



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE



UNIVERSITE M'HAMED BOUGARA BOUMERDES

Faculté des Hydrocarbures et de la Chimie

Département Génie des Procédés Chimiques et Pharmaceutiques

## MEMOIRE

Présenté par

**M<sup>me</sup> BEDDA KAHINA**

En vue de l'obtention du diplôme de

**MAGISTER en GENIE DES PROCEDES**

**OPTION : TECHNOLOGIE DES HYDROCARBURES**

Thème

*Modélisation du procédé d'hydrotraitement du gasoil à  
l'aide des réseaux de neurones artificiels*

Soutenu le 06/05/2008, devant le jury suivant :

<b>M<sup>r</sup> B. Hamada</b>	<b>Pr</b>	<b>UMBB</b>	<b>Président</b>
<b>M<sup>r</sup> K. Daoud</b>	<b>MC</b>	<b>USTHB</b>	<b>Examineur</b>
<b>M<sup>r</sup> A. Maouche</b>	<b>CC</b>	<b>UMBB</b>	<b>Examineur</b>
<b>M<sup>r</sup> N. Bentahar</b>	<b>CC</b>	<b>UMBB</b>	<b>Examineur</b>
<b>M<sup>r</sup> G. Otmanine</b>	<b>M.C</b>	<b>UMBB</b>	<b>Rapporteur</b>

Année universitaire 2007/2008



تيلمع تجرده لزيدلا يه دحاو نم مها تايلمعلا في ريركنلا ، يهف دعاستي لع نيسحت ص ناصخ دوقو لزيدلا لوصحلا ي لع تا فساوما ريباعملاو تبولطملا . نم نمض دئاوف تجذمنه هذه تيلمعلا نهكنلا تدوجب تاجوتتملا بسد تاريختم ليغشتلا (تجرد قرارحلا ، طغضلا ، عرسلا تيمجلا). كانه ددع تاينقت مجذمنلا ، نم لهنيب تقيرط قودنصلا دوسلا هذه ، تقيرطلا فوس مدختست تجذمنه تيلمع تجرده لزيدلا ي لع يوتسم ، ربخملا اذهل ضرغلا متل لامعتسا تقيرط اذلا ي عانطصلا لا ثيد نا جذومنلا لصحتما هيلع وه تكبش تبيصع نم عون تقيرط فقوتلا قبا سلا هناولأ (early stopping) نامضلا ميمعتلا ديجا جذومنلا . جئاتنلا لصحتما هيلع تبيصع م ادجو و اذخر ابتخلا نالقا نم 7٪ .

تاملكلا تبيصيرلا : تجرد هلا ، لزيدلا ، تجذمنلا ، تاكبشلا تبيصعلا تيعانطصلا .

## Résumé

Le procédé d'hydrotraitement du gasoil est l'un des procédés les plus importants en raffinage, il permet l'amélioration des caractéristiques du carburant diesel pour atteindre les spécifications requises et les normes de pollution. La modélisation de ce procédé permettrait, entre autres, de prévoir la qualité des produits en fonction des différentes variables opératoires (température, pression, vitesse spatiale). Il existe plusieurs techniques de modélisation, parmi elles la modélisation de type boîte noire ; cette méthode sera utilisée pour modéliser le procédé d'hydrotraitement à l'échelle pilote. L'approche adoptée pour ce travail est une approche par intelligence artificielle. Le modèle élaboré est un réseau de neurones de type Multi Layer Perceptron (MLP). L'apprentissage du réseau s'effectue suivant l'algorithme de rétropropagation du gradient avec momentum, la technique de l'arrêt prématuré (*early stopping*) a été utilisée pour éviter le phénomène de surajustement et garantir ainsi une bonne généralisation du modèle. Les résultats obtenus sont très satisfaisants, l'erreur de test est inférieure à 7%.

**Mots clés :** hydrotraitement, gasoil, modélisation, réseaux de neurones artificiels.

## Abstract

Hydrotreating of gasoil is one of the most important processes in the petroleum refining; it helps to improve the characteristics of diesel fuel to achieve the required specifications and standards for pollution. The modeling of this process would, among other things, to provide a quality of products based on the operating variables (temperature, pressure, velocity space). There are several modeling techniques, among them the modeling black box type, this method will be used to model the hydrotreating process at pilot scale. The approach to this work is an artificial intelligence approach. The model developed is a neural network type Multi Layer Perceptron (MLP). The learning network is carried out according to the algorithm of backpropagation gradient with momentum, early stopping technique has been used to prevent the phenomenon of overfitting and thereby guarantee a good generalization of the model. The results were very satisfactory, the error test is less than 7%.

**Keywords:** hydrotreating, gasoil, modeling, artificial neural networks.

## ***Remerciements***

Au terme de ce modeste travail je tiens à remercier en premier lieu *M G. Otmanine*, mon directeur de mémoire, pour son aide précieuse, pour ses remarques pertinentes et surtout pour le temps qu'il m'a accordé et ceci malgré ses nombreuses charges.

Mes vifs remerciements vont à *M B. Hamada* pour avoir accepté de présider ce jury.

Je tiens à remercier *M K. Daoud*, *M A. Maouche* et *M N. Bentahar*, qui ont bien voulu me faire l'honneur de participer à ce jury.

Je remercie également tous les enseignants du département Génie des Procédés de la FHC en particulier *M O. Kaddour* et *M H. Mimoun* pour leur contribution à cette formation.

Ma profonde gratitude va aussi à *M<sup>me</sup> A. Bertal* de la raffinerie de SKIKDA et à *M<sup>lle</sup> S. Kanoune* de la direction générale de NAFTEC pour leur aide précieuse.

Merci à tous ceux qui ont contribué de près ou de loin à l'élaboration de ce mémoire.

*Kahina.*

*Dédicaces*

*À l'âme de mon père,  
À ma maman, à qui je dois tout,  
À mon époux Mohamed et à mon frère Lyes,  
À tous ceux qui me sont chers.*

*Kahina*

# ***SOMMAIRE***

Remerciements

Dédicaces

Sommaire

Notations et abréviations

Liste des tableaux

Liste des figures

**Introduction générale.....01**

## **Chapitre I : Le procédé d'hydrotraitement du gasoil**

I.1 Introduction.....02

I.2 Objectifs du procédé.....03

I.3 Caractéristiques générales.....13

    I.3.1 Les réactions de l'hydrotraitement.....13

        I.3.1.1 Caractéristiques thermodynamiques.....13

        I.3.1.2 Cinétique des réactions.....14

    I.3.2 Les catalyseurs.....14

I.4 Variables opératoires.....16

    I.4.1 La pression partielle d'hydrogène.....17

    I.4.2 La température moyenne pondérée du lit catalytique.....18

    I.4.3 La vitesse spatiale horaire.....19

    I.4.4 Taux d'hydrogène ou taux de recyclage.....21

    I.4.5 Concentration en hydrogène sulfuré dans le gaz de recyclage.....21

I.5 Technologies du procédé.....23

    I.5.1 Le réacteur.....23

    I.5.2 Schémas du procédé.....24

        I.5.2.1 Le procédé d'hydrotraitement de Chevron Lummus Global LLC.....24

        I.5.2.2 Le procédé d'hydrotraitement de Axens.....24

        I.5.2.3 Le procédé ULSD de Haldor Topsøe.....25

        I.5.2.4 Le procédé HDS/HDA de Haldor Topsøe.....26

I.6 Conclusion.....27

## **Chapitre II : La modélisation en génie des procédés**

II.1 Introduction.....	28
II.2 Les usages de la modélisation dans le génie des procédés.....	29
II.3 Modèle de connaissance ou modèle « boîte blanche ».....	30
II.4 Modèle de conduite ou modèle « boîte noire ».....	31
II.4.1 Approche mathématique.....	31
II.4.2 Approche par intelligence artificielle.....	32
II.5 Modèle semi-physique ou modèle « boîte grise ».....	32
II.6 Conclusion.....	32

## **Chapitre III : Les réseaux de neurones artificiels et la modélisation de type boîte noire**

III.1 Introduction.....	33
III.2 Historique.....	33
III.3 Définitions.....	35
III.3.1 Le neurone biologique.....	35
III.3.2 Le neurone artificiel.....	36
III.3.3 Les réseaux de neurones.....	38
III.4 Applications.....	38
III.5 Classification des réseaux de neurones.....	38
III.5.1 Les réseaux de neurones non bouclés.....	38
III.5.2 Les réseaux de neurones bouclés (ou récurrents).....	39
III.6 L'apprentissage des réseaux de neurones.....	41
III.6.1 Définition.....	41
III.6.2 L'apprentissage supervisé.....	42
III.6.3 L'apprentissage non supervisé.....	42
III.7 Propriété fondamentale des réseaux de neurones à apprentissage supervisé :	
L'approximation universelle parcimonieuse.....	43
III.7.1 Réseaux non bouclés.....	43
III.7.2 Réseaux bouclés.....	44
III.8 Quand utiliser les réseaux de neurones à apprentissage supervisé ?.....	44
III.9 Modélisation statique de type boîte noire à l'aide des réseaux de neurones.....	45
III.9.1 Formulation du problème.....	45
III.9.2 Les entrées pertinentes.....	46
III.9.3 La collecte des données.....	46
III.9.4 Le nombre de neurones cachés.....	46

III.9.5 Apprentissage du réseau.....	48
III.9.5.1 Normalisation des entrées et des sorties.....	49
III.9.5.2 Evaluation du gradient de la fonction de coût par rétropropagation.....	50
III.9.5.3 Modification des paramètres en fonction du gradient de la fonction de coût.....	52
a) Initialisation des paramètres.....	52
b) La méthode du gradient simple.....	52
c) Les méthodes du gradient du second ordre.....	53
d) Choix de la méthode.....	53
e) L'algorithme de Levenberg-Marquardt.....	54
III.9.5.4 Le Momentum.....	55
III 9.5.5 Arrêt prématuré.....	55
III.10 Conclusion.....	56

#### **Chapitre IV : Application des réseaux de neurones pour la modélisation du procédé d'hydrotraitement du gasoil**

IV.1 Introduction.....	57
IV.2 Présentation des données.....	57
IV.2.1 Déroulement des expériences.....	57
IV.2.2 Résultats expérimentaux et commentaires.....	59
IV.2.3 Normalisation des données.....	62
IV.3 Détermination de la structure optimale du réseau.....	64
IV.3.1 Démarche à suivre.....	64
IV.3.2 Résultats et commentaires.....	65
IV.4 Evaluation des performances du modèle.....	69
IV.4.1 Dénormalisation des valeurs de la sortie du réseau.....	69
IV.4.2 Calcul d'erreurs.....	70
IV.5 Conclusion.....	72
<b>Conclusion et perspectives.....</b>	<b>73</b>

Références bibliographiques

Annexe : Evolutions des EQMA, EQMV et EQMT au cours des processus d'apprentissage des différents réseaux de neurones testés.

## *Notations*

$g(x, w)$	Approximation de la fonction de régression
$K_{HC}$	Constante d'adsorption de l'hydrocarbure
$K_i$	Constante d'adsorption du composé inhibiteur $i$
$k$	Constante de vitesse de la réaction
$\lambda$	Constante du momentum
$J_k(w)$	Coût partiel relatif à l'observation $k$
$r$	Facteur d'échelle
$f(i_v)$	Fonction d'activation du neurone $i$
$J(w)$	Fonction de coût des moindres carrées
$f(P_{H_2})$	Fonction plus ou moins complexe de la pression partielle d'hydrogène
$th$	Fonction sigmoïde de type tangente hyperbolique
$\nabla J$	Gradient de la fonction de coût des moindres carrées
$\delta_k$	Gradient du coût partiel par rapport au potentiel du neurone $i$
$z_p$	Grandeur relative au processus qu'on cherche à modéliser
$H(w)$	Matrice hessienne de la fonction de coût
$\mu_i$	Pas du gradient ou pas d'apprentissage
$v_i$	Potentiel du neurone $i$
$P_{HC}$	Pression partielle de l'hydrocarbure
$P_{H_2}$	Pression partielle d'hydrogène
$P_i$	Pression partielle du composé inhibiteur $i$
$q^{-l}$	Retard
$t_s$	Température à la sortie du réacteur
$t_e$	Température à l'entrée du réacteur
$Tm_{DDC}$	Température moyenne pondérée de début de cycle
$Tm_{DFC}$	Température moyenne pondérée de fin de cycle
$Tm$	Température moyenne pondérée du lit catalytique
$t$	Temps
$S$	Valeur mesurée de la teneur en soufre du gasoil hydrotraité
$S^*$	Valeur prédite par le modèle de la teneur en soufre du gasoil hydrotraité
$\{u\}$	Vecteur des entrées d'un système dynamique
$\{w\}$	Vecteur des paramètres (poids et biais)

$\{p\}$	Vecteur des perturbations
$\{y\}$	Vecteur des sorties du système
$\{x\}$	Vecteur des variables d'entrées pour un système statique ou vecteur des variables d'état pour un système dynamique
$v$	Vitesse de la réaction
VVH	Vitesse spatiale horaire
$V$	Volume total de catalyseur

### *Abréviations*

ADN	Acide désoxyribonucléique
BFGS	Broyden-Fletcher-Goldfarb-Shanoo
dB	Décibel
Dist	Distillation
EPA	Environmental Protection Agency (agence de protection de l'environnement)
EQMA	Erreur quadratique moyenne commise sur l'ensemble d'apprentissage
EQMT	Erreur quadratique moyenne commise sur l'ensemble de test
EQMV	Erreur quadratique moyenne commise sur l'ensemble de validation
HC	Hydrocarbures
HAP	Hydrocarbures aromatiques polycycliques
HDC	Hydrocraquage
HDA	Hydrodésaromatisation
HDN	Hydrodézazotation
HDO	Hydrodéoxygénation
HDS	Hydrodésulfuration
HYD	Hydrogénation
IC	Indice de cétane
LM	Levenberg-Marquardt
LCO	Light cycle oil (gasoil de craquage catalytique)
MLP	Multi-Layer Perceptron
OMS	Organisation Mondiale de la Santé
RBF	Radial Basic Function
SR	Straight run (gasoil de distillation directe)
ULSD	Ultra Low Sulfure Diesel

## *Liste des tableaux*

Tableau I.1 : Evolution des spécifications du gasoil en Europe et aux Etats-Unis.....	02
Tableau I.2 : Hydrotraitement des gasoils: Propriétés des charges et des produits.....	04
Tableau I.3 : Stabilité à l'oxydation de gasoils hydrotraités (test de stabilité: norme ASTM D 4625).....	07
Tableau I.4 : Influence de l'hydrotraitement sur les caractéristiques d'un gasoil.....	08
Tableau I.5 : Evolution des normes européennes pour les émissions de polluants appliquées aux véhicules diesel neufs.....	11
Tableau I.6 : Chaleurs de réaction de divers composés purs traités par hydrogène.....	13
Tableau I.7 : Classification des catalyseurs industriels du procédé d'hydrotraitement suivant leur activité.....	16
Tableau I.8 : Influence de la pression partielle d'hydrogène sur l'hydrotraitement d'un gasoil de distillation primaire contenant 26,7% masse d'aromatiques et 1,3% masse de soufre.....	17
Tableau I.9: Caractéristiques d'un gasoil de distillation directe avant et après le procédé d'hydrotraitement HDS/HDA de Haldor Topsøe.....	26
Tableau IV.1 : Caractéristiques des charges.....	59
Tableau IV.2 : Composition et caractéristiques d'un catalyseur commercial sulfurisé de type CoMo/alumine.....	59
Tableau IV.3 : Caractéristiques et résultats des expériences d'hydrotraitement de différents gasoils de distillation atmosphérique.....	60
Tableau IV.4 : Données normalisées du procédé.....	63
Tableau IV.5 : Valeurs des EQMA, EQMV et EQMT du réseau à un neurone caché pour différents essais d'initialisation de paramètres.....	65
Tableau IV.6 : Valeurs des EQMA, EQMV et EQMT du réseau à deux neurones cachés pour différents essais d'initialisation de paramètres.....	65
Tableau IV.7 : Valeurs des EQMA, EQMV et EQMT du réseau à trois neurones cachés pour différents essais d'initialisation de paramètres.....	66
Tableau IV.8 : Valeurs des EQMA, EQMV et EQMT du réseau à quatre neurones cachés pour différents essais d'initialisation de paramètres.....	66
Tableau IV.9 : Valeurs des EQMA, EQMV et EQMT du réseau à cinq neurones cachés pour différents essais d'initialisation de paramètres.....	66
Tableau IV.10 : Valeurs des EQMA, EQMV et EQMT du réseau à six neurones cachés pour différents essais d'initialisation de paramètres.....	66
Tableau IV.11 : Valeurs des EQMA, EQMV et EQMT du réseau à sept neurones cachés pour différents essais d'initialisation de paramètres.....	67

Tableau IV.12 : Valeurs des EQMA, EQMV et EQMT du réseau à huit neurones cachés pour différents essais d'initialisation de paramètres.....	67
Tableau IV.13 : Valeurs des EQMA, EQMV et EQMT du réseau à neuf neurones cachés pour différents essais d'initialisation de paramètres.....	67
Tableau IV.14 : Valeurs des EQMA, EQMV et EQMT du réseau à dix neurones cachés pour différents essais d'initialisation de paramètres.....	67
Tableau IV.15 : Résultats de calcul de l'erreur relative pour chaque exemple.....	71

## *Liste des figures*

Figure I.1 : Les différents gasoils à hydrotraiter.....	03
Figure I.2 : Exemple d'émissions de particules d'un moteur diesel pour différentes teneurs en soufre du gasoil.....	05
Figure I.3 : Efficacité du pot catalytique en fonction de la teneur en soufre du gasoil et de la température à l'entrée.....	06
Figure I.4 : Influence de l'ajout d'une fraction riche en composés diaromatiques ou triaromatiques sur les émissions de HAP d'un moteur diesel.....	08
Figure I.5 : Influence de l'indice de cétane sur le délai d'auto-inflammation.....	09
Figure I.6 : Relation entre le niveau de bruit d'un moteur à différents régimes et l'indice de cétane du carburant.....	10
Figure I.7 : Influence de l'indice de cétane du gasoil sur les émissions de fumées, au démarrage à froid.....	10
Figure I.8 : Influence de l'hydrotraitement du gasoil sur les émissions à l'échappement d'une voiture particulière diesel.....	12
Figure I.9 : Limitation thermodynamique de l'hydrogénation des aromatiques dans un gasoil.....	14
Figure I.10 : Influence de la Tm d'hydrotraitement sur la composition d'un gasoil pour différents rapports hydrogène/hydrocarbure.....	19
Figure I.11 : Influence de la VVH du procédé d'hydrotraitement sur la composition d'un gasoil pour différents rapports hydrogène/hydrocarbure.....	20
Figure I.12 : Effet de la concentration de H <sub>2</sub> S dans le gaz sur la composition d'un gasoil hydrotraité.....	22
Figure I.13 : Réacteur du procédé d'hydrotraitement du gasoil.....	23
Figure I.14 : Schéma du procédé d'hydrotraitement de Chevron Lummus Global LLC.....	24
Figure I.15 : Schéma du procédé d'hydrotraitement de Axens.....	25
Figure I.16 : Schéma du procédé ULSD de Haldor Topsøe.....	25
Figure I.17 : Schéma du procédé HDS/HDA de Haldor Topsøe.....	26
Figure II.1 : Nomenclature décrivant la structure d'un procédé.....	28
Figure III.1 : Le neurone biologique.....	35
Figure III.2 : Schéma du fonctionnement d'un neurone biologique.....	36
Figure III.3 : Représentation du neurone artificiel.....	37
Figure III.4 : Représentation d'un réseau de neurones non bouclé.....	39
Figure III.5 : Représentation d'un réseau de neurones bouclé.....	40
Figure III.6 : Forme canonique du réseau de neurones représenté sur la figure III.5.....	41
Figure III.7 : Exemple d'un surapprentissage d'un réseau de neurones.....	47

Figure IV.1 : Exemple d'une unité pilote d'hydrotraitement.....	58
Figure IV.2 : Influence de la température d'hydrotraitement sur la teneur en soufre du produit (P=800 psi, VVH=2,6 h )..... <sup>-1</sup>	61
Figure IV.3 : Influence de la pression d'hydrotraitement sur la teneur en soufre du produit (T=360 °C, VVH=2,8 h )..... <sup>-1</sup>	61
Figure IV.4 : Influence de la vitesse spatiale horaire d'hydrotraitement sur la teneur en soufre du produit (T=360 °C, P=800 psi).....	61
Figure IV.5 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à un neurone caché, essai N°1.....	68
Figure IV.6 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à sept neurones cachés, essai N°5.....	69
Figure IV.7 Les teneurs en soufre prédites en fonction des teneurs en soufre mesurées des gasoils hydrotraités.....	70

---

## ***INTRODUCTION GENERALE***

---

## **Introduction générale**

Le procédé d'hydrotraitement du gasoil est un procédé très important dans le domaine du raffinage. Il permet l'obtention d'un carburant diesel de haute qualité, répondant aux exigences du marché. Ce procédé a connu, ces dernières années, un développement considérable, dû essentiellement à l'évolution des normes de pollution. En effet ces normes deviennent de plus en plus sévères particulièrement en Europe et aux Etats-Unis.

La modélisation d'un tel procédé permettrait de mieux le comprendre et surtout de mieux le contrôler. En effet, il existe dans la littérature plusieurs modèles mathématiques du procédé d'hydrotraitement du gasoil ; chaque modèle est spécifique au gasoil pour lequel il a été établi. Ces modèles sont conçus à partir des lois fondamentales du génie des procédés, ils sont dits modèles de type boîte blanche.

Dans le contexte de notre travail, nous envisageons de développer un modèle plus général du procédé d'hydrotraitement du gasoil, capable de prédire les caractéristiques des produits pour différents types de gasoils et sous différentes conditions opératoires. Pour cela, nous allons utiliser une technique de modélisation complètement opposée à la première, appelée modélisation de type boîte noire. La modélisation à l'aide des réseaux de neurones artificiels semble la mieux adaptée pour la réalisation de ce travail.

