résistivité	L	O, A, B, C, D, F	O, A, B, D Après remplissage ou re-remplissage et avant mise sous tension. Ensuite après 12 mois puis tous les six ans. C, E Pas un essai de routine.	à 20°C O A D min 200GΩm B C E min 60GΩm A 190°C O A D min 16 GΩm B C min 0.2 GΩm E min 0.7 GΩm	En chercher l'origine	
Facteur de dissipation diélectrique tg δ à 90°C et 40- 60 HZ	L	O, A, B, C, D, E	O, A, B, D Après remplissage ou re-remplissage et avant mise sous tension. Ensuite après 12 mois puis tous les six ans. C, E Pas un essai de routine	O, A, D: max 0.2 B, C: max 1.0 E: max 0.3	En chercher l'origine.	Se conformer aux instructions du constructeur si une autre fréquence et d'autres limites sont proposées

Propriété	Lieu d'essai S- site L - laboratoire	Catégorie de matériel	Fréquence des essais	Limites recommandées nécessitant une intervention	Intervention	Notes
Tension interfaciale	L	O, A, B, C, D, E	O, A, B, C, D, E Après remplissage ou re-remplissage et avant mise sous tension . Ensuite après 12 mois puis tous les six ans.	O, A, B, C, D, E Min 15 mN/m	Étudier	Pas un essai de routine
Teneur en gaz	L	O A B D				Se conformer aux instructions du constructeur
Point d'éclair	L	O, A, B, C, D, E	O, A, B, C, D, E Pas un essai de routine peut être requis lorsqu'une odeur anormale est décelée à la suite d'un défaut interne ou lorsque le transformateur vient d'être rempli	O, A, B, C, D, E Abaissement maximal 15°C	Remplacer l'huile. Une inspection du matériel peut être nécessaire	

CATÉGORIES DE MATÉRIELS

Le matériel électrique est classé en différentes catégories comme suit :

Catégorie O

Transformateurs de puissance de tension de réseau maximale supérieure à 420kV.

Catégorie A

Transformateurs de puissance de tension de réseau maximale supérieure à 170kV et inférieure ou égale à 420kV. Sont également inclus les transformateurs de puissance, quelle que soit la tension assignée, dont la continuité de service est essentielle et le matériel similaire d'applications particulières travaillant dans des conditions de coûts élevés.

Catégorie B

Transformateurs de puissance de tension de réseau maximale supérieure à 72,5 kV et inférieure ou égale à 170kV (autre que ceux appartenant à la catégorie A).

Catégorie C

Transformateurs de puissance de tension de réseau maximale inférieure ou égale à 72,5 kV (autres que ceux appartenant à la catégorie A). interrupteurs remplis d'huile, sélecteurs et régulateurs, en courant alternatif, à enveloppe métallique.

Catégorie D

Réducteurs de mesure de tension de réseau maximale supérieure à 170kV.

Catégorie E

Réducteurs de mesure de tension de réseau maximale inférieure ou égale à 170kV.

Catégorie F

Cuve du rupteur des changeurs de prise en charge.

Catégorie G

Disjoncteur à bain d'huile.

Notes

Les petits transformateurs jusqu'à 1MVA et 36kV ne sont pas compris dans cette classification. Un programme de contrôle est considéré comme n'étant pas économique pour ce type de matériel. Lorsqu'un programme de contrôle est exigé pour ces transformateurs, il convient que les directives données dans la catégorie C soient suivies.

ملخص

في الجزء الأول من هذا العمل قمنا بإنشاء برامج تستخدم الشبكات العصبية الاصطناعية الموجهة و غير الموجهة في تشخيص و تصنيف زيت المحولات المستخدمة من طرف سونالغاز و المسماة بوراك22 وذلك باستعمال التحاليل الفيزيائية -الكيميائية. قمنا بإنجاز أربع شبكات عصبية لمساعدة المهندس المحلل ذو الخبرة القليلة لاتخاذ قرارات صائبة و تتمثل هذه الشبكات في: شبكة الانتقال الارتدادي بمعامل وثب , شبكة الانتقال الارتدادي بمعامل وثب و تكييف نسبة التعلم , شبكة ليفنبارغ - ماركرد و شبكة التعديل البايزية. بالنسبة لموضوع التصنيف , فلقد استخدمنا الشبكات التنافسية و خريطة كوهونان ذاتية التنظيم و أيضا شبكات التعديل البايزية. في الجزء الثاني أنشأنا شبكات موجهة و غير موجهة أيضا للتنبؤ بالشيخوخة الحرارية للعازل PVC المصنوع ب'كابال' و للبوراك22 و أيضا للتنبؤ بالضغط التمزقي لمجال الهواء المتكون من النظام رأس حاجز سطح. غالبا ما تكون معرفة التشخيص للعازل غير كافية , فعلى المهندس أن تكون لديه فكرة عن الوقت المتبقي للتجهيز قبل توقفه عن العمل و إذا كان هذا التقييم دقيقا يستطيع أن يبدأ برنامج الصيانة التي ستوفر كثيرا من الأموال. من أجل هذا أنشأنا شبكات ذات أساس قطري و ذات دالة غوصية : الأولى ممرنة بطريقة الانتقال الارتدادي و الثانية بطريقة التحسين العشوائي (ط.ت.ع) و الثالثة هي طريقة استحدثناها و تتمثل في قرن طريقة (ط.ت.ع) بطريقة متوسط التصنيفي و التي تعرض المراكز الغوصية في بعدين.