A cet effet, nous avons divisé ce mémoire en quatre chapitres :

- Dans le premier chapitre nous présentons une description détaillée du procédé d'hydrotraitement du gasoil, à savoir la thermodynamique et la cinétique des réactions, les différents catalyseurs employés, l'effet des variables opératoires sur le procédé, le type de réacteur utilisé ainsi que des exemples de schémas technologiques rencontrés dans l'industrie.
- Le deuxième chapitre donne un aperçu sur les usages de la modélisation dans le domaine du génie des procédés ainsi que les différentes méthodes utilisées pour l'élaboration de ces modèles.
- Le troisième chapitre est consacré aux réseaux de neurones artificiels. Nous allons commencer par définir cette technique ainsi que ses propriétés fondamentales, qui sont l'apprentissage et l'approximation universelle parcimonieuse. La méthodologie de modélisation statique de type boîte noire à l'aide de ces réseaux est ensuite présentée en détail.
- Le chapitre IV est dédié à l'application des réseaux de neurones artificiels de type Multi Layer Perceptron, pour la modélisation du procédé d'hydrotraitement du gasoil à l'échelle pilote. Le modèle établi permet la prédiction de la teneur en soufre du produit en fonction des conditions opératoires, du nombre d'étages du procédé et de la teneur en soufre de la charge.

Une conclusion générale et des perspectives d'amélioration de ce présent travail parachèvent ce mémoire.

## ***CHAPITRE I***

---

### ***LE PROCEDE D'HYDROTRAITEMENT DU GASOIL***

---

## **I.1 Introduction**

Les procédés d'hydrotraitement consistent à traiter sous pression d'hydrogène une coupe pétrolière à l'aide d'un catalyseur solide. Ces procédés concernent toutes les coupes pétrolières d'une raffinerie, de l'essence légère au résidu sous vide [1,2,3].

Les procédés d'hydrotraitement peuvent être divisés en deux principaux groupes [4] :

- Les premiers sont les procédés d'hydropurification ou d'hydrorafinage, qui visent essentiellement à éliminer les hétéroatomes tels que le soufre, l'azote, l'oxygène, le vanadium et le nickel des molécules organiques qui les contiennent, c'est le cas de l'hydrodésulfuration (HDS) pour le soufre, l'hydrodéazotation (HDN) pour l'azote, l'hydrodéoxygénation (HDO) pour l'oxygène et l'hydrodémétallisation pour le nickel, le vanadium, l'arsenic ...etc.
- Les seconds sont les procédés d'hydroconversion qui permettent le changement du squelette et de la structure des molécules organiques, c'est le cas de l'hydrogénation (HYD), de l'hydrodéaromatization (HDA) et de l'hydrocraquage (HDC).

Les conditions opératoires des procédés d'hydrotraitement sont très variables, elles dépendent de la nature de la charge et de l'objectif du procédé : les pressions varient de 10 à plus de 200 bar et les températures de réactions de 250 à 450°C [1].

Les procédés d'hydrotraitement sont de loin les outils les plus utilisés dans les schémas de raffinage des coupes pétrolières. Deux rôles essentiels leurs sont dévolus [1] : d'une part améliorer les caractéristiques des produits finis en terme de spécifications requises pour atteindre la qualité des produits et les normes de pollution et d'autre part, préparer les charges pour d'autres unités de transformation ou de conversion de la raffinerie (isomérisation, reformage, craquage catalytique ...etc.) dont les catalyseurs sont sensibles aux impuretés.

Il apparaît, ces dernières années que les problèmes d'environnement sont devenus une vraie force motrice pour la recherche et le développement des procédés d'hydrotraitement [5,6,7], en particulier pour le gasoil dont les spécifications deviennent de plus en plus sévères comme le montre le tableau I.1. Cette évolution des réglementations incite les raffineurs à améliorer le procédé d'hydrotraitement du gasoil à fin de répondre aux exigences qualitatives du marché.

<b>Caractéristiques du gasoil</b>	<b>Europe (a)</b>			<b>Etats-Unis (b)</b>	
	<b>1996</b>	<b>2000</b>	<b>2005</b>	<b>1993</b>	<b>2006</b>
Indice de cétane	49	51	51-58	40	40
Aromatiques (% vol)				35	35
Polyaromatiques (% poids)	Non spécifié	11	11-2		
Soufre (ppm)	500	350	50-10	500	15

Tableau I.1 : Evolution des spécifications du gasoil en Europe et aux Etats-Unis (a) [5], (b) [6].

## I.2 Objectifs du procédé

Le gasoil à hydrotraiter est obtenu soit, de la distillation primaire du pétrole brut, soit des procédés de conversion (viscoréduction, cokéfaction, craquage catalytique), la figure I.1 rappelle la situation en matière d'indice de cétane, teneurs en soufre et en aromatiques des différents gasoils à hydrotraiter [5].

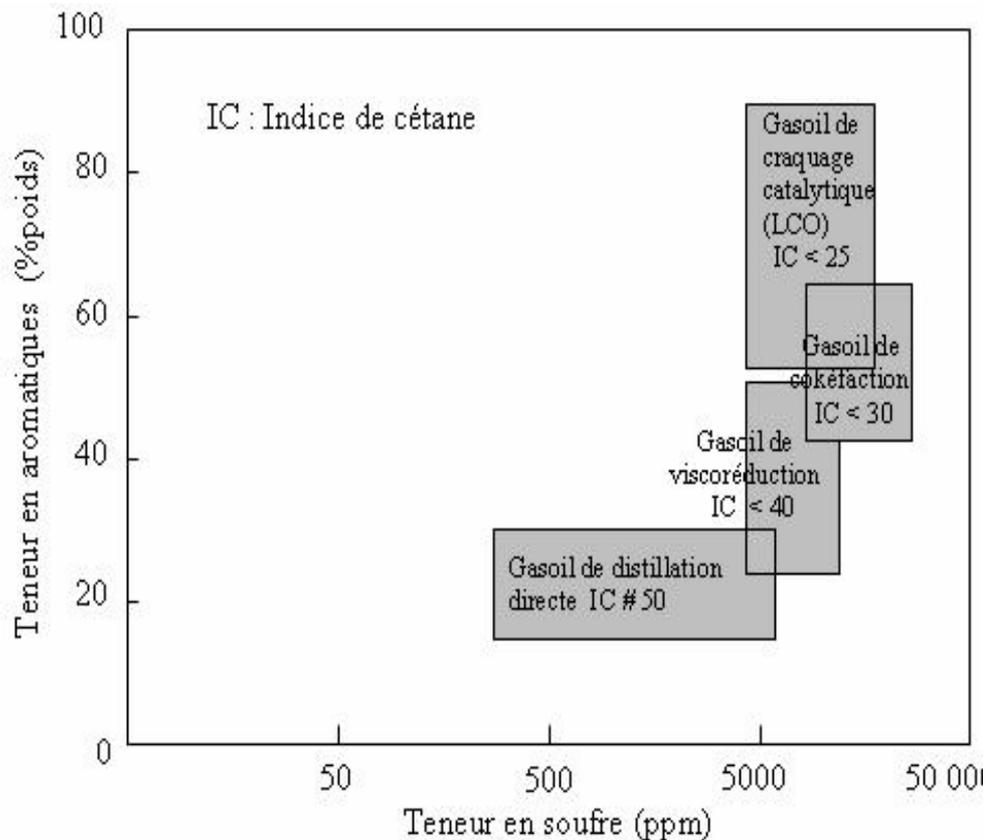


Figure I.1 : Les différents gasoils à hydrotraiter [5].

Le but de l'hydrotraitement de la fraction gasoil est principalement la réduction des teneurs en soufre, azote, aromatiques et oléfines ainsi que l'amélioration de la stabilité et de l'indice de cétane de cette fraction [1]. Le tableau I.2 indique les propriétés des charges et des produits obtenus pour différents cas typiques d'hydrotraitement de gasoils.

	<b>Gasoil de distillation directe</b>	<b>Gasoil + LCO 80% / 20%</b>	<b>LCO</b>
<b>propriétés des charges :</b>			
Densité	0,846	0,863	0,941
Soufre (%masse)	1,31	1,50	2,24
Azote (ppm)	70	245	940
Viscosité à 50°C (mm <sup>2</sup> /s)	3	2,8	2,4
Aromatiques (%masse)	27	37	78
Indice de cétane	53	47	20
Indic de brome (g/100g)	1	4	15
Dist.ASTM D86 (°C) 10% vol	257	253	236
50% vol	294	288	276
90% vol	341	338	346
PF	358	358	374
<b>Rendement après hydrotraitement :</b>			
H <sub>2</sub> S + NH <sub>3</sub> (%masse)	1,34	1,55	2,45
C <sub>1</sub> -C <sub>4</sub> (%masse)	0,30	0,35	0,50
C <sub>5</sub> -150°C (%masse)	3,50	3,60	4,50
150°C + (%masse)	95,26	95,0	93,9
<b>Total</b> (%masse)	<b>100,40</b>	<b>100,50</b>	<b>101,35</b>
<b>Consommation en hydrogène :</b>			
Chimique (%masse)	0,40	0,50	1,35
Pertes (%masse)	0,06	0,06	0,10
<b>Total</b> (%masse)	<b>0,46</b>	<b>0,56</b>	<b>1,45</b>
<b>Propriétés des produits :</b>			
Densité	0,833	0,848	0,898
Soufre (ppm)	500	500	500
Azote (ppm)	45	150	50
Viscosité à 50°C (mm <sup>2</sup> /s)	2,8	2,7	2,1
Aromatiques (%masse)	26	34	71
Indice de cétane	56	51	32
Couleur ASTM	< 0,5	< 1,0	2
Stabilité couleur	Bonne	Bonne	Limite

Tableau I.2 : Hydrotraitement des gasoils: Propriétés des charges et des produits [1].

La teneur en soufre du gasoil fait partout l'objet de réglementations précises. Celles-ci ont pour objectif essentiel la limitation des émissions gazeuses de  $\text{SO}_2$  à l'échappement [8,9]. En effet, le dioxyde de soufre est un polluant majeur. Dans l'atmosphère, il se mélange à la vapeur d'eau produisant l'acide sulfurique. Cette pollution acide peut être transportée par le vent sur plusieurs centaines de kilomètres, et être déposée en tant que pluies acides. Le dioxyde de soufre provoque aussi des maladies respiratoires et peut aggraver les maladies cardio-vasculaires existantes [10].

Par ailleurs la désulfuration du gasoil est très souhaitable pour diminuer les émissions de particules. En effet, une partie du soufre se transforme en  $\text{SO}_3$ , puis en acide sulfurique hydraté sur le filtre destiné à recueillir les particules. Comme l'indique la figure I.2 cette influence perturbatrice est d'autant plus importante que la température des gaz d'échappement, arrivant sur le pot catalytique, est élevée. Ainsi, le traitement catalytique des gaz d'échappement des moteurs diesel ne peut se réaliser qu'avec des gasoils très désulfurés [8,9].

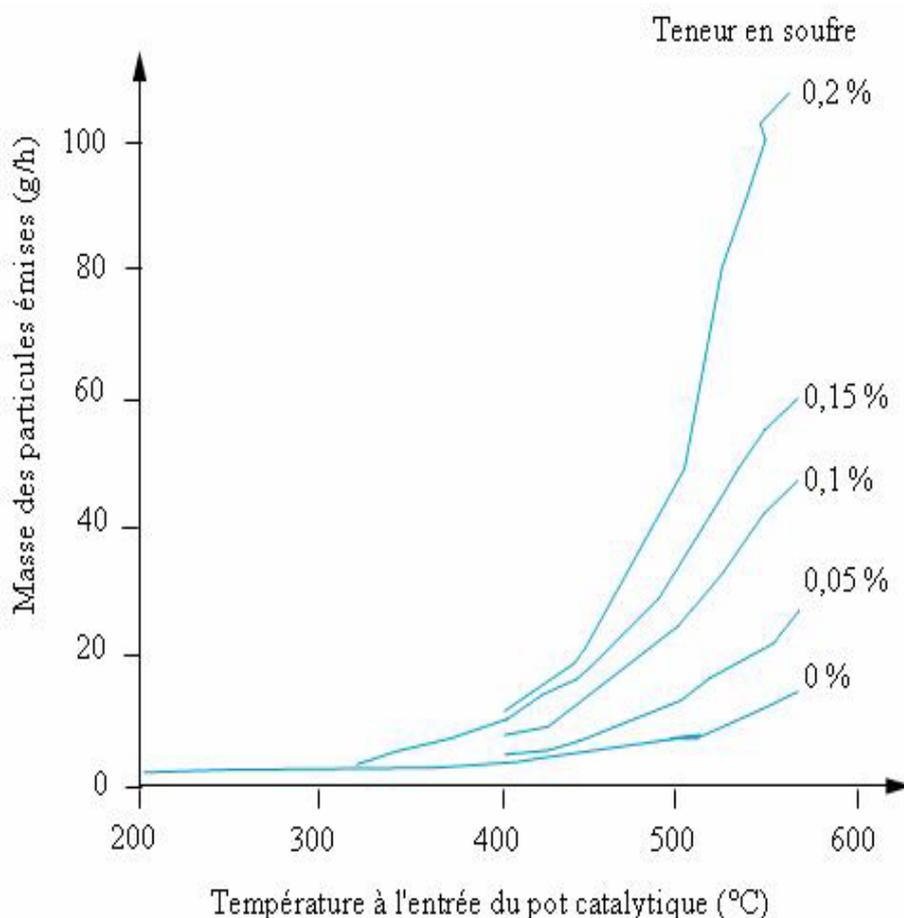


Figure I.2 : Exemple d'émissions de particules d'un moteur diesel pour différentes teneurs en soufre du gasoil [8].

D'autre part, le soufre exerce une action néfaste sur le fonctionnement même du pot catalytique [11]. On voit par exemple sur la figure I.3, que sa température d'amorçage croît avec la teneur en soufre du gasoil, même entre 0,01% et 0,05%. Or dans le moteur diesel, caractérisé par des températures d'échappement relativement faibles, la mise en action du catalyseur constitue un paramètre déterminant. C'est pourquoi on a envisagé une désulfuration du gasoil jusqu'à des teneurs inférieures à 0,05%.

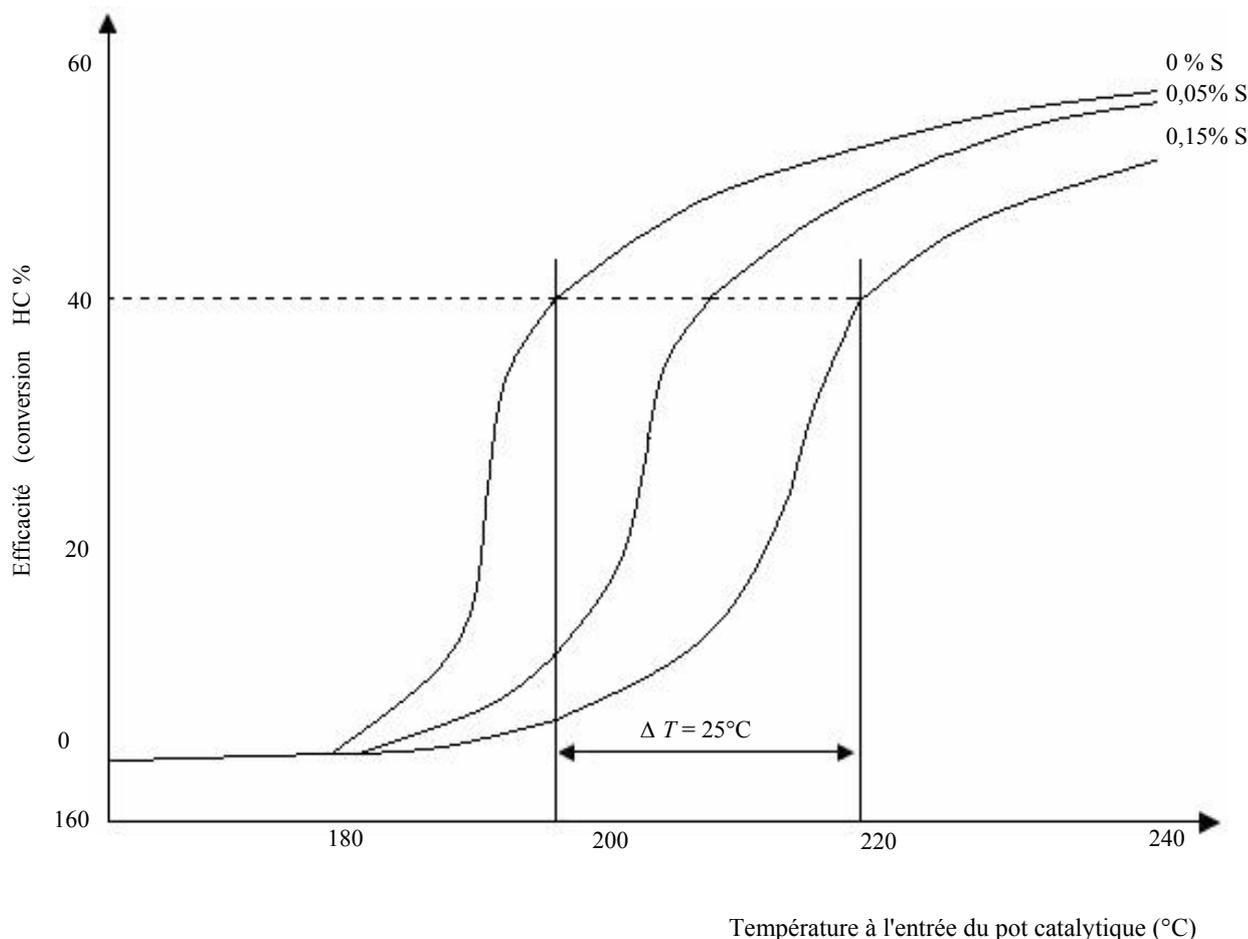


Figure I.3 : Efficacité du pot catalytique en fonction de la teneur en soufre du gasoil et de la température à l'entrée [11].

La réduction du taux de soufre dans le gasoil est considérée comme bénéfique vis-à-vis des risques d'usure corrosive du moteur diesel. D'un autre côté il a été observé pour les gasoils à très faible teneur en soufre (inférieure à 100 ppm) une usure des systèmes d'injection, les incidents se manifestent par des grippages des pompes d'injection provoqués par des usures adhésives-abrasives et survenant après des périodes d'utilisation courtes (15000 Km), ceci peut être expliqué par la réduction du pouvoir lubrifiant due à la disparition du soufre [12].

Entre sa sortie de raffinerie et sa combustion dans le moteur, le gasoil subit un certain nombre d'opérations de stockage et de transport. A l'intérieur même du véhicule, il est soumis à la chaleur, au contact de l'oxygène et de l'humidité de l'air. Ces situations peuvent conduire à une dégradation du carburant entraînant des incidents plus ou moins graves tels que le colmatage des filtres et l'encrassement des injecteurs. On désigne par le terme générique de stabilité, la faculté que doit posséder le gasoil, à se maintenir dans le même état de composition et de caractérisation au cours du temps.

L'hydrotraitement constitue la meilleure technique d'amélioration de la stabilité des gasoils. Le tableau I.3 montre, à ce sujet, le comportement en matière de stabilité de quelques coupes, avant et après hydrotraitement. Cette opération est d'ailleurs absolument nécessaire sur des coupes issues des procédés de conversion thermique (viscoréduction, cokéfaction, craquage), de nature très peu stable. La manifestation la plus nette des améliorations obtenues est la réduction de la teneur en constituants azotés et oléfiniques qui sont respectivement les précurseurs de formation des sédiments et des gommages. L'hydrotraitement de ces coupes s'accompagne évidemment d'une réduction de la teneur en soufre et en oxygène ce qui influe bénéfiquement sur leur stabilité [12].

Caractéristiques	Gasoils											
	LCO				Mélange LCO/SR				Mélange SR/Viscoréduit			
	Réf.*	25 bar	35 bar	45 bar	Réf.*	25 bar	35 bar	45 bar	Réf.*	25 bar	35 bar	45 bar
Masse volumique (Kg/dm <sup>3</sup> )	0,941	0,909	0,905	0,900	0,895	0,873	0,870	0,868	0,847	0,834	0,833	0,832
Rendement HDS (%)		89,7	91,4	93,6		91,6	93,2	95,6		92,0	99,5	96,3
Rendement HDN (%)		58,0	70,0	82,1		48,4	67,7	82,0		46,9	66,0	78,4
<b>Test de stabilité :</b>												
Couleur avant test	2,5	2	1,5	1,5	1,5	1,5	1,5	1,5	2,5	0,5	0,5	0,5
Couleur après test	8	3	2	2	8	2	2	2	3	0,5	0,5	0,5
Sédiments (mg/100 ml)												
Libres	0,9	0,1	0,2	0,2	9,0	0,3	0,2	0,2	0,1	0,1	0,2	0,1
Adhérents (aux parois du récipient)	28,0	0,2	0,3	0,4	7,4	2,3	0,9	0,9	1,7	1,7	1,0	0,4
Totaux	28,9	0,3	0,5	0,6	16,4	2,6	1,1	1,1	1,8	1,8	1,2	0,5

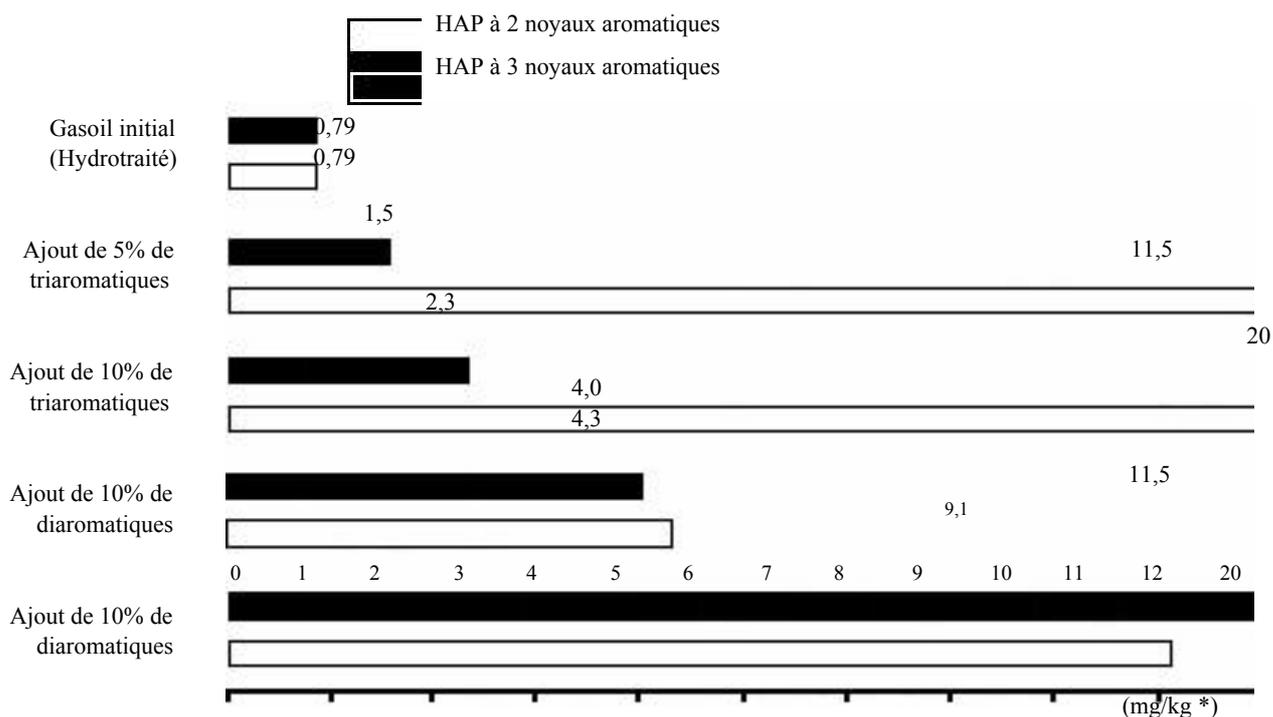
\* Référence sans hydrotraitement.

Tableau I.3 : Stabilité à l'oxydation de gasoils hydrotraités (test de stabilité: norme ASTM D 4625) [12].

Les hydrocarbures aromatiques polycycliques, communément appelés HAP, sont une famille d'hydrocarbures dont la structure comprend au moins deux cycles aromatiques condensés. Depuis de nombreuses années, les HAP sont très étudiés car ce sont des composés présents dans tous les milieux environnementaux et qui montrent une forte toxicité, d'ailleurs c'est l'une des raisons qui a conduit à leur ajout dans la liste des polluants prioritaires par l'agence de protection de l'environnement des Etats-Unis (EPA US Environmental Protection Agency), dès 1976. Aujourd'hui ils font également partie des listes de l'OMS (Organisation Mondiale de la Santé) et de la communauté européenne [13].

La formation des HAP peut avoir de nombreuses origines [14,15], parmi elles la combustion des gasoils dans le moteur diesel. Il est logique de penser que l'existence de constituants aromatiques ou polyaromatiques dans le gasoil, constitue un facteur favorable à la formation des HAP, d'ailleurs plusieurs travaux montrent une liaison nette entre la teneur en di et triaromatiques du gasoil et les émissions de HAP. Les résultats de la figure I.4 montrent, par exemple, que l'adjonction d'une coupe riche, soit en composés diaromatiques, soit en composés triaromatiques, dans le gasoil hydrotraité provoque une augmentation nette des émissions de HAP à deux ou trois noyaux.

Le procédé d'hydrotraitement du gasoil permet de réduire considérablement la teneur en hydrocarbures aromatiques et par conséquent réduire les émissions de HAP. Le tableau I.4 indique les modifications de la composition chimique d'un gasoil résultant de plusieurs hydrotraitements de plus en plus profonds, la teneur en aromatiques passe dans cet exemple de 39,1% à 6,8%.



\* Les émissions sont exprimées en milligrammes par kilogramme de carburant.

Figure I.4 : Influence de l'ajout d'une fraction riche en composés diaromatiques ou triaromatiques sur les émissions de HAP d'un moteur diesel [15].

Caractéristiques du produit	A	A <sup>+</sup>	A <sup>++</sup>	A <sup>+++</sup>	A <sup>++++</sup>
Masse volumique (Kg/dm <sup>3</sup> ) à 15°C	0,862	0,850	0,849	0,838	0,827
Viscosité à 20°C (mm <sup>2</sup> /s)	5,55	5,34	5,22	5,12	4,90
Teneur en soufre (ppm)	11600	640	230	22	4
Teneur en azote (ppm)	216	150	135	17	0,2
Indice de cétane	49,0	50,4	49,0	53,9	60,2
<b>Composition (%masse)</b>					
Paraffines	36,5	36,2	36,8	37,0	41,4
Naphtènes	24,3	24,5	36,5	37,7	51,8
Monoaromatiques	14,2	23,1	21,9	20,2	6,0
Diaromatiques	15,4	12,8	12,6	4,5	0,8
Triaromatiques	1,8	1,0	0,9	0,4	0,0
Composés thiophéniques	7,7	2,4	1,4	0,3	0,0
<b>Total aromatiques</b>	39,1	39,3	36,8	25,4	6,8

Tableau I.4 : Influence de l'hydrotraitement sur les caractéristiques d'un gasoil [15].

Dans le moteur diesel, il est nécessaire que le gasoil présente une structure chimique favorable à l'autoinflammation. Cette qualité s'exprime par l'indice de cétane. Sur le moteur diesel, on estime qu'un indice de cétane proche de 50 constitue une valeur minimale acceptable pour réaliser un contrôle satisfaisant des différentes phases de combustion. Le procédé d'hydrotraitement permet d'améliorer cette caractéristique (tableaux I.2 et I.4), ce qui engendre [11,12,15] :

- un démarrage plus facile du véhicule dû à la diminution du délai d'auto-inflammation (figure I.5) ;
- une réduction du bruit en particulier au ralenti à froid (figure I.6) ;
- une émission moins élevée de fumée au démarrage (figure I.7) ;
- une diminution des émissions d'oxyde de carbone CO et d'hydrocarbures imbrûlés (HC), à l'échappement. En effet une augmentation de 8 points de l'indice de cétane d'un gasoil alimentant une voiture particulière, entraîne des diminutions de plus de 25% des émissions de CO et HC.

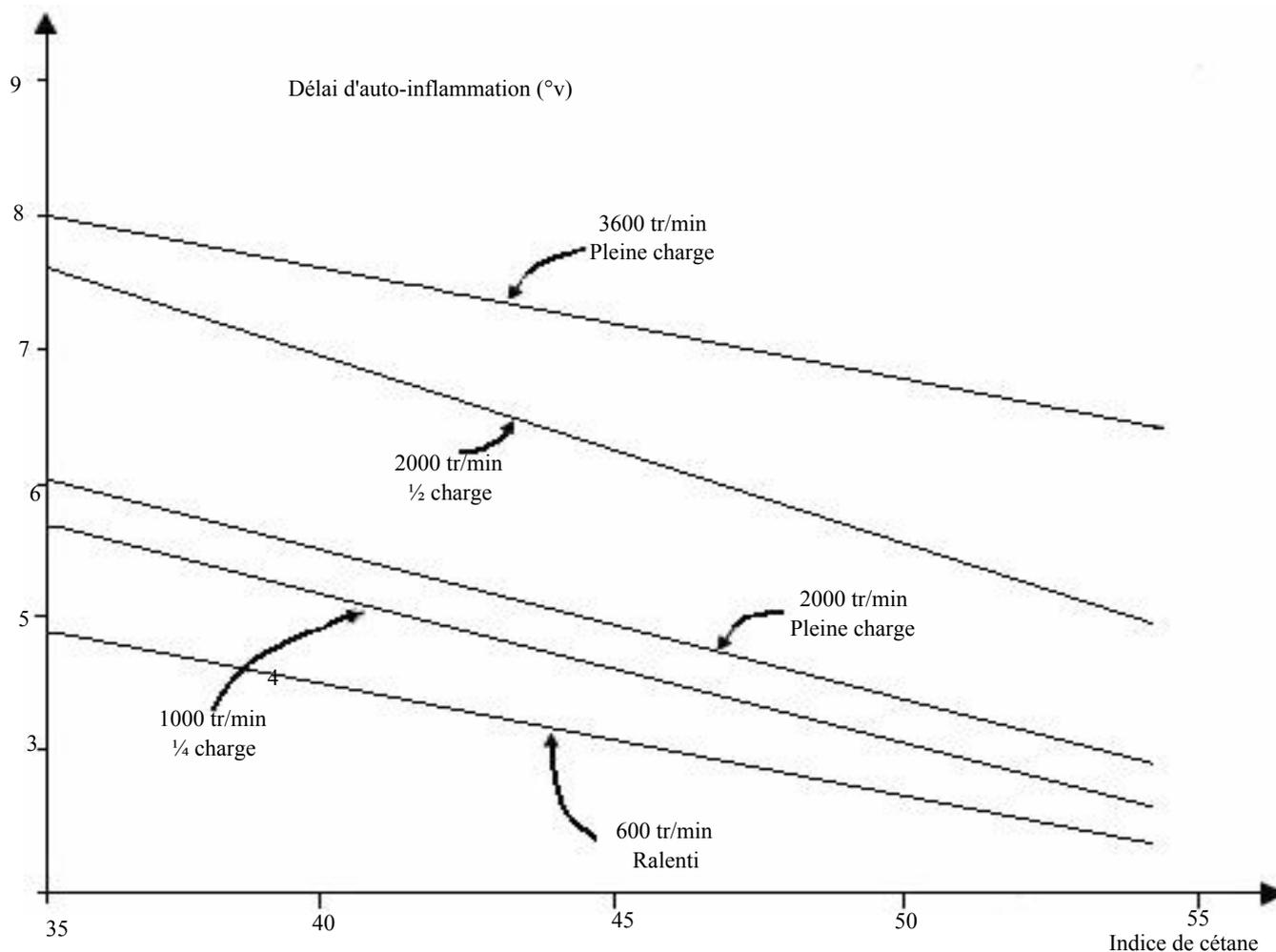


Figure I.5 : Influence de l'indice de cétane sur le délai d'auto-inflammation [12].

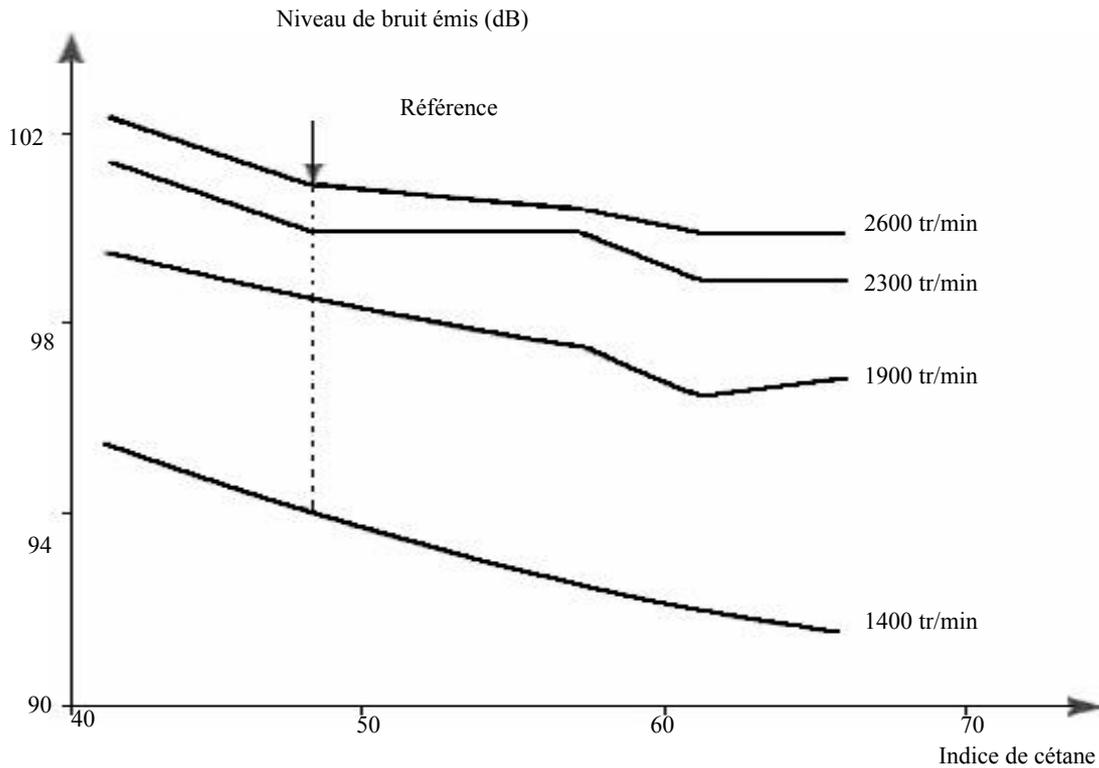


Figure I.6 : Relation entre le niveau de bruit d'un moteur à différents régimes et l'indice de cétane du carburant (moteur diesel à injection directe fonctionnant en pleine charge. Relevé de bruit à 1 m devant le moteur) [12].

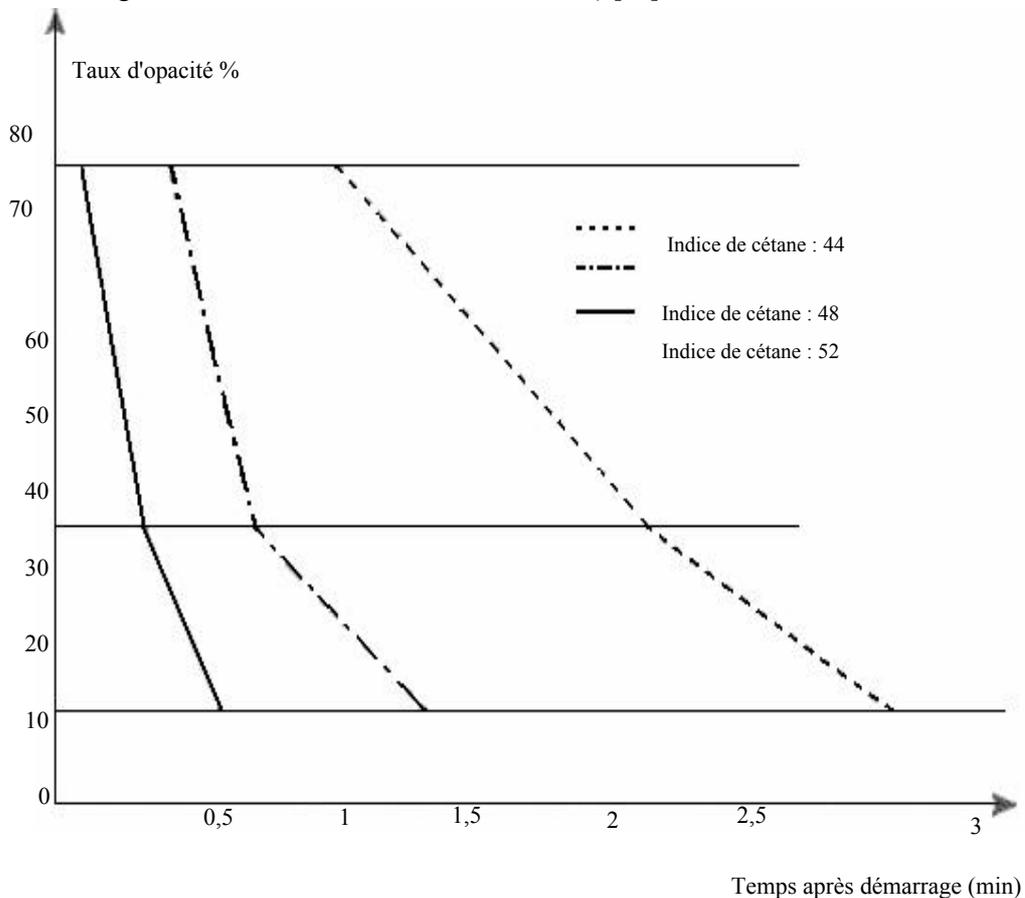


Figure I.7 : Influence de l'indice de cétane du gasoil sur les émissions de fumées, au démarrage à froid. Moteur diesel à injection directe suralimenté (poids lourd). Démarrage à -10°C [11].

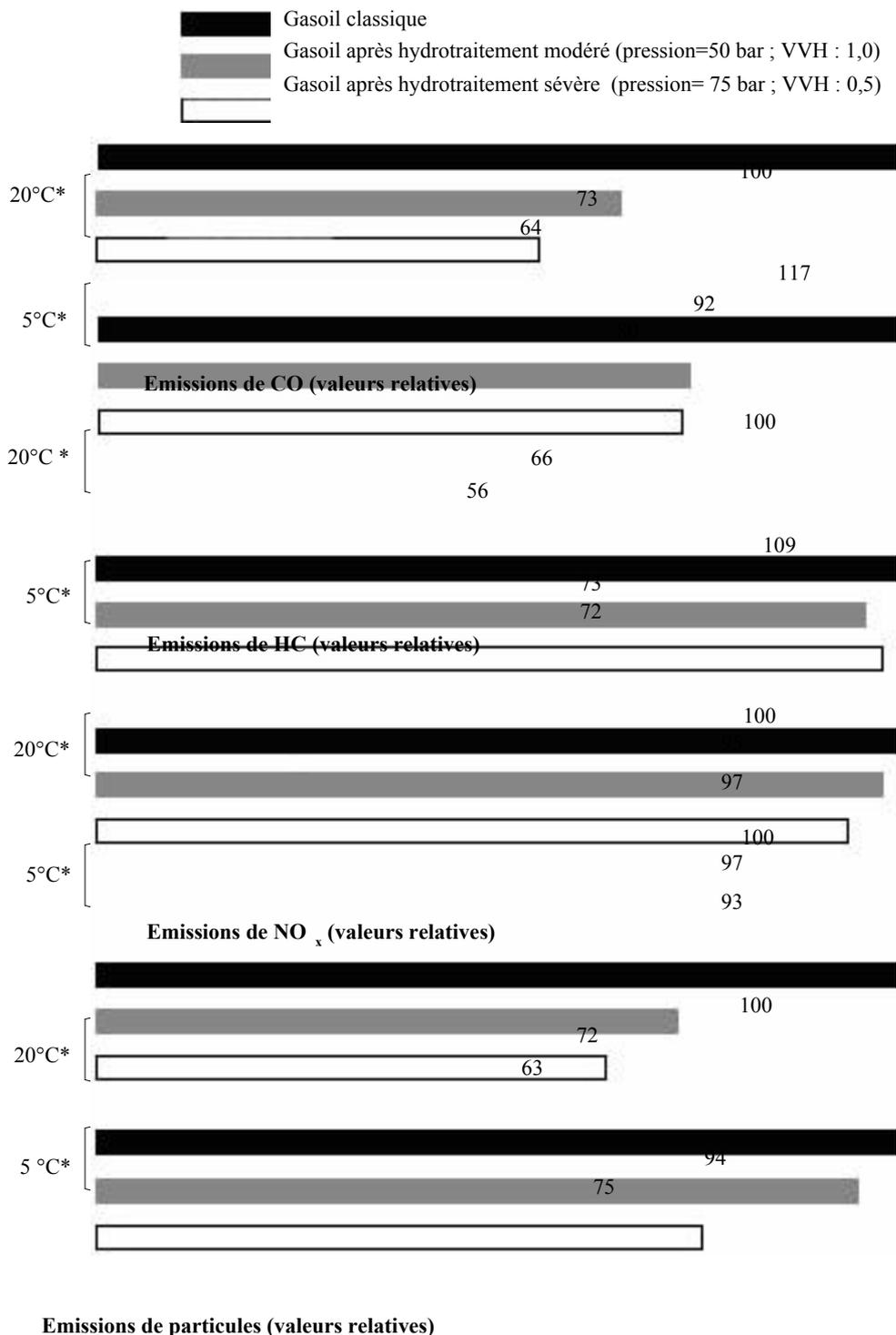
Les gaz d'échappement ont un impact sur la qualité de l'air et la santé humaine, notamment dans les zones urbaines où la circulation automobile est particulièrement dense. Pour réduire cet impact, les voitures doivent respecter certaines normes relatives aux gaz d'échappement avant qu'elles ne soient autorisées à la vente. Le tableau I.5 montre l'évolution de ces normes en Europe depuis 1996 à 2005 pour les véhicules diesel. Ces normes, de plus en plus sévères, ont obligé les constructeurs automobiles à équiper les véhicules diesel de pots catalytiques et de filtres à particules, mais également les raffineurs à améliorer la qualité du gasoil principalement par le procédé d'hydrotraitement [16,17,18]. En effet ce dernier, permet (figure I.8) la réduction des émissions de polluants [15], cette réduction est de l'ordre de :

- 30% pour CO ;
- 40% pour HC ;
- 25% pour les particules.

Par ailleurs, les émissions des oxydes d'azote  $\text{NO}_x$  restent pratiquement inchangées.

<b>Norme</b>	<b>Euro 2 (1996)</b>	<b>Euro 3 (2000)</b>	<b>Euro 4 (2005)</b>
Oxydes d'azote $\text{NO}_x$ (mg/Km)	-	500	250
Monoxyde de carbone CO (mg/Km)	700/900	560	300
Hydrocarbures (mg/Km)	-	-	300 ( $\text{NO}_x$ +HC)
Particules (mg/Km)	80-100	50	25

Tableau I.5 : Evolution des normes européennes pour les émissions de polluants appliquées aux véhicules diesel neufs [17].



133

\*Les tests sont effectués à deux températures ambiantes : 20°C et 5°C

Figure I.8 : Influence de l'hydrotraitement du gasoil sur les émissions à l'échappement d'une voiture particulière diesel [15].

### I.3 Caractéristiques générales

#### I.3.1 Les réactions de l'hydrotraitement

Les différents hétéroatomes présents dans la coupe gasoil sont enlevés des molécules qui les contiennent par réaction avec l'hydrogène à la surface du catalyseur. Ces réactions conduisent à l'élimination du soufre, de l'azote et de l'oxygène sous la forme d'H<sub>2</sub>S, d'NH<sub>3</sub> et d'H<sub>2</sub>O. En parallèle aux réactions d'hydroraffinage, les hydrocarbures insaturés, oléfines et aromatiques seront en partie hydrogénés.

##### I.3.1.1 Caractéristiques thermodynamiques

Toutes les réactions de l'hydrotraitement sont des réactions exothermiques comme l'indiquent les valeurs négatives des chaleurs de réaction du tableau I.6.

Les réactions de rupture des liaisons carbone hétéroatomes ainsi que les réactions d'hydrogénation des oléfines, sont généralement quasi complètes dans les conditions habituelles de l'hydrotraitement. En revanche pour les réactions d'hydrogénation des aromatiques, elles peuvent atteindre l'équilibre thermodynamique, la figure I.9 montre la limitation thermodynamique atteinte lors de l'hydrogénation d'un gasoil. Une augmentation de la température de réaction ne provoque pas une décroissance de la teneur en aromatiques mais au contraire une augmentation. La figure 1.9 montre également que plus l'aromatique est lourd et donc le nombre de noyaux élevé, plus la limite thermodynamique sera placée vers une basse température pour une pression d'hydrogène donnée. L'hydrogénation poussée des aromatiques nécessite donc une forte pression d'hydrogène [1,19].