كلمات مفتاحية: الشبكات العصبية الاصطناعية، تشخيص ، تصنيف، تنبؤ، العوازل المستخدمة في أنظمة الضغط العالي

Abstract

In a first part of this work, supervised and unsupervised neural networks have been applied in the diagnosis and classification of transformer oil used by SONELGAZ and named BORAK22 using physico-chemical tests. To help young inexperienced analysts making good diagnosis, we have applied four neural networks: Back-Propagation (BP) with Momentum, Back-Propagation with Momentum and Adaptive learning rate and Levenberg-Marquardt (L.M) method and a regularization Bayesian network. The optimization of these nets offers the best performances to the Bayesian and L.M nets. Competitive layer and Kohonen's self-organization maps and a Bayesian net have been used in the oil classification. In a second part, supervised and unsupervised neural networks have been also applied in the prediction of PVC (manufactured at CABEL) and BORAK22 thermal ageing, and of breakdown voltage of point-barrier-plane air gap. Knowing the diagnosis is not enough for most cases, a field engineer must have an idea of how far away the equipment is from the potential outage. If the estimation can be made very clearly, he can start planning maintenance and system solution in advance. A large amount of money can be saved if he takes proper actions. For this purpose, some Radial Basis Function Gaussian (RBF) networks have been used: an RBF trained by BP, an RBF trained by the Random Optimization Method (ROM) and an original method combining supervised and unsupervised algorithms. This network is an RBF trained by ROM combined with the K-means clustering method which presents the Gaussian centres in two dimensions.

Keywords: artificial neural networks, diagnosis, classification, prediction, high voltage insulation

Résumé

Dans une première partie de ce travail, des réseaux de neurones supervisés et non supervisés sont appliqués au diagnostic et à la classification de l'huile de transformateur utilisée par SONELGAZ et nommée BORAK22 en utilisant les tests physico-chimiques. Pour aider les jeunes ingénieurs analystes à prendre de bonnes décisions, quatre réseaux de neurones (RN) sont appliqués: la Back-Propagation (BP) avec Momentum, BP avec Momentum et à adaptation du taux d'apprentissage, méthode de Levenberg-Marquardt (L.M) et les réseaux de régularisation bayésienne. L'optimisation de tous ces réseaux a offert les meilleures performances aux réseaux bayésiens et L.M. Les réseaux compétitifs, les cartes auto organisées de Kohonen et les réseaux à régularisation bayésienne sont utilisés pour la classification de l'huile de transformateurs. Dans une deuxième partie, des réseaux supervisés et non supervisés sont aussi appliqués dans la prédiction du vieillissement thermique du PVC (fabriqué à CABEL) et de la BORAK22 et de la tension disruptive des intervalles d'air pointe-barrière-plan. En général, la connaissance du diagnostic n'est pas suffisante; un ingénieur de terrain doit avoir une idée sur le temps restant du bon fonctionnement de l'équipement. Si l'estimation est assez précise, il peut commencer à faire un programme de maintenance et avec des actions correctes, beaucoup d'argent pourra être épargné. Dans ce but nous avons utilisé des réseaux à base radiale et à fonction gaussienne (RBF): un RBF entraîné par la BP, un RBF entraîné par la méthode d'optimisation aléatoire (ROM) et une méthode originale combinant des algorithmes supervisés et non supervisés. Ce réseau est un RBF entraîné par la ROM couplée avec la méthode des K-means clustering qui présente les centres gaussiens en deux dimensions.