Réaction	Chaleur de réaction (KJ/mol)
2-propanethiol + H <sub>2</sub> ——— propane + H <sub>2</sub> S	-54,3
R-S-R + 2 H <sub>2</sub> ——— 2 RH + H <sub>2</sub> S	-117,0
R-S-S-R + 3 H <sub>2</sub> ——— 2 RH + 2 H <sub>2</sub> S	-209,0
Cyclohexanethiol + H <sub>2</sub> ——— n-butane + H <sub>2</sub> S	-112,8
Thiophène + 4 H <sub>2</sub> ——— n-butane + H <sub>2</sub> S	-284,2
Dibenzothiophène + 2 H <sub>2</sub> ——— biphényle + H <sub>2</sub> S	-46,0
R-NH <sub>2</sub> + H <sub>2</sub> ——— RH + NH <sub>3</sub>	-79,4
o-éthylaniline + H <sub>2</sub> ——— éthylbenzène + NH <sub>3</sub>	-58,5
Indole + 3 H <sub>2</sub> ——— éthylbenzène + NH <sub>3</sub>	-204,8
Quinoléine + 2 H <sub>2</sub> ——— 1,2,3,4-tétrahydroquinoléine	-133,8
Quinoléine + 4 H <sub>2</sub> ——— propylbenzène + NH <sub>3</sub>	-271,7
R-OH + H <sub>2</sub> ——— RH + H <sub>2</sub> O	-100,3
Phénol + H <sub>2</sub> ——— benzène + H <sub>2</sub> O	-66,9
Furane + 4 H <sub>2</sub> ——— n-butane + H <sub>2</sub> O	-351,1
Benzofurane + 3 H <sub>2</sub> ——— éthylbenzène + H <sub>2</sub> O	-104,5
Benzène + 3H <sub>2</sub> ——— cyclohexane	-209,0
Naphtalène + 2 H <sub>2</sub> ——— tétraline	-125,4

Tableau I.6 : Chaleurs de réaction de divers composés purs traités par hydrogène [1].

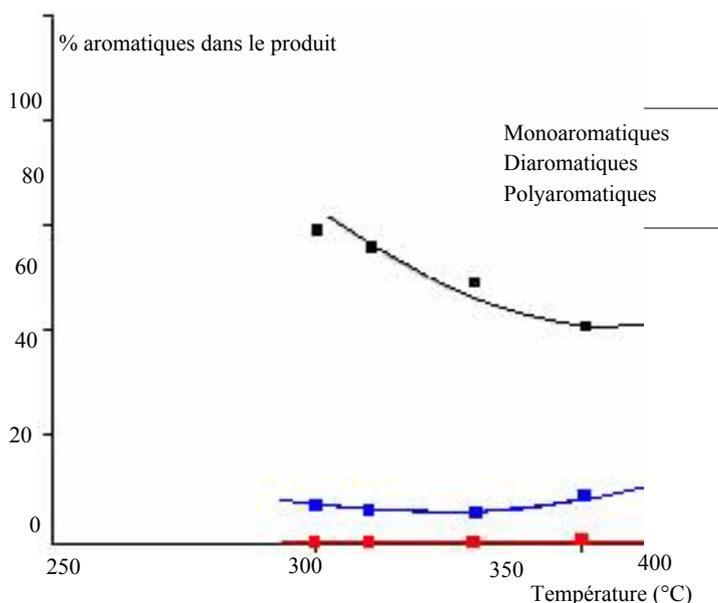


Figure I.9 : Limitation thermodynamique de l'hydrogénation des aromatiques dans un gasoil. Catalyseur NiMo/alumine, P<sub>H<sub>2</sub></sub> = 60 bar, 1000 ppm de S dans la charge [1].

### I.3.1.2 Cinétique des réactions

De très nombreuses lois cinétiques ont été proposées pour les réactions d'HDS, d'HDN et d'HYD de composés modèles. Elles dépendent entre autre, de la nature des réactifs et des conditions opératoires [1]. Généralement, la vitesse de réaction pour les réactifs simples est de la forme suivante:

$$v = \frac{k K_{HC} P_{HC}}{(1 + K_{HC} P_{HC} + \sum_i K_i P_i)} f(P_{H_2}) \dots\dots\dots(I.1)$$

Où HC désigne le réactif, H<sub>2</sub> l'hydrogène, i l'ensemble des composés inhibiteurs, P<sub>i</sub> la pression partielle et K<sub>i</sub> les constantes d'adsorption des produits i, k la constante de vitesse de la réaction et f une fonction plus ou moins complexe de la pression partielle d'hydrogène P<sub>H<sub>2</sub></sub>.

En présence d'H<sub>2</sub>S et de composés fortement adsorbés et donc inhibiteurs, on obtient souvent des lois d'ordre un (01) par rapport à l'hydrocarbure et à l'hydrogène de type :

$$v = k' P_{HC} P_{H_2} \dots\dots\dots(I.2)$$

### I.3.2 Les catalyseurs

Les catalyseurs d'hydrotraitement sont composés d'un support oxyde et d'une phase active sous la forme de sulfure de molybdène ou de tungstène promu par le cobalt ou le nickel. Les formules communément employées sont les associations CoMo, NiMo et NiW pour la phase active et l'alumine γ de grande aire spécifique pour le support [1,20,21].

Les catalyseurs d'hydrotraitement se présentent sous différentes formes billes, extrudés cylindriques, extrudés polylobés, anneaux,...etc. La forme la plus communément employée est l'extrudé cylindrique ou polylobé de petit diamètre : 1,2 à 1,6 mm. Il existe de nombreuses méthodes de préparation de ce type de catalyseurs, les plus courantes sont l'imprégnation du support oxyde ou le malaxage et l'extrusion du mélange du gel d'oxyde et des éléments actifs [1,21].

Le catalyseur d'hydrotraitement est fabriqué avec des métaux sous forme d'oxyde, il subit dans le réacteur d'hydrotraitement, avant son utilisation, une sulfuration du précurseur oxyde généralement représenté par une formule de type CoO-MoO<sub>3</sub>/alumine pour un catalyseur CoMo par exemple [1,21,22]. La réaction de sulfuration est très exothermique, elle consomme de l'hydrogène sulfuré, de l'hydrogène et génère de l'eau selon la réaction :



La sulfuration transforme la phase oxyde en une phase sulfure de structure totalement différente. Ainsi, la phase active des catalyseurs d'hydrotraitement est constituée d'une phase sulfure mixte. Ces phases mixtes sont souvent symbolisées par CoMoS, NiMoS et NiWS.

Il existe différentes méthodes industrielles de sulfuration, on distingue :

- la sulfuration en phase gazeuse, qui utilise en général un mélange d'hydrogène et d'hydrogène sulfuré ou parfois un sulfure organique tel que le méthyl mercaptan ou le diméthyl sulfure;
- la sulfuration en phase liquide, qui consiste à injecter sous hydrogène une coupe essence lourde ou gasoil à laquelle on a rajouté le sulfure organique ;
- la sulfuration ex-situ, qui est de plus en plus utilisée à cause de sa simplicité, elle consiste à faire activer le catalyseur par dépôt d'un composé sulfuré lourd par imprégnation suivi d'un chauffage. Cette opération s'effectue hors site par des sociétés spécialisées tels que la société Eurecat.

En plus de l'alumine  $\gamma$ , il existe d'autres supports oxyde tels que les silice-alumines, les titane-alumines et les dioxydes de titane. Le support oxyde possède plusieurs rôles, il doit fournir une grande aire spécifique pour permettre de déposer une grande quantité de phase active sous la forme la plus dispersée possible. Il doit ensuite conférer au catalyseur des propriétés mécaniques telles que la résistance à l'écrasement et à l'attrition.

Parmi les principales caractéristiques des supports on distingue la dimension des grains et la taille des pores, ces deux caractéristiques doivent être optimisées pour s'affranchir des limitations diffusionnelles. Une autre caractéristique importante des supports est l'acidité, cette fonction favorise les réactions de craquage et de formation de coke, ce dernier constitue la principale cause de désactivation des catalyseurs, c'est pour cette raison que l'acidité doit être limitée [1,4,23].

La désactivation des catalyseurs d'hydrotraitement par les dépôts de coke est favorisée par une haute température et une faible pression d'hydrogène. Ce type de désactivation est réversible car le carbone peut être brûlé pendant la régénération.

Le catalyseur peut être également désactivé par les dépôts d'impuretés métalliques tels que Ni, V, Si, Pb, As...etc. Ces poisons sont irréversibles mais en général ils se déposent en tête de lit, on peut donc procéder à un écrémage de la tête et rajouter du catalyseur neuf si l'on souhaite économiser le catalyseur. La durée de cycle d'un catalyseur d'hydrotraitement du gasoil varie entre 1 et 2 ans et sa durée de vie peut aller jusqu'à 6 ans, tout dépend de la composition de la charge et des conditions opératoires du procédé [1,21].

Les catalyseurs de type CoMo sont en général employés pour l'hydrodésulfuration, tandis que les catalyseurs de type NiMo sont utilisés pour l'hydrodézotation et l'hydrodéoxygénation. Pour l'hydrogénation des aromatiques et des oléfines on utilise le plus souvent les catalyseurs de type NiW. Le tableau I.7 indique la classification des différents catalyseurs industriels suivant leur activité pour les réactions d'hydrotraitement [4].

Réactions	catalyseurs
Hydrodésulfuration	CoMo > NiMo > NiW
Hydrodéazotation	NiMo = NiW > CoMo
Hydrodéoxygénation	NiMo > CoMo > NiW
Hydrogénation des aromatiques et des oléfines	NiW > NiMo > CoMo

Tableau I.7 : Classification des catalyseurs industriels du procédé d'hydrotraitement suivant leur activité [4].

Dans l'industrie, le catalyseur le plus employé pour l'hydrotraitement du gasoil est le CoMo/alumine, mais actuellement et avec les nouvelles spécifications, ce catalyseur ne suffit plus, on a donc pensé à l'améliorer en le dopant par le phosphore. En effet un nouveau catalyseur à été commercialisé au Japon de formule CoMoP/Al<sub>2</sub>O<sub>3</sub> et de référence C-606A [24]. Il se caractérise par une grande activité et une grande stabilité, il permet d'ailleurs la production d'un gasoil avec des teneurs en soufre inférieures à 10 ppm.

Une autre méthode a été développée pour l'amélioration des catalyseurs d'hydrotraitement, c'est l'imprégnation par les métaux nobles tels que le Pt et le Ru [4,25,26,27]. En effet plusieurs catalyseurs ont été mis au point mais ne sont toujours pas utilisés dans l'industrie, on cite par exemple le PtMo/Al<sub>2</sub>O<sub>3</sub>, RuMo/Al<sub>2</sub>O<sub>3</sub> et le RuCoMo/Al<sub>2</sub>O<sub>3</sub>.

Il existe actuellement des procédés d'hydrotraitement du gasoil utilisant des catalyseurs à base de Pt ou de Pd déposés sur zéolites [28,29]. Pour palier la rapide désactivation des métaux nobles par le soufre, la charge est traitée, dans une première étape, par un catalyseur classique généralement de type CoMo, la teneur en soufre du gasoil obtenu est d'environ 20 ppm. Ensuite la charge subit, dans une deuxième étape, un hydrotraitement plus profond qui permet la production d'un gasoil avec des teneurs en soufre inférieures à 5 ppm. Les teneurs faibles en soufre et en aromatiques du gasoil ainsi améliorées sont obtenues par utilisation de pressions d'hydrogène élevées d'où un coup de revient relativement plus élevé.

Une autre technique permet l'obtention d'un gasoil de bonne qualité et ceci sans la modification de la composition du catalyseur, c'est l'hydrotraitement en deux stages, utilisant à chaque fois le même catalyseur CoMo/Al<sub>2</sub>O<sub>3</sub>. D'excellents résultats ont été obtenus en particulier pour l'hydrodésulfuration, on arrive d'ailleurs à obtenir des gasoils avec des teneurs en soufre inférieures à 10 ppm [29].

## I.4 Variables opératoires

Les variables opératoires cinétiques et thermodynamiques principales du procédé d'hydrotraitement du gasoil sont [1] :

- la pression partielle d'hydrogène ( $P_{H_2}$  en bar) ;
- la température moyenne pondérée du réacteur ( $T_m$  en °C) ;
- la vitesse spatiale horaire (VVH en m<sup>3</sup> de charge/m<sup>3</sup> de catalyseur/h).

De plus deux variables secondaires sont importantes :

- le taux de recyclage et la pureté de l'hydrogène recyclé ;
- la pression partielle d'H<sub>2</sub>S.

Le choix de ces variables conduit à la détermination :

- de la pression opératoire ;
- du schéma du procédé ;
- de l'investissement et du coût opératoire.

Un facteur important pour la définition des conditions opératoires est la nature de la charge. Divers paramètres peuvent être utilisés pour définir la charge tels que la densité, l'indice de réfraction, la température à 95% de distillation...etc.

#### **I.4.1 La pression partielle d'hydrogène**

L'hydrogène étant l'un des réactifs principaux, le choix de son niveau de pression est capital pour le bon fonctionnement du procédé. La pression d'hydrogène nécessaire croit avec la lourdeur de la charge. Pour un gasoil donné, par exemple de distillation directe, elle varie avec les performances requises [1]:

Hydrodésulfuration	15-25 bar
Hydrostabilisation (hydrogénation modérée)	25-35 bar
Hydrodéaromatisation (hydrogénation profonde)	35-60 bar

L'amélioration de l'indice de cétane et la réduction de la teneur en aromatiques exigent des pressions partielles d'hydrogène particulièrement élevées. La consommation chimique d'hydrogène, qui dépend des performances requises, est également importante comme l'indique le tableau I.8.

La pression d'hydrogène a donc pour but de déplacer les réactions dans le sens favorable et de minimiser les réactions parasites telles que la production de gaz et la formation de dépôts de coke. Ce qui permet l'amélioration du rendement du procédé et l'augmentation de la durée de vie du catalyseur [1].

<b>Pression partielle d'hydrogène (bar)</b>	<b>Teneur en aromatiques dans l'effluent (%masse)</b>	<b>Consommation chimique d'hydrogène (%masse)</b>
30	25,7	0,4
40	20,0	0,6
65	10,0	0,9
85	5,0	1,1

Tableau I.8 : Influence de la pression partielle d'hydrogène sur l'hydrotraitement d'un gasoil de distillation primaire contenant 26,7% masse d'aromatiques et 1,3% masse de soufre [1].

### I.4.2 La température moyenne pondérée du lit catalytique

A fin d'étudier l'effet de la température sur le procédé d'hydrotraitement du gasoil, il est nécessaire de définir la notion de température moyenne pondérée du lit catalytique. Etant donné que les réactions d'hydrotraitement sont des réactions exothermiques, il s'établit dans le lit un gradient croissant de température. A défaut de connaître exactement la température en chaque point du lit catalytique, le positionnement de thermocouples permet de mesurer les températures de plusieurs tranches du lit catalytique et on peut ainsi définir une température moyenne pondérée du lit par :

$$T_m = \frac{t_1 v_1 + t_2 v_2 + \dots + t_n v_n}{V} \dots\dots\dots(I.3)$$

Avec :  $V$  le volume total de catalyseur.

$$v_i \text{ le volume de chaque tranche : } V = v_1 + v_2 + \dots + v_n \dots\dots\dots(I.4)$$

Pour un réacteur contenant un seul lit de catalyseur, la température moyenne pondérée peut être déterminée approximativement par :

$$T_m = \frac{t_e + t_s}{2} = t_e + \frac{1}{2} \Delta t = t_s - \frac{1}{2} \Delta t \dots\dots\dots(I.5)$$

Avec  $\Delta t = t_s - t_e \dots\dots\dots(I.6)$

Où  $t_e$  est la température à l'entrée du réacteur.

$t_s$  est la température à la sortie du réacteur.

Pour montrer l'effet de la température moyenne pondérée sur le procédé d'hydrotraitement, une étude a été réalisée à l'aide d'un réacteur pilote utilisant comme charge un gasoil de distillation directe et comme catalyseur le NiMo/γ-Al<sub>2</sub>O<sub>3</sub> [30]. La figure I.10 indique l'influence de la  $T_m$  sur les caractéristiques du produit pour différents rapports hydrogène/charge. Le domaine de variation de la température correspond à celui généralement rencontré en industrie c'est-à-dire 340-360°C. D'après ces courbes, on constate qu'une augmentation de la température moyenne pondérée engendre une diminution des teneurs en soufre, azote et aromatiques du produit, cette variation de température est donc bénéfique.

Les unités d'hydrotraitement fonctionnent à performance constante et à pression constante. Compte tenu de la désactivation progressive du catalyseur, le maintien du niveau de performance est assuré par l'augmentation de la température de réaction [1]. On définit ainsi une température moyenne pondérée de début de cycle ( $Tm_{DDC}$ ) et une température moyenne pondérée de fin de cycle ( $Tm_{DFC}$ ).

La température moyenne pondérée de fin de cycle est la température maximale que peut tolérer le catalyseur, elle est en général de 400 à 410 °C en fonction de la nature de la charge et des qualités requises pour le produit (teneur en soufre, teneur en aromatiques, stabilité...etc.). En effet lorsque la température devient supérieure à 410 °C, le craquage thermique et la formation de coke deviennent importants et ne peuvent plus être modérés par une augmentation de la pression partielle d'hydrogène.

La température moyenne pondérée de début de cycle est la variable principale d'une unité en fonctionnement, elle a une influence très importante sur les performances et sur la durée de cycle du catalyseur. Elle dépend des caractéristiques physicochimiques de la charge. Par exemple, lorsque le poids moléculaire moyen du gasoil augmente, la quantité de composés soufrés à traiter et la difficulté de la désulfuration augmentent aussi. Ceci se traduit par une diminution de la vitesse spatiale horaire et une augmentation de la température de début de cycle.

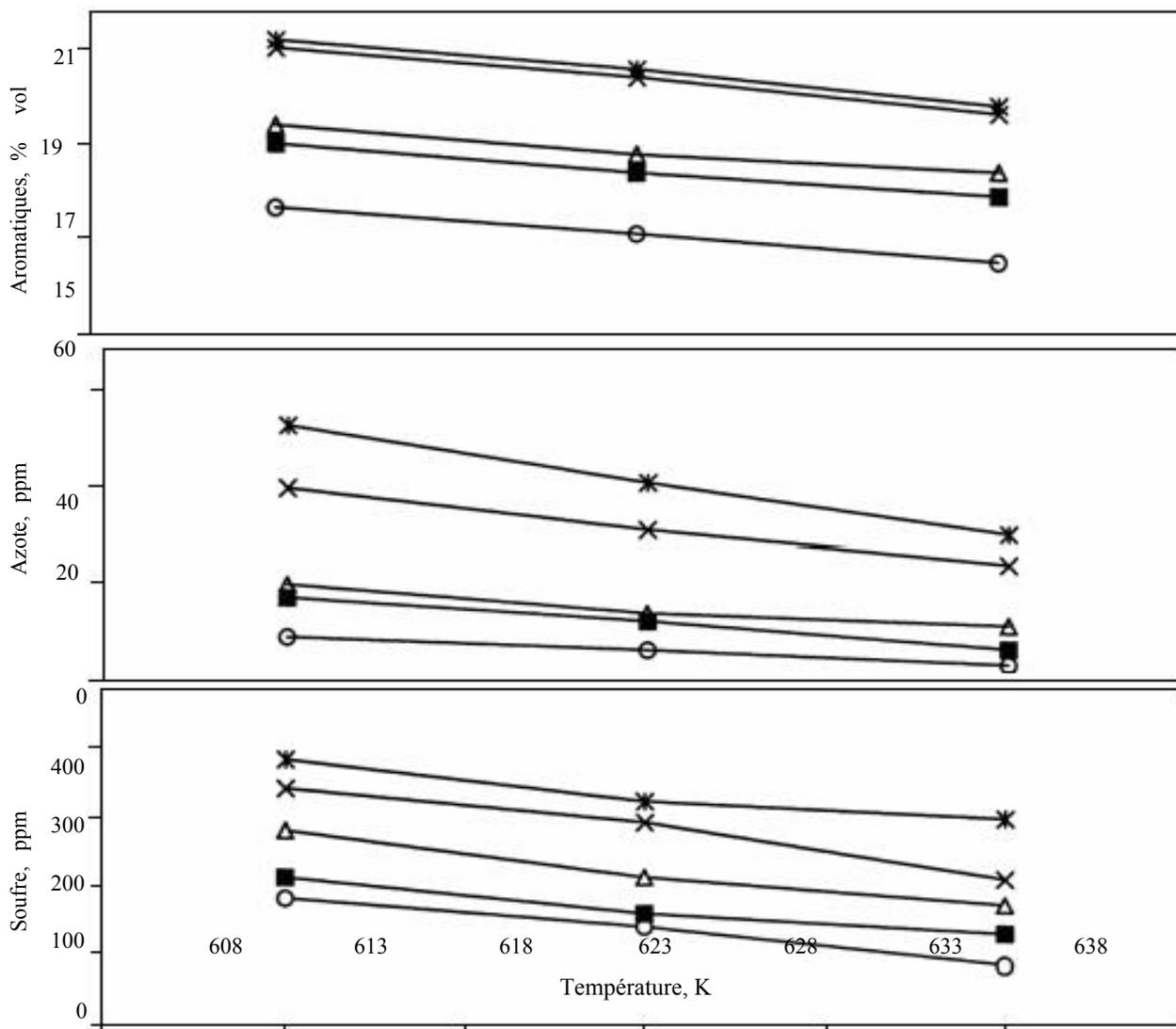


Figure I.10 : Influence de la Tm d'hydrotraitement sur la composition d'un gasoil pour différents rapports hydrogène/hydrocarbure. (○) 500 ml.ml , (■) 450 ml.ml , (Δ) 400 ml.ml , (×) 350 ml.ml , (\*) 300 ml.ml [30].

-1

### I.4.3 La vitesse spatiale horaire

La vitesse spatiale horaire (VVH) permet de connaître la quantité de catalyseur à mettre en place pour obtenir une performance donnée. Elle s'exprime en m<sup>3</sup> de charge fraîche (à 15°C) par m<sup>3</sup> de catalyseur et par heure [1,30]. L'inverse de la VVH est généralement appelé temps de séjour. La VVH dépend de l'activité du catalyseur, de la nature de la charge et de la pression partielle d'hydrogène. Pour une charge donnée il y a donc un compromis entre  $P_{H_2}$ , VVH et  $\Delta T$  entre début et fin de cycle.

3

L'influence de la VVH sur les caractéristiques d'un gasoil (de distillation directe) hydrotraité, pour différents rapports hydrogène/charge, est matérialisée par les courbes sur la figure I.11. En effet, une augmentation de la vitesse spatiale se traduit par une dégradation de la qualité du produit, on observe une augmentation des teneurs en soufre, azote et aromatiques.

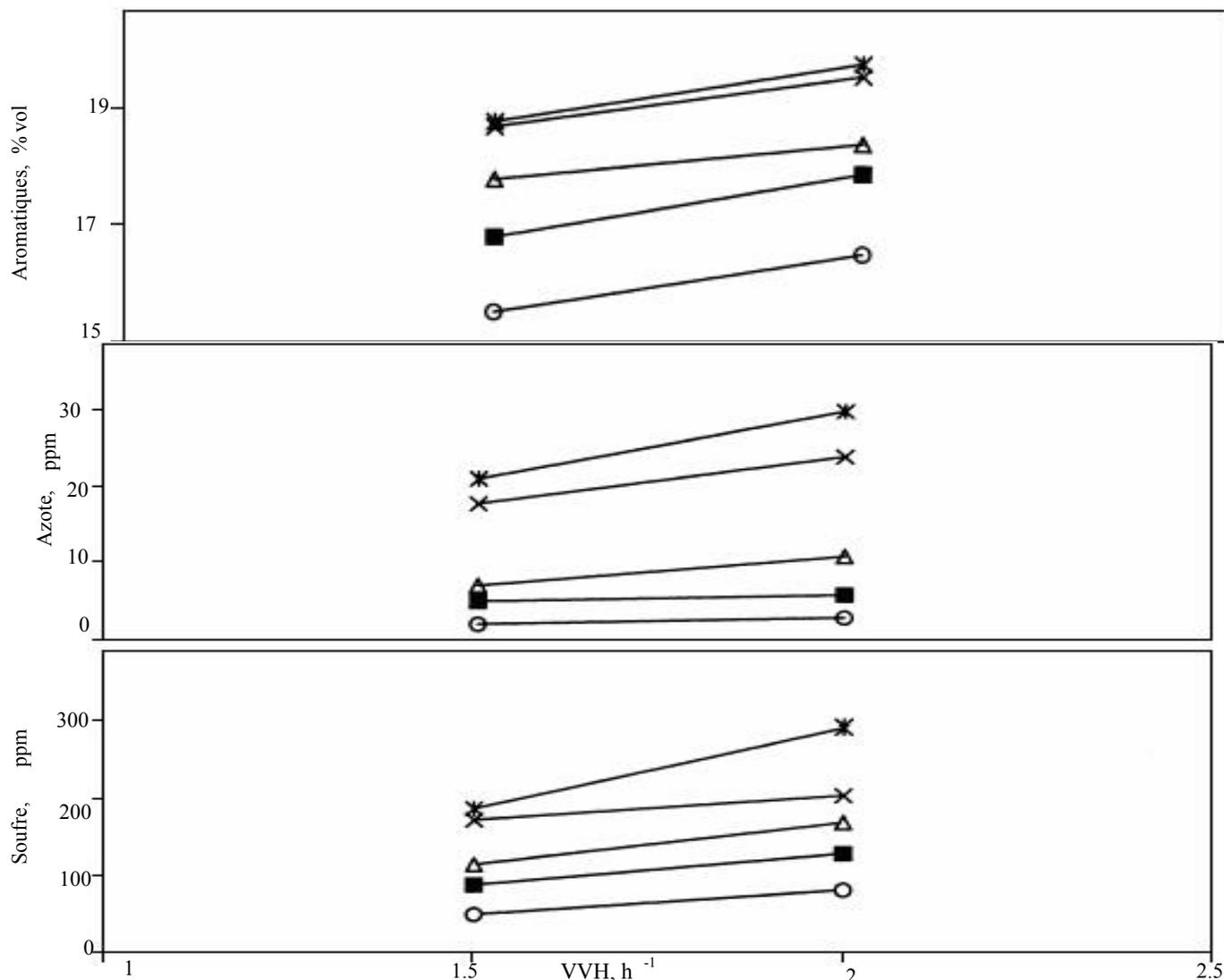


Figure I.11 : Influence de la VVH du procédé d'hydrotraitement sur la composition d'un gasoil pour différents rapports hydrogène/hydrocarbure. (○) 500 ml.ml , (■) 450 ml.ml , (Δ) 400 ml.ml , (×) 350 ml.ml , (\*) 300 ml.ml [30].

#### I.4.4 Taux d'hydrogène ou taux de recyclage

Au cours du procédé d'hydrotraitement du gasoil, il est nécessaire d'introduire avec la charge des quantités d'hydrogène notablement supérieures à celles qui correspondent à la consommation chimique. Les figures I.10 et I.11 indiquent clairement que plus le taux d'hydrogène par rapport à la charge est élevé plus la qualité du produit est meilleure. A fin de minimiser les besoins en hydrogène, un recyclage de gaz est prévu ; ce dernier est bien évidemment complété par un appoint [1,30].

Dans le réacteur, en dehors du maintien de la pression partielle d'hydrogène, le gaz de recyclage assure aussi la turbulence nécessaire au contact des réactifs dans le réacteur. De plus le gaz recyclé à l'entrée de l'unité augmente la turbulence dans les échangeurs et le four, il améliore ainsi les conditions de transfert de chaleur et prévient dans une certaine mesure le dépôt de salissures dans ces appareils [1].

Les réactions du procédé d'hydrotraitement du gasoil sont des réactions exothermiques, il est donc nécessaire de contrôler le gradient de température dans le réacteur. En effet le catalyseur est reparti sur plusieurs lits, entre deux lits on introduit une trempe (quench) de gaz recyclé froid ceci permet d'opérer au plus proche de l'isotherme et de minimiser le vieillissement prématuré du catalyseur [1].

#### **I.4.5 Concentration en hydrogène sulfuré dans le gaz de recyclage**

L'effet de la concentration en hydrogène sulfuré dans le gaz de recyclage sur le procédé d'hydrotraitement du gasoil a été étudié dans une unité pilote, utilisant comme catalyseur le  $\text{CoMo}/\text{Al}_2\text{O}_3$  et comme charge un mélange de 85% volumique de gasoil SR et 15% volumique de gasoil LCO [31].

Les résultats de cette étude, sont matérialisés par les courbes de la figure I.12, en effet une augmentation de la concentration en hydrogène sulfuré engendre une augmentation des teneurs en soufre, azote et aromatiques des produits hydrotraités. Ceci peut être expliqué par l'effet inhibiteur de  $\text{H}_2\text{S}$  sur les réactions d'hydrodésulfuration, d'hydrodéazotation et d'hydrogénation [1,31,32].

Selon les performances requises, un lavage partiel ou total du gaz de recyclage aux amines peut être installé. Cependant il est utile de conserver une certaine pression partielle d' $\text{H}_2\text{S}$  afin de maintenir l'état de sulfuration de la phase catalytique [1].

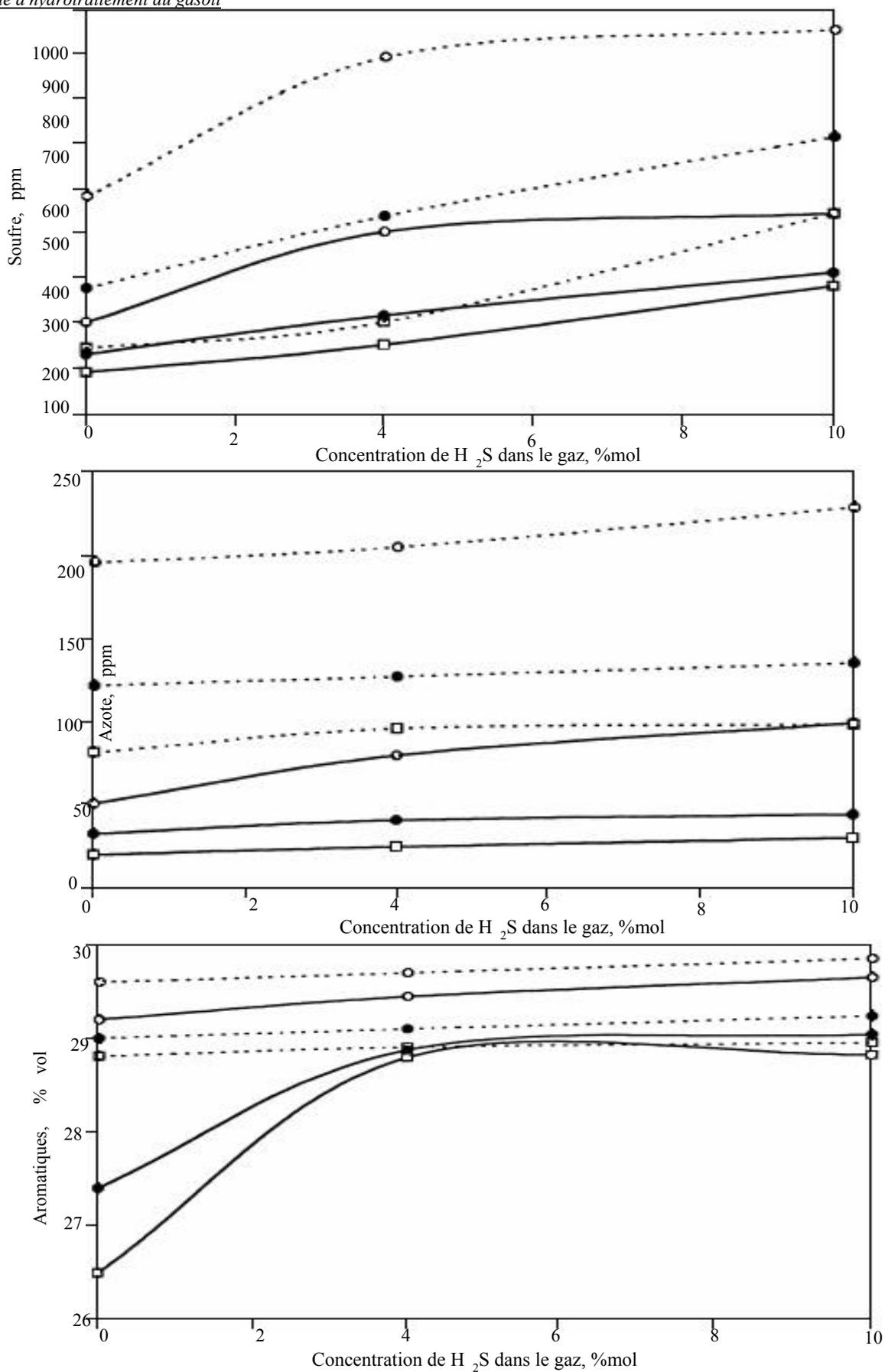


Figure I.12 : Effet de la concentration de H<sub>2</sub>S dans le gaz sur la composition d'un gazoil hydrotraité. (o) 350°C, (●) 360°C, (□) 370°C, (—) VVH=1h, (---) VVH=2h [31].

## I.5 Technologies du procédé

La technologie du procédé d'hydrotraitement du gasoil ainsi que le catalyseur sont adaptés aux performances requises pour les différents gasoils à hydrotraiter. Parmi les équipements qui constituent le procédé (réacteurs, fours, compresseurs, séparateurs, échangeurs...etc.), le réacteur et ses équipements internes sont le cœur de la technologie.

### I.5.1 Le réacteur

Le réacteur du procédé d'hydrotraitement du gasoil est un réacteur à lit fixe, opérant en phase mixte liquide-gaz. L'aménagement de ce type de réacteur est représenté sur la figure I.13, il est conçu pour répartir au mieux le mélange diphasique sur la section du réacteur. Il possède de nombreux équipements internes, qui permettent le contrôle des températures, le contrôle des pertes de charges et l'obtention ainsi des performances recherchées [1,33].

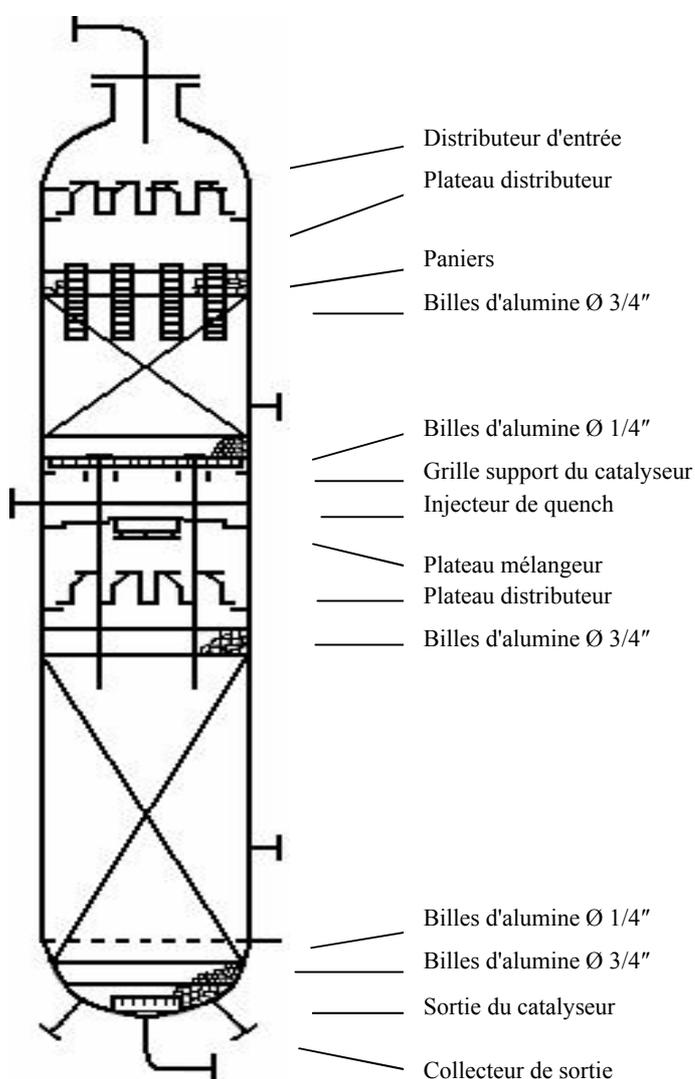


Figure I.13 : Réacteur du procédé d'hydrotraitement du gasoil [1].

## I.5.2 Schémas du procédé

Il existe dans l'industrie plusieurs schémas technologiques du procédé d'hydrotraitement du gasoil, établis par différentes firmes. Le choix entre un schéma et un autre dépend de la composition de la charge et de la qualité des produits recherchés. Parmi ces procédés on distingue :

### I.5.2.1 Le procédé d'hydrotraitement de Chevron Lummus Global LLC [34]

Ce procédé permet le traitement des différents types de gasoils (SR, LCO, de cokéfaction et de viscoréduction) séparément ou mélangés. La charge est mélangée au gaz de recyclage riche en hydrogène (figure I.14), puis chauffée et introduit dans le réacteur (1). Le réacteur contient un ou plusieurs lits catalytiques, on utilise généralement comme catalyseur le CoMo ou le NiMo. Le réacteur fonctionne en mode isotherme ; les trempes de gaz recyclé (froid) entre les lits catalytiques, permettent le maintien de la température pratiquement constante le long du réacteur.

Les effluents du réacteur refroidis entrent dans le ballon de flash (2). Le gaz riche en hydrogène subit un lavage à l'eau puis aux amines (3) pour éliminer  $H_2S$ . Les effluents liquides subissent un deuxième flash (4) à fin d'éliminer la totalité du gaz, puis sont fractionnés dans la colonne (5).

Ce procédé permet l'obtention de gasoils avec de faibles teneurs en soufre (< 10 ppm). Le même procédé est utilisé pour l'hydrotraitement des naphtas et kérosènes ainsi que pour le prétraitement du gasoil lourd pour le procédé de craquage catalytique. On compte aujourd'hui plus de 60 unités dans le monde.

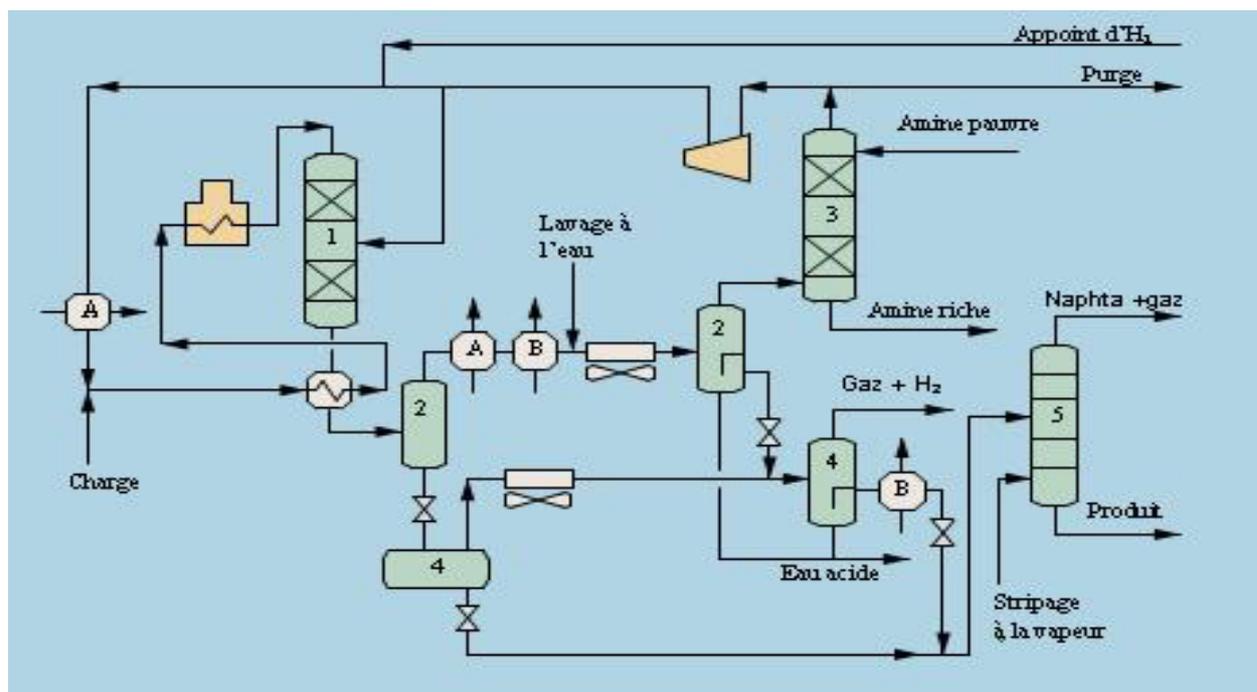


Figure I.14 : Schéma du procédé d'hydrotraitement de Chevron Lummus Global LLC [34].

### I.5.2.2 Le procédé d'hydrotraitement de Axens [35,36]

Le procédé d'hydrotraitement de Axens, traite uniquement les gasoils, il permet d'améliorer leurs caractéristiques à des pressions relativement faibles. Ceci grâce à la grande activité des catalyseurs utilisés le HR526 CoMo et le HR538/HR548 NiMo. Ce procédé se caractérise par l'utilisation de deux réacteurs (figure I.15), dans le second le volume de catalyseur est plus important. On compte plus de 56 unités dans le monde, permettant la production de gasoils avec des teneurs en soufre inférieures à 50 ppm.

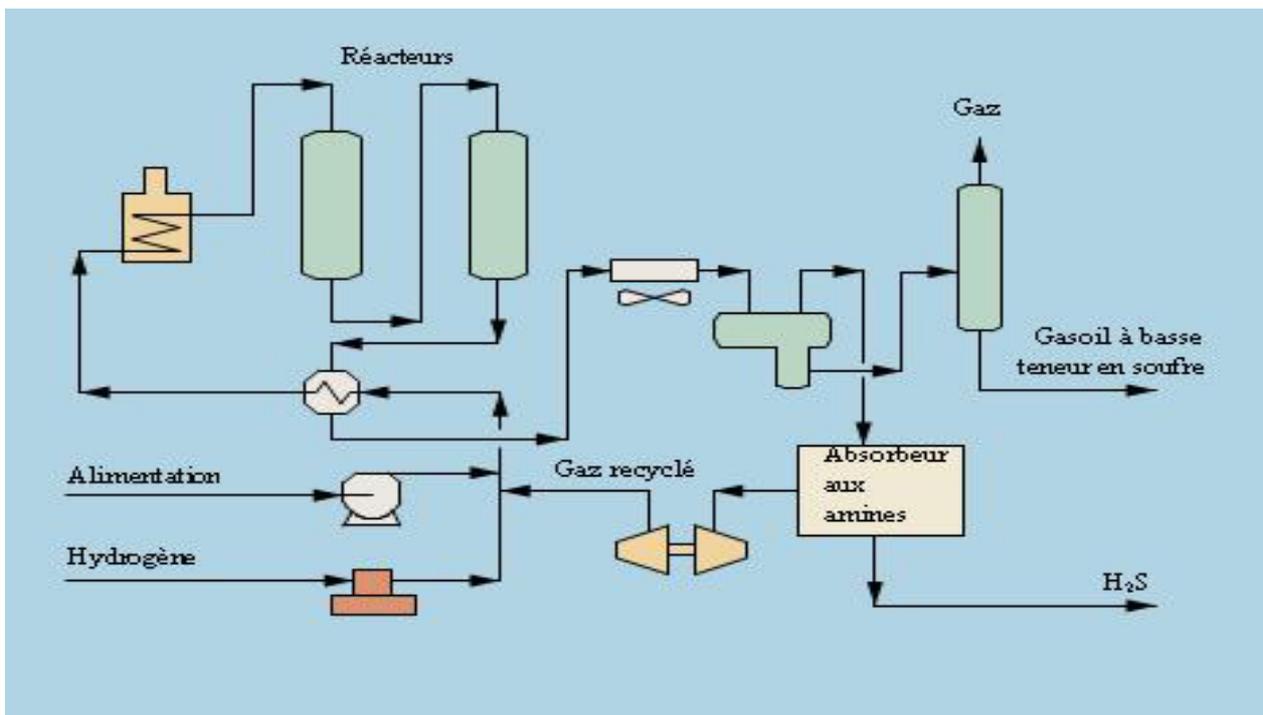


Figure I.15 : Schéma du procédé d'hydrotraitement de Axens [36].