Mots clés : réseaux de neurones artificiels, diagnostic, classification, prédiction, isolations de haute tension

ملخص

في الجزء الأول من هذا العمل قمنا بإنشاء برامج تستخدم الشبكات العصبية الاصطناعية الموجهة و غير الموجهة في تشخيص و تصنيف زيت المحولات المستخدمة من طرف سونالغاز و المسماة بـ BORA22 وذلك باستعمال التحليل الفيزيائية -الكيميائية. قمنا بإنجاز أربع شبكات عصبية لمساعدة المهندس المحلل ذو الخبرة القليلة لاتخاذ قرارات صائبة و تتمثل هذه الشبكات في: شبكة الانتقال الارتدادي بمعامل وثب , شبكة الانتقال الارتدادي بمعامل وثب و تكييف نسبة التعلم , شبكة ليفنبارغ - ماركرد و شبكة التعديل البايزية. بالنسبة لموضوع التصنيف, فلقد استخدمنا الشبكات التنافسية و خريطة كوهونان ذاتية التنظيم و أيضا شبكات التعديل البايزية. في الجزء الثاني أنشأنا شبكات موجهة و غير موجهة أيضا للتنبؤ بالشيخوخة الحرارية للعازل PVC المصنوع بـ 'كابال' و للبوراك22 و أيضا للتنبؤ بالضغط التمزقي لمجال الهواء المتكون من النظام زاس-حاجز سطح. غالبا ما تكون معرفة التشخيص للعازل غير كافية , فعلى المهندس أن تكون لديه فكرة عن الوقت المتبقي للتجهيز قبل ثوقفه عن العمل و إذا كان هذا التقييم دقيقا يستطيع أن يبدأ برنامج الصيانة التي ستوفر كثيرا من الأموال. من أجل هذا أنشأنا شبكات ذات أساس قطري و ذات دالة غوسية : الأولى ممرنة بطريقة الانتقال الارتدادي و الثانية بطريقة التحسين العشوائي (ط.ت.ع) و الثالثة هي طريقة استحداثها و تتمثل في قرن طريقة (ط.ت.ع) بطريقة متوسط التصنيفي و التي تعرض المراكز الغوسية في بعدين.

Abstract

In a first part of this work, supervised and unsupervised neural networks have been applied in the diagnosis and classification of transformer oil used by SONELGAZ and named BORA22 using physico-chemical tests. To help young inexperienced analysts making good diagnosis, we have applied four neural networks: Back-Propagation (BP) with Momentum, Back-Propagation with Momentum and Adaptive learning rate and Levenberg-Marquardt (L.M) method and a regularization Bayesian network. The optimization of these nets offers the best performances to the Bayesian and L.M nets. Competitive layer and Kohonen's self-organization maps and a Bayesian net have been used in the oil classification. In a second part, supervised and unsupervised neural networks have been also applied in the prediction of PVC (manufactured at CABEL) and BORA22 thermal ageing, and of breakdown voltage of point-barrier-plane air gap. Knowing the diagnosis is not enough for most cases, a field engineer must have an idea of how far away the equipment is from the potential outage. If the estimation can be made very clearly, he can start planning maintenance and system solution in advance. A large amount of money can be saved if he takes proper actions. For this purpose, some Radial Basis Function Gaussian (RBF) networks have been used: an RBF trained by BP, an RBF trained by the Random Optimization Method (ROM) and an original method combining supervised and unsupervised algorithms. This network is an RBF trained by ROM combined with the K-means clustering method which presents the Gaussian centres in two dimensions.

Résumé

Dans une première partie de ce travail, des réseaux de neurones supervisés et non supervisés sont appliqués au diagnostic et à la classification de l'huile de transformateur utilisée par SONELGAZ et nommée BORA22 en utilisant les tests physico-chimiques. Pour aider les jeunes ingénieurs analystes à prendre de bonnes décisions, quatre réseaux de neurones (RN) sont appliqués: la Back-Propagation (BP) avec Momentum, BP avec Momentum et à adaptation du taux d'apprentissage, méthode de Levenberg-Marquardt (L.M) et les réseaux de régularisation bayésienne. L'optimisation de tous ces réseaux a offert les meilleures performances aux réseaux bayésiens et L.M. Les réseaux compétitifs, les cartes auto organisées de Kohonen et les réseaux à régularisation bayésienne sont utilisés pour la classification de l'huile de transformateurs. Dans une deuxième partie, des réseaux supervisés et non supervisés sont aussi appliqués dans la prédiction du vieillissement thermique du PVC (fabriqué à CABEL) et de la BORA22 et de la tension disruptive des intervalles d'air pointe-barrière-plan. En général, la connaissance du diagnostic n'est pas suffisante; un ingénieur de terrain doit avoir une idée sur le temps restant du bon fonctionnement de l'équipement. Si l'estimation est assez précise, il peut commencer à faire un programme de maintenance et avec des actions correctes, beaucoup d'argent pourra être épargné. Dans ce but nous avons utilisé des réseaux à base radiale et à fonction gaussienne (RBF): un RBF entraîné par la BP, un RBF entraîné par la méthode d'optimisation aléatoire (ROM) et une méthode originale combinant des algorithmes supervisés et non supervisés. Ce réseau est un RBF entraîné par la ROM couplée avec la méthode des K-means clustering qui présente les centres gaussiens en deux dimensions.