### I.5.2.3 Le procédé ULSD de Haldor Topsøe [37]

Le procédé ULSD (Ultra Low Sulfure Diesel) est un procédé d'hydrodésulfuration des gasoils issus de la distillation primaire et des procédés de conversion thermique, la teneur en soufre des produits varie entre 5 et 50 ppm. Les catalyseurs utilisés sont à base de CoMo et de NiMo, ils sont très stables et très actifs mais leur formule exacte demeure secrète. Il existe actuellement plus de 50 unités d'hydrotraitement ULSD dont 40 permettent l'obtention de gasoils avec des teneurs en soufre inférieures à 10 ppm (figure I.16).

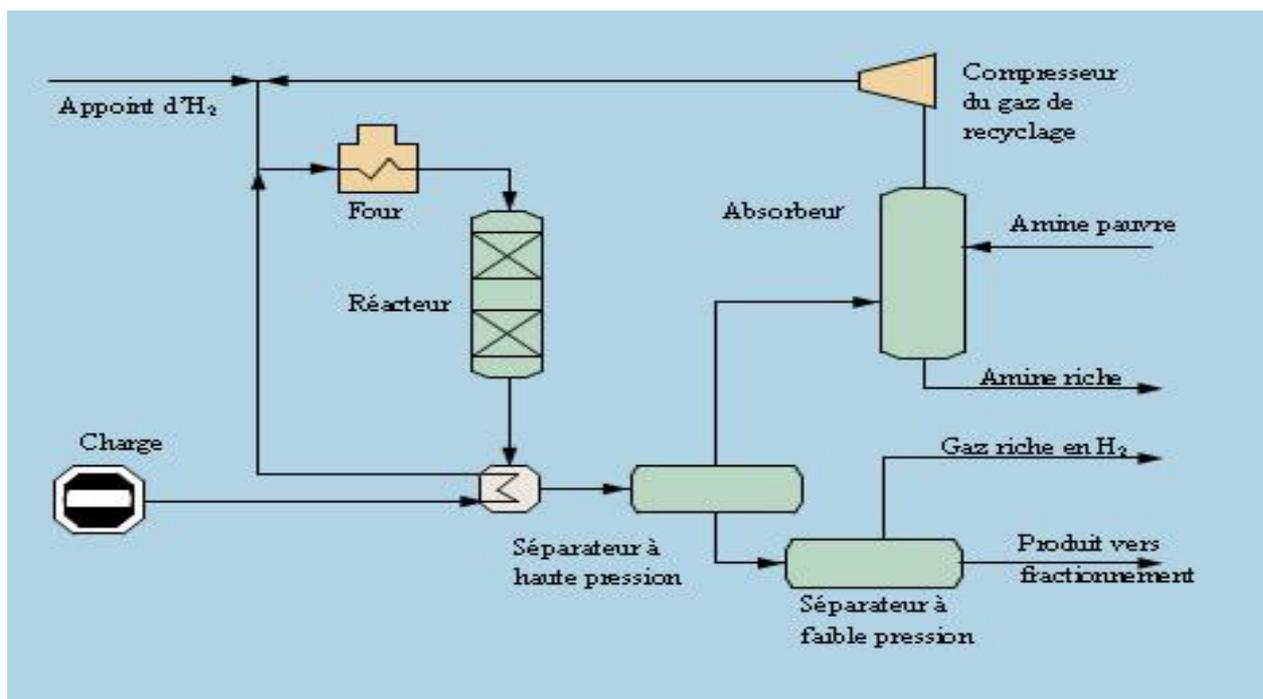


Figure I.16 : Schéma du procédé ULSD de Haldor Topsøe [37].

### I.5.2.4 Le procédé HDS/HDA de Haldor Topsøe [38]

Ce procédé est très performant, il permet l'amélioration de toutes les caractéristiques du gasoil d'une manière spectaculaire, il comprend quatre sections (figure I.17) :

- ¼ L'hydrodésulfuration dans le premier réacteur, qui contient un catalyseur classique de référence TK-575. La pression dans ce réacteur varie entre 20 et 60 bar, et la température entre 320 et 400°C ;
- ¼ Un stripage intermédiaire ;
- ¼ L'hydrodésaromatization dans le second réacteur, contenant un catalyseur à base de métaux nobles de référence TK-907/TK-911 ou TK- 915, la température dans ce cas varie entre 260 et 330°C et la pression entre 20 et 60 bar ;
- ¼ Et enfin un stripage des produits.

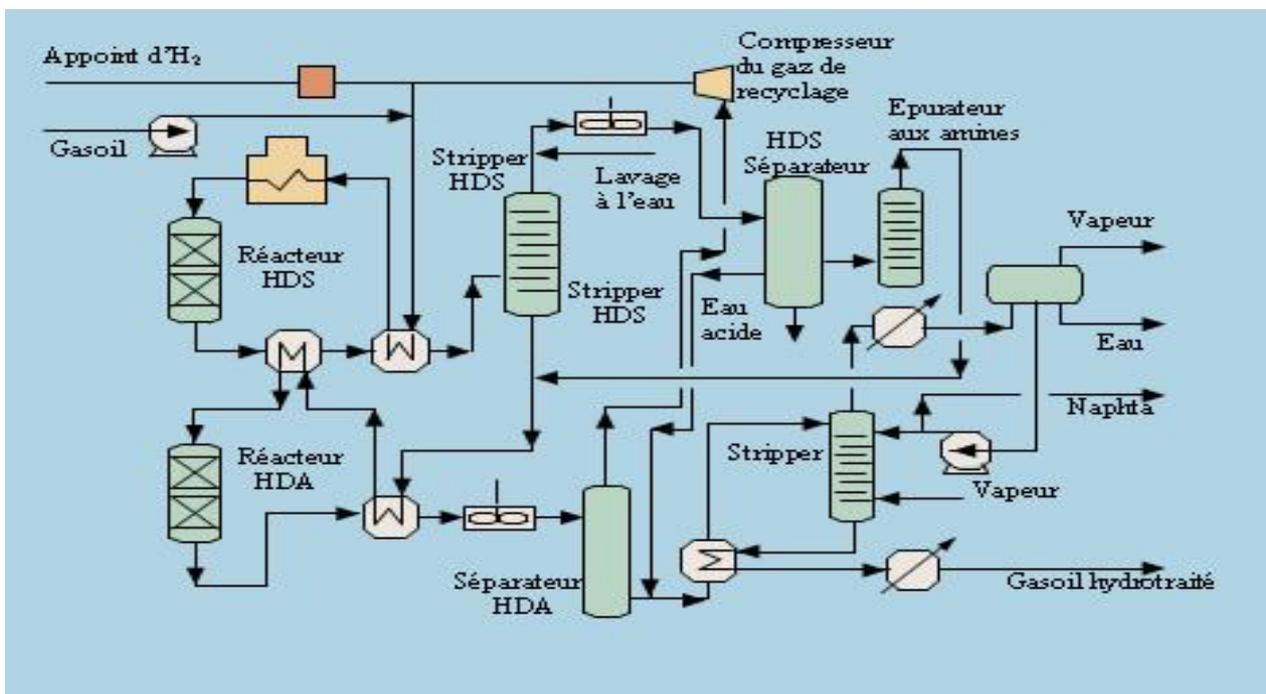


Figure I.17 : Schéma du procédé HDS/HDA de Haldor Topsøe [38].

Un exemple du traitement HDS/HDA de Haldor Topsøe pour un gasoil de distillation directe est donné sur le tableau I.9.

Caractéristiques	charge	produit
Teneur en soufre (ppm)	3000	1
Teneur en azote (ppm)	400	<1
Teneur en aromatiques (%masse)	30	<10
Cétane index	49	57

Tableau I.9: Caractéristiques d'un gasoil de distillation directe avant et après le procédé d'hydrotraitement HDS/HDA de Haldor Topsøe [38].

Ce procédé traite en plus des différents gasoils, le kérosène pour la production du carburéacteur. Malgré les performances de ce procédé on compte aujourd'hui un total de 7 unités seulement dans le monde, ceci est dû à son coup très élevé.

## **I.6 Conclusion**

Dans ce chapitre nous avons effectué une synthèse générale sur le procédé d'hydrotraitement du gasoil, ce dernier permet non seulement la production d'un carburant diesel de haute qualité mais il contribue également à la protection de l'environnement.

Il existe actuellement une grande concurrence entre les plus grandes compagnies de raffinage et entre les laboratoires de recherche les plus réputés, pour améliorer et développer la technologie ainsi que les catalyseurs du procédé d'hydrotraitement du gasoil, ce ci confirme l'importance d'un tel procédé et la nécessité de sa présence dans chaque raffinerie.

## ***CHAPITRE II***

---

### ***LA MODELISATION EN GENIE DES PROCEDES***

---

## II.1 Introduction

Définir la modélisation en génie des procédés est une tâche délicate tant elle recouvre d'actions et d'objectifs divers. On pourrait néanmoins lui donner la définition synthétique suivante [39,40] :

«La modélisation est une démarche qui permet, à partir de faits expérimentaux, de construire un outil mathématique reliant les sorties d'un système à ses entrées.»

On entend par système une structure physique prise de manière isolée, un cristalliseur, un réacteur chimique, une colonne de distillation...etc. Ses entrées sont les paramètres qui agissent sur son comportement. Elles peuvent être contrôlées ou non. Ses sorties sont les paramètres résultant des valeurs données aux entrées.

On distingue dans un procédé (figure II.1) :

- Le vecteur des entrées ( $u_1, u_2, \dots, u_k$ ) ;
- Le vecteur des sorties ( $y_1, y_2, \dots, y_m$ ) ;
- Le vecteur des perturbations ( $p_1, p_2, \dots, p_\ell$ ) ;

Ce dernier ensemble de valeurs (les perturbations) correspond aux paramètres influant sur le comportement statique ou dynamique du procédé sans que l'on puisse les maîtriser ou que l'on veuille les prendre en compte dans le modèle. Ce concept de perturbations est très important pour les problèmes de modélisation dynamique et/ou d'automatisation.

En fait, la description précédente restreint le problème aux bornes du système, c'est-à-dire aux entrées et sorties. Le plus souvent une succession de phénomènes internes intervient dans la relation entre les entrées et les sorties. Par exemple dans une colonne à distiller pour laquelle sont définis débit de reflux et puissance de rebouillage, les caractéristiques (débit et composition) du distillat et du résidu sont la conséquence des phénomènes de transfert de matière et d'équilibre intervenant sur chaque plateau. Ces variables internes sont souvent appelées variables d'état du procédé ( $x_1, x_2, \dots, x_n$ ). Elles peuvent dépendre du temps (lorsqu'elles décrivent un procédé non stationnaire) ou non (lorsqu'elles décrivent un procédé en régime permanent).

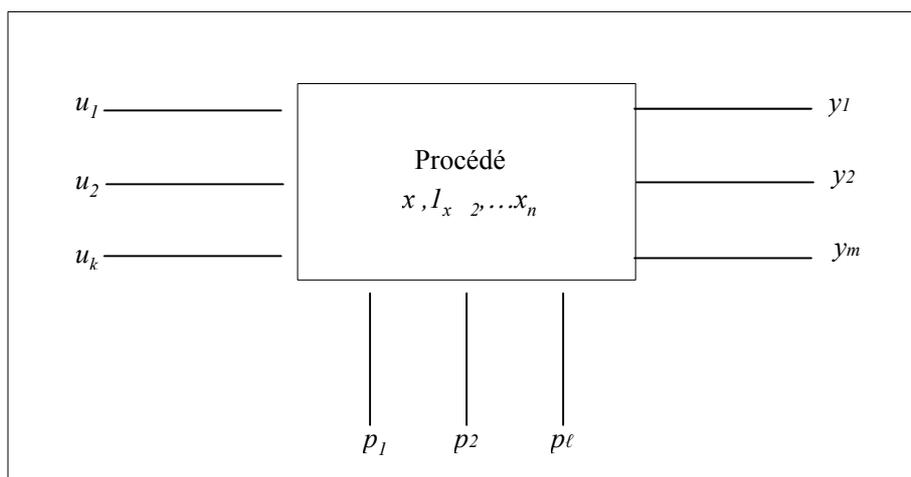


Figure II.1 : Nomenclature décrivant la structure d'un procédé [39].

Généralement, toute description mathématique d'une opération unitaire ou d'un procédé peut se ramener au système d'équations suivantes [39] :

$$\left\{ f \left( \frac{dx(t)}{dt}, x(t), u(t), p(t), a(t), t \mid = \phi \right) \dots\dots\dots (II.1) \right.$$

$$\left. \left\{ g \left( \frac{dy(t)}{dt}, y(t), x(t) \mid \right) = 0 \dots\dots\dots (II.2) \right. \right.$$

*f* et *g* sont des fonction qui décrivent le procédé.  
*t* représente le temps et *a(t)* un vecteur de paramètres permettant d'ajuster le modèle à la réalité expérimentale. Là encore, leurs valeurs ne sont pas a priori constantes au cours du temps. Un exemple typique est l'encrassement des échangeurs de chaleur au cours du temps.

Un tel ensemble n'est pas toujours accessible ou se révèle impossible à résoudre. Le problème de base lié à une démarche de modélisation est de faire des choix :

- relatifs à la structure de base du modèle (définition des entrées, des sorties et des perturbations) ;
- relatifs à la structure des équations.

Le choix des entrées d'un modèle est extrêmement arbitraire. En fait, ce sont les variables dont on est obligé de fixer les valeurs pour rendre le système soluble. Il est clair qu'il est impossible d'aborder le premier item de manière concrète. En effet, ces choix dépendent essentiellement du procédé, de la sensibilité des entrées et de l'importance relative des sorties à l'égard des objectifs du procédé.

La structure d'un modèle peut être choisie arbitrairement (modèle entrées-sorties) ou établie à partir de bilans massiques ou énergétiques (modèle de connaissance ou modèle semi-physique).

## **II.2 Les usages de la modélisation dans le génie des procédés**

La modélisation dans le génie des procédés a des usages multiples. Les modèles sont utilisés pour [41,42] :

1. la recherche et le développement, par exemple la détermination des cinétiques des réactions complexes, l'étude de l'influence des différentes variables opératoires du procédé sur ses performances...etc.
2. éviter d'effectuer des mesures qui peuvent être coûteuses ou qui peuvent perturber le fonctionnement d'un procédé ;
3. l'optimisation du fonctionnement des procédés ;
4. la conception de nouvelles installations ;
5. la synthèse de loi de commande ;
6. le développement de capteurs intelligents ;
7. la formation des opérateurs ;
8. la détection d'anomalies de fonctionnement ;
9. la prévision et la détermination de politique de maintenance ;
10. la gestion des pannes et des stocks ;
11. etc.

### **II.3 Modèle de connaissance ou modèle « boîte blanche »**

En génie des procédés la modélisation est principalement basée sur les lois fondamentales de la physique et de la chimie telles que la conservation de masse, le premier principe de la thermodynamique, les équilibres entre phases, les lois de transfert de matière, les lois de la cinétique chimique...etc [42,43]. Dans leur formulation la plus générale, ceux-ci s'écrivent :

(ce qui rentre) + (ce qui se crée) = (ce qui sort) + (ce qui s'accumule).

Le terme de création correspond par exemple à la production d'espèces chimiques dans un réacteur ou à l'augmentation de l'entropie dans un bilan entropique. Il peut être négatif pour traduire une consommation d'espèce chimique dans un réacteur.

Le terme d'accumulation n'intervient que dans les systèmes instationnaires ; il traduit les fluctuations de la quantité de produit en réponse à un changement de rétention.

Les deux autres termes représentent la contribution des flux convectifs et des phénomènes de transport.

Dans ce type d'approche, les transferts de matière ou de chaleur et les pertes de charge sont définis par des corrélations se référant à des critères hydrodynamiques macroscopiques (nombre de Reynolds) et des constantes physicochimiques.

Ces modèles ainsi obtenus sont dits modèles de connaissance ou modèles de type boîte blanche. Ils se distinguent des modèles entrées-sorties (modèles de type boîte noire) par le fait qu'ils décrivent aussi le fonctionnement interne du système considéré. Ce type de modélisation est le plus répandu en génie des procédés. On le retrouve par exemple dans la modélisation des procédés de transformation tels que le craquage catalytique fluidisé du gasoil [44], le craquage thermique du résidu sous vide du procédé de viscoréduction [45], l'hydrocraquage des fractions lourdes [46]...etc. On le retrouve également dans les procédés de séparation tels que l'extraction liquide-liquide [47], la distillation [42]...etc ; mais aussi pour la modélisation de différents équipements intervenants dans les procédés tels que les échangeurs de chaleur [48].

Cette modélisation est principalement utilisée dans les logiciels commerciaux de simulation. Ces logiciels extrêmement performants s'appuient sur des modèles thermodynamiques calculant les propriétés de mélange à partir de celles des corps purs (équilibre entre phases, entropie, enthalpie) et sur des variables intensives (température, pression et composition) [39]. Parmi ces logiciels on distingue :

- ASPEN, commercialisé par Aspen Tech, est extrêmement complet ; il possède une base de données très importante et orientée vers la chimie ;
- PRO II, distribué par SimSci (Simulation Science Inc), a les mêmes potentiels ; il est plus orienté vers l'industrie pétrolière ;
- HYSIM, distribué par Hyprotech, a la particularité d'avoir été développé spécifiquement pour micro-ordinateur ; il bénéficie donc d'une interface très conviviale ; il est assez orienté vers l'industrie pétrolière.

Ces logiciels sont surtout employés pour effectuer un prédimensionnement des appareils. Les résultats obtenus sont utilisés pour des estimations économiques préliminaires mais ils doivent être validés par des méthodes spécifiques avant toute construction.

Les modèles de connaissance offrent des ressources importantes pour aider à la compréhension d'un procédé et à la définition d'une structure de procédé. Mais cette puissance est redoutable car elle impose une parfaite maîtrise des lois fondamentales qui régissent le génie des procédés. Trop d'erreurs ont pu être faites à cause d'une information insuffisante sur les propriétés

des corps purs [39] ou bien à cause de la complexité des systèmes [42], d'ailleurs parfois l'ingénieur se voit obligé de les simplifier en ignorant quelques phénomènes secondaires tels que les réactions parasites ou les volumes morts dans un réacteur.

## **II.4 Modèle de conduite ou modèle « boîte noire »**

La modélisation de processus complexes de natures très diverses est souvent réalisée par des méthodes purement empiriques, basées sur l'utilisation d'observations expérimentales mathématisées. Ce type de modèle est souvent appelé modèle de conduite, modèle empirique, modèle entrées-sorties ou bien modèle de type boîte noire [40,41,43]. Cette technique est de plus en plus utilisée en génie des procédés, on la retrouve par exemple dans la modélisation des procédés tels que le craquage catalytique fluidisé de l'essence [49], les procédés de traitement des eaux [50], le procédé de pasteurisation [51]...etc mais aussi pour la modélisation des accessoires tels que les échangeurs de chaleur [52].

Un modèle de type boîte noire se réduit donc à une description du procédé dénuée de toute signification physique. De nombreuses formes peuvent lui être données depuis la formulation mathématique pure reliant les entrées et les sorties jusqu'à une description qualitative comme les techniques modernes d'intelligence artificielle l'autorisent [39]. Quel que soit le type d'approche adopté, la priorité est d'atteindre la meilleure qualité de représentation des réponses du système en fonction des entrées, ou plus exactement le rapport optimal qualité/complexité du modèle. En effet il est préférable d'utiliser un modèle moins performant mais plus simple, de manière à avoir une garantie sur sa robustesse.

### **II.4.1 Approche mathématique**

Deux problèmes sont à résoudre lorsque l'on cherche à créer des relations mathématiques entre entrées et sorties d'un système. Le premier consiste à sélectionner les variables influant le plus sur la sortie étudiée. Le second consiste à chercher la forme mathématique la plus appropriée.

Bien que le premier puisse être traité directement par identification, des outils statistiques peuvent aider [39], à partir de l'analyse des valeurs expérimentales, à déterminer les variables ayant le plus d'influence (composantes principales).

La recherche des relations est généralement empirique et se limite, le plus souvent, à l'emploi de relations polynomiales [39,40,43]. Pour identifier le nombre optimal de paramètres (qui doit être le plus faible possible), des tests statistiques sur le résidu entre valeurs expérimentales et valeurs calculées sont nécessaires. Il est souvent intéressant d'utiliser d'autres fonctions mathématiques sélectionnées pour leurs caractéristiques géométriques, citons par exemple la fonction sigmoïde définie par la relation [39]:

$$\frac{\exp(ax) - 1}{\exp(ax) + 1} \dots\dots\dots(\text{II.3})$$

Lorsque le paramètre  $a$  est grand, l'évolution de cette fonction est brutale et tend vers la fonction de Heaviside  $H$ .

$$H(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases} \dots\dots\dots(\text{II.4})$$

## **II.4.2 Approche par intelligence artificielle**

Considérée par certains comme une technique d'intelligence artificielle et par d'autres comme une technique de modélisation de type boîte noire, la modélisation par réseaux neuronaux a pris une extension récente justifiée par les performances de cette approche [49,50,51]. Une telle approche est extrêmement intéressante car, en entreprenant un faible effort en terme d'identifications, des modèles non linéaires multivariés sont construits et conduisent à un excellent accord entre valeurs expérimentales et valeurs calculées pour des processus fortement non linéaires [39].

D'autres outils sont disponibles ou disparaissent complètement les équations. Citons ici, les systèmes experts, la logique floue, la physique qualitative...dont la caractéristique de base est de fournir des réponses de type décisionnel à partir des informations fournies. Les principes de tous ces outils sont très proches. A partir de connaissances obtenues sur le procédé, des règles logiques reliant des faits et des événements sont établies ; lorsque cette relation est stricte, la technique est appelée système expert. Lorsque les conséquences ne le sont pas, nous tombons dans le domaine de la logique floue. Lorsque la définition des événements n'est plus précise mais repérée par de simples niveaux de valeurs, nous abordons le principe de la physique qualitative. A partir de cette base de règles, un outil (moteur d'inférence pour les systèmes experts) explore toutes les possibilités et donne une solution. Comme nous l'avons remarqué ci-avant, ces techniques, bien adaptées pour identifier des décisions, le sont beaucoup moins lorsqu'une approche plus quantitative des interactions entrées/sorties est recherchée [39].

Il est important de préciser que, quelques soient les techniques utilisées, les modèles de type boîte noire sont des modèles de régression, ils ne représentent de manière fiable que le domaine expérimental sur lequel ils ont été calés. Les utiliser en extrapolation est un risque qu'il faut savoir maîtriser [39,43].

## **II.5 Modèle semi-physique ou modèle « boîte grise »**

Entre la boîte noire et le modèle de connaissance se situe le modèle semi-physique, ou modèle boîte grise, qui contient à la fois des équations résultant d'une théorie, et des équations purement empiriques, résultant d'une modélisation de type boîte noire [39,43]. On retrouve cette technique dans la modélisation du procédé de polymérisation de l'éthylène [53].

## **II.6 Conclusion**

Dans ce chapitre nous avons vu brièvement l'utilité de la modélisation dans le génie des procédés. Plusieurs exemples élaborés par différentes techniques ont été cités. La modélisation par les réseaux de neurones est de plus en plus utilisée, en particulier pour les procédés complexes, ce qui nous ramène à étudier en détail cette méthode, à fin de l'exploiter pour la modélisation du procédé d'hydrotraitement du gasoil.

## ***CHAPITRE III***

---

### ***LES RESEAUX DE NEURONES ARTIFICIELS ET LA MODELISATION DE TYPE BOITE NOIRE***

---

### **III.1 Introduction**

Le rêve de créer une machine dotée d'une forme d'intelligence est présent depuis fort longtemps dans l'imagination humaine. Comment l'homme fait-il pour penser, raisonner ou même éprouver des sentiments ? Toutes ces questions ont conduit les chercheurs à l'étude des réseaux de neurones.

Un réseau de neurones artificiel (ou *Artificial Neural Network* en anglais) est un modèle de calcul dont la conception est très schématiquement inspiré du fonctionnement de vrais neurones humains. Les réseaux de neurones sont considérés par certains comme un outil statistiques et pare d'autres comme une technique d'intelligence artificielle ; ils permettent de traiter des problèmes, jadis qualifiés de complexes, en s'appuyant davantage sur la perception que sur le raisonnement logique formel [54,55,56].

Il faut en réalité distinguer deux types de réseaux neuroniques. D'une part, il y a ceux qui sont censés fournir un modèle idéalisé permettant d'étudier le fonctionnement et les propriétés des systèmes nerveux biologiques. D'autre part, il y a les réseaux qui sont des systèmes de traitement dont le but est de réaliser certaines tâches technologiques bien précises [57,58,59]. C'est à ce dernier type de réseaux qu'on s'intéressera dans ce chapitre. Nous aborderons les principales architectures des réseaux que l'on retrouve dans la littérature. Il ne s'agit pas de les étudier toutes, car elles sont trop nombreuses, mais plutôt d'en comprendre les mécanismes internes fondamentaux et de savoir comment et quand les utiliser en particulier pour la modélisation statique de type boîte noire.

### **III.2 Historique**

Deux ingrédients sont à la base de tout avancement des connaissances. Premièrement, il importe de posséder un nouveau concept, ou un nouveau point de vue à propos d'un sujet, qui vient jeter une lumière là où il n'y avait qu'obscurité. Deuxièmement, il importe aussi de posséder des outils technologiques permettant de construire des systèmes concrets. L'histoire des réseaux de neurones est donc tissée à travers des découvertes conceptuelles et des développements technologiques survenus à diverses époques.

Les premières recherches remontent à la fin du 19<sup>e</sup> et au début du 20<sup>e</sup> siècle. Ils consistent en divers travaux en physique, en psychologie et en neurophysiologie par des scientifiques tels que Hermann Von Helmholtz, Ernst Mach et Ivan Pavlov [57,59]. A cette époque, il s'agissait de théories plutôt générales sans modèle mathématique précis d'un neurone. On s'entend pour dire que la naissance du domaine des réseaux de neurones artificiels remonte aux années 1940 avec les travaux de Warren McCulloch et Walter Pitts.

#### **Le début [54,56,57,58,59]**

- **1943** : W.McCulloch et W.Pitts laissent leurs noms à une modélisation du neurone biologique (un neurone au comportement binaire). Ceux 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).
- **1949** : D.Hebb, physiologiste américain, a proposé une théorie fondamentale pour l'apprentissage. La règle qu'il propose connue sous le nom de règle de Hebb, est presque partout présente dans les modèles actuels, même les plus sophistiqués.

#### **Les premiers succès [54,57,59]**

- **1954** : 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.

- **1960** : B.Windrow, 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 rétropropagation du gradient très utilisé aujourd'hui avec les perceptrons multicouches.
- **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 implicitement ces limitations à tous les modèles de réseaux de neurones artificiels. Leur objectif est atteint, il y a abandon financier des recherches dans le domaine (surtout aux USA), d'autant plus qu'on ne disposait pas à l'époque d'ordinateurs suffisamment puissants pour effectuer des calculs complexes.

#### **L'ombre [57,59]**

- **1967-1982** : Toutes les recherches ne sont, bien sur, pas interrompues. Elles se poursuivent, mais déguisées, 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 que S.Grossberg, T.Kohonen, J.Anderson...etc.

#### **Le renouveau [54,57]**

- **1982** : J.J.Hopfield est un physicien reconnu à qui l'on doit le renouveau d'intérêt pour les réseaux de neurones. A cela plusieurs raisons :

Au travers d'un article court, clair et bien écrit, il présente une théorie du fonctionnement et des possibilités des réseaux de neurones. Il faut remarquer la présentation non-conformiste de son article. Alors que les auteurs s'acharnent jusqu'alors à proposer une structure et une loi d'apprentissage, puis à étudier les propriétés émergentes ; J.J.Hopfield fixe préalablement le comportement à atteindre pour son modèle et construit à partir de là, la structure et la loi d'apprentissage correspondant au résultat espéré. Ce modèle est aujourd'hui encore très utilisé pour des problèmes d'optimisation.

D'autre part, entre les mains de ce physicien distingué, la théorie des réseaux de neurones devient respectable. Elle n'est plus réservée exclusivement à un certain nombre de psychologues et neurobiologistes dépassés.

Enfin, une petite phrase, placée en commentaire dans son article initial, met en avant l'isomorphisme de son modèle avec le modèle d'Ising (magnétisme). Cette idée va drainer un flot de physiciens vers les réseaux de neurones artificiels.

Notons qu'à cette date, l'intelligence artificielle est l'objet d'une certaine déception, elle n'a pas répondu à toutes les attentes et s'est même heurtée à de sérieuses limitations. Malgré que les limitations du perceptron mises en avant par M.Minsky ne soient pas levées par le modèle de Hopfield, les recherches sont relancées.

#### **La levée des limitations [54,57,59]**

- **1983** : La machine de Boltzman est le premier modèle connu capable de traiter de manière satisfaisante les limitations recensées dans le cas du perceptron. Mais l'utilisation pratique s'avère difficile car la convergence de l'algorithme est extrêmement longue (les temps de calcul sont considérables).
- **1985** : La rétropropagation du gradient apparaît. C'est un algorithme d'apprentissage adapté aux réseaux de neurones multicouches (aussi appelés perceptrons multicouches). Sa découverte a été réalisée par trois groupes de chercheurs indépendants. Dès cette découverte, nous avons eu 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.

Depuis ce temps, le domaine des réseaux de neurones artificiels bouillonne constamment de nouvelles théories, de nouvelles structures et de nouveaux algorithmes, qui trouvent leurs applications dans divers domaines que nous citerons ultérieurement. Cependant, les réseaux multicouches et la rétropropagation du gradient reste le modèle le plus étudié et le plus productif au niveau des applications. Nous lui avons consacré la plus grande partie de ce chapitre.

### **III.3 Définitions**

#### **III.3.1 Le neurone biologique**

Les modèles de neurones artificiels sont, à l'origine, une imitation du fonctionnement du cerveau. Ce dernier contient chez l'homme environ 10 milliards de neurones, et chacun est connecté à environ 10 000 autres neurones. On voit ainsi sa complexité étonnante [56,60].

Dans le cerveau, les neurones sont reliés entre eux par l'intermédiaire d'axones et de dendrites (figure III.1). En première approche, on peut considérer que ces sortes de filaments sont conducteurs d'électricité et peuvent ainsi véhiculer des messages depuis un neurone vers un autre. Les dendrites représentent les entrées du neurones et son axone sa sortie [61].

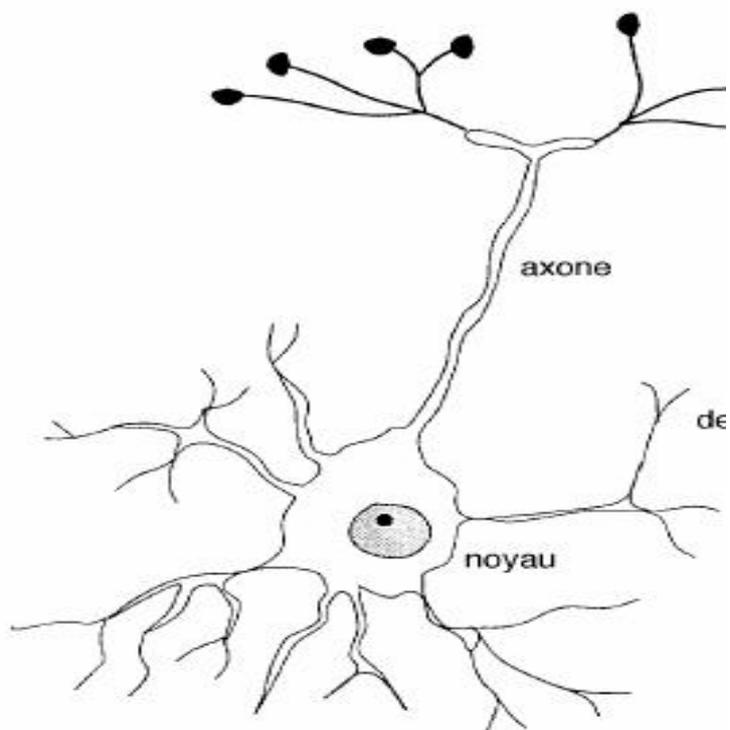


Figure III.1 : Le neurone biologique [60].

Un neurone émet un signal en fonction des signaux qui lui proviennent des autres neurones. On observe en fait au niveau d'un neurone, une intégration des signaux reçus au cours du temps, c'est-à-dire une sorte de sommation des signaux. En général, quand la somme dépasse un certain seuil, le neurone émet à son tour un signal électrique (figure III.2). Ce mécanisme complexifie la façon dont les informations sont transmises : un neurone ne se borne pas à faire passer l'information, il la filtre [56,61].

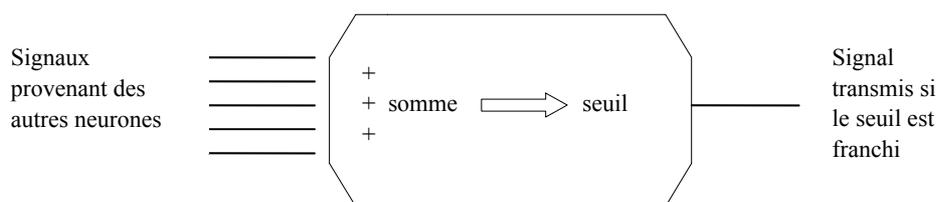


Figure III.2 : Schéma du fonctionnement d'un neurone biologique [56].

La notion de synapse explique la transmission des signaux entre un axone et une dendrite. Au niveau de la jonction (c'est-à-dire de la synapse), il existe un espace vide à travers lequel le signal ne peut pas se propager. La transmission se fait alors par l'intermédiaire de substances chimiques, les neuromédiateurs. Quant un signal arrive au niveau de la synapse, il provoque l'émission de neuromédiateurs qui vont se fixer sur des récepteurs de l'autre côté de l'espace synaptique. Quant suffisamment de molécules se sont fixées, un signal électrique est émis et on a donc une transmission. En fait, suivant le type de la synapse, l'activité d'un neurone peut renforcer ou diminuer l'activité de ces voisins. On parle ainsi de synapse excitatrice ou inhibitrice [61].

Si l'on ramène la contribution d'un neurone au cerveau tout entier, on se rend compte que chacun effectue un travail très simple par rapport au résultat obtenu. En effet, les neurones réalisent des opérations basiques, et pourtant, lorsque l'on en met 10 milliards ensemble, on peut créer une entité pensante ! Cependant, cela n'est pas suffisant : un cerveau ne peut rien faire s'il n'a pas de quoi apprendre. Il a besoin d'informations venant de l'extérieur. C'est pour cela qu'il est relié aux différents organes du corps. Par exemple il reçoit des images provenant des yeux, les sons, les douleurs...etc. Grâce à ces informations il est capable de faire son apprentissage ; lorsqu'une action a provoqué une douleur, il doit changer l'organisation des neurones afin de ne pas répéter la même erreur. L'objectif des réseaux de neurones artificiels est donc de modéliser le fonctionnement des neurones réels, mais aussi de permettre un apprentissage [56].

### **III.3.2 Le neurone artificiel**

Un neurone artificiel est une représentation mathématique d'un neurone biologique. Il possède généralement plusieurs entrées et une sortie qui correspondent respectivement aux dendrites et à l'axone du neurone biologique. Les actions excitatrices et inhibitrices des synapses sont représentées, la plupart du temps, par des coefficients numériques associés aux entrées. Les valeurs numériques de ces coefficients (paramètres ou poids) sont ajustées dans une phase d'apprentissage [62].

Mathématiquement, le neurone artificiel est une fonction algébrique, paramétrée, à valeurs bornées. Les variables de la fonction sont les entrées du neurone et la valeur de la fonction est sa sortie [62,63]. Il est commode de représenter graphiquement un neurone comme indiqué sur la figure III.3. Cette représentation est le reflet de l'inspiration biologique qui a été à l'origine de la première vague d'intérêt pour les neurones artificiels.

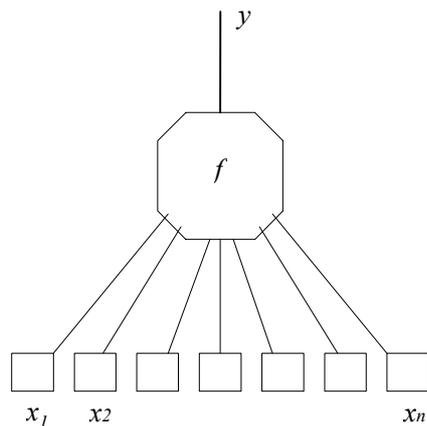


Figure III.3 : Représentation du neurone artificiel [63].

La fonction  $f$  peut être paramétrée de manière quelconque. Deux types de paramétrages sont fréquemment utilisés [63] :

- Les paramètres sont attachés aux entrées du neurone : la sortie du neurone est une fonction non linéaire d'une combinaison des entrées  $\{x_i\}$  pondérées par les paramètres  $(w_i)$ , qui sont alors souvent désignés sous le nom de « poids » ou « poids synaptiques ». Conformément à l'usage, cette combinaison linéaire est appelée « potentiel ». Le potentiel le plus fréquemment utilisé est la somme pondérée, à laquelle s'ajoute un terme constant ou « biais » :

$$v = w_0 + \sum_{i=1}^{n-1} w_i x_i \dots\dots\dots(III.1)$$

La fonction  $f$  est appelée fonction d'activation, il est recommandé d'utiliser pour  $f$  une fonction sigmoïde (c'est-à-dire une fonction sous forme de « S »), telle que la tangente hyperbolique ou la fonction arctangente. Ainsi, dans la très grande majorité des applications, la sortie d'un neurone a pour équation :

$$y = th \left[ w_0 + \sum_{i=1}^{n-1} w_i x_i \right] \dots\dots\dots(III.2)$$

- Les paramètres sont attachés à la non linéarité du neurone : ils interviennent directement dans la fonction  $f$ ; cette dernière peut être une fonction radiale ou RBF (en anglais Radial Basic Function) qui tire son origine de la théorie de l'approximation, ou encore une fonction ondelette qui provient de la théorie du signal.

Par exemple, la sortie d'un neurone RBF à non linéarité gaussienne a pour équation :

$$y = \exp \left[ - \frac{\sum_{i=1}^n (x_i - w_i)^2}{2 w_{n+1}^2} \right] \dots\dots\dots(III.3)$$

Où les paramètres  $w_i, i=1$  à  $n$  sont les coordonnées du centre de la gaussienne, et  $w_{n+1}$  est son écart type.

La différence pratique essentielle entre les deux types de neurones que nous venons d'introduire est la suivante : les neurones tels que les RBF ou les ondelettes ont des non-linéarités locales, qui tendent vers zéro dans toutes les directions de l'espace des entrées ; leur zone d'influence est donc limitée dans l'espace, ce qui n'est pas le cas des neurones à potentiel à fonction d'activation sigmoïde [63].

### **III.3.3 Les réseaux de neurones**

Il n'y a pas de définition universellement acceptée de « réseau de neurones ». On considère généralement qu'un réseau de neurones est constitué d'un grand ensemble d'unités (ou neurones), ayant chacune une petite mémoire locale. Ces unités sont reliées par des canaux de communication (les connexions), qui transportent des données numériques. Les unités peuvent uniquement agir sur leurs données locales et sur les entrées qu'elles reçoivent par leurs connexions [57,64].

La plupart des réseaux ont une certaine capacité d'apprentissage. Cela signifie qu'ils apprennent à partir d'exemples, de même que les enfants apprennent à distinguer les chiens des chats à partir d'exemples. Le réseau peut ensuite dans une certaine mesure être capable de généraliser, c'est-à-dire de produire des résultats corrects sur de nouveaux cas qui ne lui avaient pas été présentés au cours de l'apprentissage [64].

### **III.4 Applications**

Les réseaux de neurones servent aujourd'hui à toutes sortes d'applications dans divers domaines, il n'est donc pas possible de détailler toutes ces applications, d'ailleurs plusieurs livres n'y suffiraient pas. On se contentera de citer brièvement quelques exemples :

- En robotique, par exemple pour la modélisation de l'actionneur hydraulique d'un bras de robot ou le pilotage de robots mobiles [55,65,66].
- La reconnaissance de formes, en effet on utilise les réseaux de neurones pour la lecture automatique des codes postaux et des chèques bancaires, ainsi que pour la reconnaissance de visages [54,59,64,67].
- Dans le contrôle et la commande de processus tel que le système de soudure par laser [63,66,68].
- En biologie, par exemple pour la classification des espèces animales suivant leur ADN (Acide Désoxyribonucléique) ou la détection de maladies pour les plantes [54,69].
- Dans le contrôle de l'environnement, par exemple pour la prédiction de la pollution de l'air [70].
- Dans la modélisation des procédés industriels tels que la distillation du pétrole brut [55], le craquage catalytique fluidisé [49], la pasteurisation [51], la calcination [55]...etc.
- Dans l'économie et les finances, par exemple pour la prédiction des prix, l'évaluation des risques en assurance, l'analyse des marchés boursiers ...etc. [54,67,68,71].
- Dans la synthèse et la reconnaissance de la parole [59,66,67].

### **III.5 Classification des réseaux de neurones**

On distingue deux familles de réseaux de neurones : les réseaux non bouclés et les réseaux bouclés.

#### **III.5.1 Les réseaux de neurones non bouclés**

Un réseau de neurones non bouclé est présenté graphiquement par un ensemble de neurones connectés entre eux, l'information circulant des entrées vers les sorties sans retour en arrière : si l'on représente le réseau comme un graphe dont les nœuds sont les neurones et les arêtes les connexions entre eux-ci, le graphe d'un réseau non bouclé est acyclique : si l'on se déplace dans le réseau, à partir d'un neurone quelconque, en suivant les connexions, on ne peut pas revenir au neurone de départ [55,57,63].

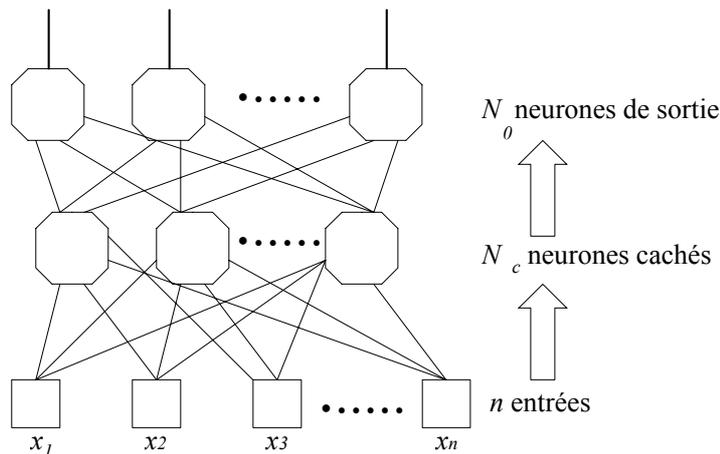


Figure III.4 : Représentation d'un réseau de neurones non bouclé [63].

La très grande majorité des applications des réseaux de neurones mettent en jeux des réseaux à couches, dont un exemple est représenté sur la figure III.4. Ce réseau de neurones non bouclé à  $n$  entrées,  $N_c$  neurones cachés et  $N_o$  neurones de sortie, réalise  $N_o$  fonctions non linéaires de ses  $n$  variables d'entrée par composition des  $N_c$  fonctions algébriques réalisées par ses neurones cachés [63].

Il convient d'insister sur le fait que le temps ne joue aucun rôle fonctionnel dans un réseau de neurones non bouclé : si les entrées sont constantes, les sorties le sont également. Le temps nécessaire pour le calcul de la fonction réalisée par chaque neurone est négligeable et, fonctionnellement, on peut considéré ce calcul comme instantané. C'est pour cette raison que les réseaux non bouclés sont souvent appelés « réseaux statiques », par opposition au réseaux bouclés ou « dynamiques » que nous introduirons plus bas [55,63,65].

Il est important de noter que les réseaux de neurones non bouclés à couches, dont les neurones cachés ont une fonction d'activation sigmoïde, sont souvent appelés perceptrons multicouches ou Multi-Layer Perceptron (MLP) [63].

### III.5.2 Les réseaux de neurones bouclés (ou récurrents)

L'architecture la plus générale pour les réseaux de neurones est celle des réseaux bouclés, dont le graphe des connexion est cyclique : lorsqu'on se déplace dans le réseau en suivant le sens des connexions, il est possible de trouver au moins un chemin qui revient à son point de départ (un tel chemin est désigné sous le terme de cycle). La sortie d'un neurone du réseau peut donc être fonction d'elle-même ; cela n'est évidemment concevable que si la notion de temps est explicitement prise en considération [55,58,63,65].

Ainsi, à chaque connexion d'un réseau bouclé (ou à chaque arête de son graphe) est attaché, outre un poids comme pour les réseaux non bouclés, un retard, multiple entier (éventuellement nul) de l'unité de temps choisie. Une grandeur à un instant donné, ne pouvant pas être fonction de sa propre valeur au même instant, tout cycle du graphe du réseau doit avoir un retard non nul [59,63].

La figure III.5 représente un exemple d'un réseau de neurones bouclé. Les chiffres dans les carrés indiquent le retard attaché à chaque connexion, exprimé en multiple de l'unité de temps (ou période d'échantillonnage)  $T$ . Ce réseau contient un cycle qui part du neurone 3 et revient à celui-ci en passant par le neurone 4 ; la connexion de 4 vers 3 est attachée à un retard de 1 unité de temps [55].

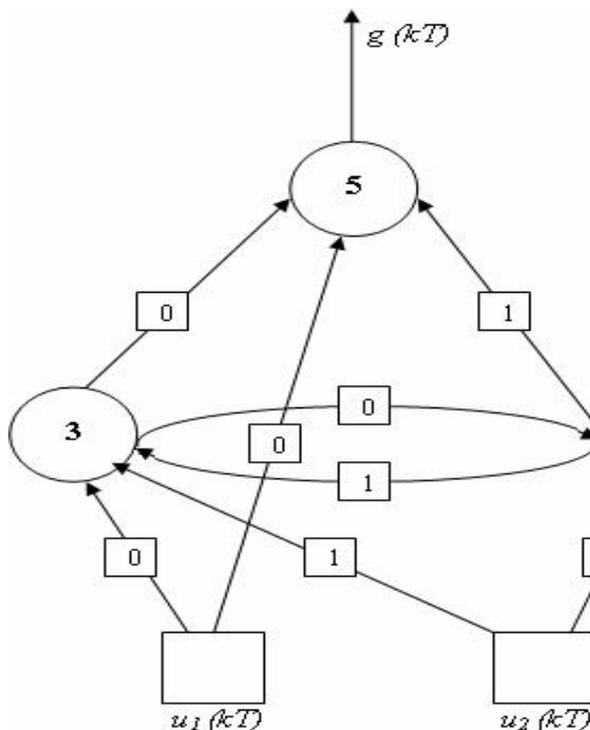


Figure III.5 : Représentation d'un réseau de neurones bouclé [55].

A l'instant  $kT$  : le neurone 3 reçoit en entrée les quantités  $u_1(kT)$ ,  $u_2[(k-1)T]$ ,  $y_4[(k-1)T]$  (où  $k$  est un entier positif et  $y_4(kT)$  est la sortie du neurone 4 à l'instant  $kT$ ), et il calcule sa sortie  $y_3(kT)$ . Le neurone 4 reçoit  $u_2(kT)$  et  $y_3(kT)$  et il calcule sa sortie  $y_4(kT)$ . Le neurone 5 reçoit  $y_3(kT)$ ,  $u_1(kT)$ ,  $y_4[(k-1)T]$ , et il calcule sa sortie qui est la sortie du réseau  $g(kT)$ .

Un réseau de neurones bouclé à temps discret réalise donc une (ou plusieurs) équations aux différences non linéaires, par composition des fonctions réalisées par chacun des neurones et des retards associés à chacune des connections [63].

Tout réseau de neurones bouclé, aussi complexe soit-il, peut être mis sous forme canonique, comportant un réseau de neurones non bouclé dont certaines sorties (les variables d'état) sont ramenées aux entrées par des bouclages de retard unité ( $q^{-1}$ ). Les variables d'état sont un ensemble de variables, en nombre minimal, qui permet de décrire complètement le système à l'instant  $k$  si l'on connaît leurs valeurs initiales et si l'on connaît les valeurs des entrées à tout instant compris entre 0 et  $k-1$ . Le nombre de variables d'état est appelé ordre du système [55,63,65].

Par exemple, le réseau de neurones bouclé représenté sur la figure III.5, peut être mis sous la forme canonique indiquée sur la figure III.6. Ce réseau présente une seule variable d'état  $x$ , il est donc de premier ordre. La partie grise de la forme canonique est un réseau de neurones non bouclé.

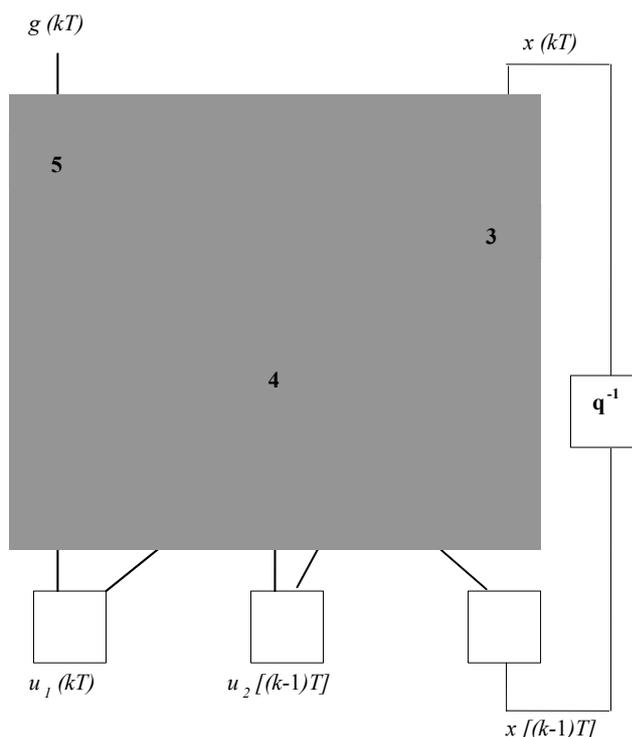


Figure III.6 : Forme canonique du réseau de neurones représenté sur la figure III.5 [55].

A l'instant  $kT$  : le neurone 4 reçoit  $u_2[(k-1)T]$  et  $x[(k-1)T] = y_3[(k-1)T]$ , il calcule donc la quantité  $y_4[(k-1)T]$  ; comme dans la forme non canonique, le neurone 3 reçoit en entrée les quantités  $u_1(kT)$ ,  $u_2[(k-1)T]$  et  $y_4[(k-1)T]$ , il calcule donc la quantité  $y_3(kT)$  ; le neurone 5 reçoit  $y_3(kT)$ ,  $u_1(kT)$  et  $y_4[(k-1)T]$ , il calcule donc sa sortie, qui est la sortie du réseau de neurones  $g(kT)$ . Les deux réseaux sont ainsi bien équivalents fonctionnellement.

## III.6 L'apprentissage des réseaux de neurones

### III.6.1 Définition

Parmi les propriétés désirables pour un réseau de neurones, la plus fondamentale est sûrement la capacité d'apprendre de son environnement, d'améliorer sa performance à travers un processus d'apprentissage. Mais que signifie donc l'apprentissage ? Malheureusement, il n'existe pas de définition générale, universellement acceptée, car ce concept touche à trop de notions distinctes qui dépendent du point de vue que l'on adopte [57,59].

Dans le contexte des réseaux de neurones artificiels, nous adopterons un point de vue pragmatique en proposant la définition suivante : *l'apprentissage est un processus dynamique et itératif permettant de modifier les paramètres d'un réseau en réaction avec le stimuli qu'il reçoit de son environnement. Le type d'apprentissage est déterminé par la manière dont les changements des paramètres surviennent* [59].

Cette définition implique qu'un réseau se doit d'être stimulé par un environnement, qu'il subisse des changements en réaction avec cette stimulation, et que ceux-ci provoquent dans le futur une réponse nouvelle vis-à-vis de l'environnement. Ainsi le réseau peut s'améliorer avec le temps.

Dans la plus part des architectures, l'apprentissage se traduit par une modification de l'efficacité synaptique, c'est-à-dire par un changement dans la valeur des poids qui relient les neurones d'une couche à l'autre. Un ensemble de règles bien définies permettent de réaliser un tel

processus d'adaptation des poids, qui constitue ce qu'on appelle l'algorithme d'apprentissage du réseau. Selon ces algorithmes, il a été défini deux grandes classes d'apprentissage : l'apprentissage supervisé et l'apprentissage non supervisé. Cette distinction repose sur la forme des exemples d'apprentissage [57,59].

### **III.6.2 L'apprentissage supervisé**

Nous avons vu, qu'un réseau de neurone réalise une relation algébrique entre ses entrées et ses sorties. On peut donc affecter à un tel réseau la tâche qui consiste à réaliser une fonction algébrique non linéaire [63], celle-ci peut être :

- Connue analytiquement : le réseau réalise alors une tâche d'approximation de fonction.
- Inconnue analytiquement, mais pour laquelle on dispose de valeurs, en nombre fini, qui sont entachées de bruits si ces valeurs résultent de mesures effectuées sur un processus physique, chimique, économique, biologique ..etc : le réseau réalise alors une tâche de modélisation ou de régression.

C'est essentiellement pour ce dernier type d'applications que sont mis en œuvre les réseaux de neurones à apprentissage supervisé.

On connaît donc, en tous points ou seulement en certains points, les valeurs que doit avoir chaque sortie du réseau en fonction des entrées correspondantes : c'est en ce sens que l'apprentissage est supervisé, métaphore qui signifie qu'un professeur peut fournir au réseau des exemples de ce que celui-ci doit faire [63,64,66].

Les réseaux de neurones non bouclés à apprentissage supervisé sont généralement utilisés pour la modélisation statique de type boîte noire et la classification ; tandis que les réseaux de neurones bouclés à apprentissage supervisé sont utilisés pour la modélisation dynamique de type boîte noire et la commande de processus [63].

### **III.6.3 L'apprentissage non supervisé**

Un réseau de neurone peut également être utilisé dans un but de visualisation ou d'analyse de données : par exemple à partir de données décrites par des vecteurs de grande dimension, trouver une représentation de ces données dans un espace de dimension beaucoup plus faible tout en conservant les proximités ou ressemblances entre ces données. Il n'y a donc pas là de professeur, puisque c'est au réseau de découvrir les ressemblances entre les éléments de la base de données, et de les traduire par une proximité qu'il doit produire. On remarque donc, que contrairement à l'apprentissage supervisé où l'on dispose d'exemples entrées-sorties associés, les exemples pour l'apprentissage non supervisé sont composés uniquement de valeurs des entrées. Les réseaux à apprentissage non supervisé les plus étudiés et utilisés sont les cartes auto-organisatrices ou cartes de Kohonen [57,64,66].

Les réseaux de neurones sont également utilisés pour l'optimisation. Ce type de réseau ne nécessite aucun apprentissage, ils sont dits réseaux sans apprentissage [57,63].

### **III.7 Propriété fondamentale des réseaux de neurones à apprentissage supervisé : L'approximation universelle parcimonieuse**

La propriété d'*approximation universelle parcimonieuse* justifie, dans une large mesure, l'utilisation des réseaux de neurones dans le domaine des Sciences de l'Ingénieur. Le caractère générique des algorithmes d'apprentissage constitue, comme nous l'avons vu dans le paragraphe précédent, la seconde justification de cette utilisation.

#### **III.7.1 Réseaux non bouclés**

Nous avons vu dans le paragraphe précédent que tout réseau de neurones peut être mis sous une forme canonique comprenant un réseau de neurones statique ; les propriétés générales des réseaux de neurones, bouclés ou non, dépendent donc des propriétés des réseaux de neurones non bouclés.

La propriété d'approximation universelle des réseaux de neurones peut s'énoncer comme suit : *pour toute fonction déterministe suffisamment régulière, il existe au moins un réseau de neurones non bouclé, possédant une couche de neurones cachés et un neurone de sortie linéaire, qui réalise une approximation de cette fonction et de ses dérivées successives, au sens des moindres carrés, avec une précision arbitraire.* Ce théorème ne s'applique que si les fonctions d'activation des neurones cachés possèdent des propriétés que nous ne développons pas ici ; les fonctions sigmoïdes et les fonctions à base radiale remplissent les conditions d'application du théorème [55,65].

Il va de soi que la propriété d'approximation universelle n'est pas spécifique aux réseaux de neurones : les polynômes, les séries de Fourier, les fonctions splines, possèdent cette même particularité. Ce qui différencie les réseaux de neurones des autres approximateurs universels usuels, c'est leur *parcimonie* : pour obtenir une approximation d'une précision donnée, les réseaux de neurones utilisent moins de paramètres que les approximateurs usuels. En particulier, le nombre de paramètres varie essentiellement de manière linéaire en fonction du nombre de variables de la fonction que l'on cherche à approcher, alors qu'il varie beaucoup plus rapidement (voire exponentiellement) avec la dimension de l'espace des entrées dans le cas des approximateurs usuels [55,63].

En pratique, les réseaux de neurones sont donc avantageux, par rapport aux approximateurs conventionnels, lorsque l'on cherche à résoudre un problème à plus d'une ou deux variables [55]. Cet avantage est particulièrement intéressant du point de vue du temps de calcul, et surtout du point de vue de la quantité d'informations nécessaires pour le calcul des coefficients :

- du point de vue du temps de calcul, l'estimation des coefficients du réseau (l'apprentissage), est d'autant plus rapide que le nombre de paramètres à calculer est petit ;
- lorsque la fonction que l'on cherche à approcher n'est pas connue analytiquement, mais seulement par l'intermédiaire de l'ensemble d'apprentissage, la taille de l'échantillon (donc le nombre d'exemples) nécessaire croît avec le nombre de poids. Le fait d'utiliser moins de coefficients que les méthodes classiques de régression permet donc une économie en nombre d'exemples, ce qui peut être particulièrement important lorsque l'acquisition des exemples est coûteuse ou lente.

En résumé, l'avantage pratique des réseaux de neurones par rapport aux techniques classiques de modélisation non linéaire réside dans leur capacité à réaliser des modèles de précision équivalente avec moins de données expérimentales (ou dans leur capacité à constituer des modèles plus précis à partir du même nombre d'exemples). De manière générale, les réseaux de neurones

permettent de tirer le meilleur parti des données numériques disponibles, pour construire des modèles à plusieurs entrées.

### **III.7.2 Réseaux bouclés**

La propriété d'approximation universelle des réseaux de neurones prend un sens différent s'il s'agit d'un réseau bouclé. En effet, l'approximation d'un système dynamique par un réseau de neurones bouclé est restreinte à un domaine de l'espace des entrées (comme pour les réseaux non bouclés), mais aussi à un domaine de l'espace des variables d'état, et à un intervalle de temps fini : un tel approximateur peut donc ne pas refléter des caractéristiques fondamentales du processus qu'il est censé approcher, sa stabilité par exemple [55].

### **III.8 Quand utiliser les réseaux de neurones à apprentissage supervisé ?**

Nous avons expliqué précédemment le fondement théorique de l'utilisation des réseaux de neurones à apprentissage supervisé et la propriété d'approcher, de manière parcimonieuse, toute fonction de régression non linéaire suffisamment régulière. Il peut donc être avantageux de mettre en œuvre ces réseaux pour toute application nécessitant de trouver, une relation non linéaire entre des données numériques [63].

Sous quelles conditions peut on utiliser une telle approche ?

- Une première condition, nécessaire mais non suffisante : il faut disposer d'échantillons de taille suffisamment grande, et bien représentatifs.
- Une fois que ces données ont été recueillies, il faut s'assurer de l'intérêt réel d'un modèle non linéaire pour l'application considérée. En effet, la mise en œuvre d'un modèle linéaire (ou affine) est toujours plus simple et moins coûteuse en temps de calcul, que celle d'un réseau de neurones. Par conséquent en l'absence de toute connaissance a priori sur l'intérêt d'un modèle non linéaire, l'ingénieur doit d'abord utiliser les méthodes simples d'élaboration d'un modèle linéaire ; si il s'avère que la précision du modèle est insuffisante bien que tous les facteurs pertinents soient présents dans les entrées du modèle, alors le concepteur de modèle doit envisager la mise en œuvre de modèles non linéaires tels que les réseaux de neurones.
- Si les données sont disponibles, et si l'on s'est assuré qu'un modèle non linéaire est utile, il faut s'interroger sur l'opportunité d'utiliser un réseau de neurones de préférence à une autre famille de fonctions non linéaires, les polynômes par exemple. C'est ainsi qu'intervient la notion de parcimonie abordée dans le paragraphe III.7. En effet lorsque le nombre de variables est supérieur à 2, il est beaucoup plus avantageux d'utiliser les réseaux de neurones qu'un approximateur conventionnel.

Bien entendu, si les données ne sont pas numériques (mais linguistiques par exemple), les réseaux de neurones ne peuvent les traiter directement. Il faut avoir recours à des prétraitements permettant de quantifier ces données [63].

### III.9 Modélisation statique de type boîte noire à l'aide des réseaux de neurones artificiels

#### III.9.1 Formulation du problème

Les propriétés mathématiques décrites dans les sections précédentes sont fondamentales, elles donnent une base solide à l'application des réseaux de neurones multicouches, non bouclés à apprentissage supervisé pour la modélisation de type boîte noire de processus statiques.

Le plus souvent, le problème qui se pose à l'ingénieur est le suivant : on dispose d'un ensemble de variables mesurées  $\{x^k, k=1 \text{ à } N\}$  et d'un ensemble de mesures  $\{y^k(x^k), k=1 \text{ à } N\}$  d'une grandeur  $z_p$  relative à un processus de nature quelconque (physique, chimique, économique, financier...). On suppose qu'il existe une relation mathématique entre le vecteur des variables  $\{x\}$  et la grandeur  $z_p$ , cette relation est appelée fonction de régression, et on cherche à déterminer une approximation de cette fonction, sachant que :

- Les mesures sont en nombre fini ;
- Les mesures sont certainement entachées de bruits.

Cette approximation ne sera pas une approximation uniforme, l'ingénieur cherchera plutôt une approximation au sens des moindres carrés, c'est-à-dire qu'il cherchera une fonction paramétrée  $g(x,w)$ , réalisée à l'aide d'un réseau de neurones, pour laquelle la fonction de coût des moindres carrés  $J(w)$  soit minimale [55,57,59,63,66,72].

$$J(w) = \frac{1}{2} \sum_{k=1}^N [y_p^k - g(x^k, w)]^2 \dots\dots\dots(III.4)$$

L'apprentissage d'un réseau ayant un nombre donné de neurones cachés consiste donc, dans ce cas en une procédure de minimisation qui permet de trouver un vecteur de paramètres  $\{w\}$  satisfaisant.

Ainsi, un réseau de neurones peut, à partir d'exemples, prévoir la valeur que peut prendre une grandeur qui dépend de plusieurs variables qui ne font pas partie de la base d'exemples utilisée pour calculer les coefficients du réseau, cette propriété est connue sous le terme de généralisation. Il faut bien comprendre que cette généralisation est néanmoins limitée, elle ne peut pas s'étendre au-delà des limites de la région de l'espace des entrées dans laquelle se trouvent les exemples qui servent à l'apprentissage [57,59,63,66,72].

La réalisation d'un tel réseau nécessite les étapes suivantes :

- La détermination des entrées pertinentes, c'est-à-dire les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser ;
- La collecte des données nécessaires à l'apprentissage et à l'évaluation des performances du réseau de neurones,
- La détermination du nombre de neurones cachés nécessaires pour obtenir une approximation satisfaisante,
- L'estimation des valeurs des paramètres correspondant à un minimum de la fonction de coût, c'est-à-dire d'effectuer un apprentissage ;
- L'évaluation des performances du réseau de neurones à l'issue de l'apprentissage.

En fonction des résultats obtenus, il peut être nécessaire d'effectuer plusieurs itérations de toutes ou d'une partie de cette procédure. Nous allons aborder successivement ces différents points.

### **III.9.2 Les entrées pertinentes**

Le problème de la détermination des entrées pertinentes se pose de différentes manières selon les applications envisagées.

Si le processus que l'on veut modéliser est un processus industriel conçu par des ingénieurs, le problème est important mais pas crucial, car, en général, on connaît bien les grandeurs qui interviennent et les relations causales entre celles-ci. Il peut être néanmoins utile de faire un choix parmi ces grandeurs, en ne prenant en considération, en entrée du modèle, que celles qui agissent de manière très significative sur le processus (c'est-à-dire celles dont l'effet est plus important que l'incertitude de la mesure) [63].

En revanche, si le processus à modéliser est un processus naturel complexe (écologique par exemple), ou un processus économique, social ou financier, la détermination des entrées pertinentes peut être beaucoup plus délicate. Ce problème n'est pas spécifique aux réseaux de neurones, il se pose pour toutes les techniques de modélisation. Il existe cependant, pour résoudre ce type de problème, plusieurs techniques statistiques telle que la méthode du descripteur sonde que nous ne développerons pas ici [63].

### **III.9.3 La collecte des données**

Nous avons vu que, pour pouvoir effectuer l'apprentissage, il faut disposer de données. Celles-ci doivent être en nombre suffisant, et être suffisamment représentatives des données qui seront présentées au modèle lors de son utilisation. Lorsque la grandeur que l'on veut modéliser dépend de nombreux facteurs, c'est-à-dire lorsque le modèle possède de nombreuses entrées, il n'est pas possible de réaliser un pavage régulier dans tout le domaine de variation des entrées, il faut donc trouver une méthode permettant de réaliser uniquement des expériences qui apportent une information significative pour l'apprentissage du modèle, il faut réaliser un plan d'expériences. Pour les modèles linéaires l'élaboration des plans d'expériences est bien maîtrisée ; ce n'est malheureusement pas le cas pour les modèles non linéaires. Cependant, il existe dans la littérature quelques méthodes qui permettent de construire itérativement un plan d'expériences pour un modèle neuronal, ces méthodes ne seront pas traitées dans ce chapitre [63].

### **III.9.4 Le nombre de neurones cachés**

L'écart entre l'approximation réalisée par un réseau de neurones et la fonction à approcher est inversement proportionnel au nombre de neurones cachés ; malheureusement, ce résultat n'est pas constructif, dans le sens qu'il ne peut que donner des estimations, ou des bornes larges du nombre de neurones cachés nécessaires. Il n'existe pas à l'heure actuelle, de résultat théorique permettant de prévoir le nombre de neurones cachés nécessaires pour obtenir une performance spécifiée du modèle, compte tenu des données disponibles [55,63,73]. Il faut donc nécessairement mettre en œuvre une procédure itérative, selon laquelle le nombre de neurones est déterminé suivant le succès de l'apprentissage [55,63].

L'idée la plus naturelle consisterait à choisir le nombre de neurones cachés le plus grand possible, puisque c'est ce qui assure la plus grande précision à l'approximation uniforme d'une fonction. Ce serait oublié que le problème réel que l'on se pose n'est pas un problème d'approximation uniforme, mais un problème d'ajustement d'une fonction à un nombre fini de points ; il faut donc, non seulement que la fonction réalisée par le réseau passe le plus près possible, au sens des moindres carrés, des points utilisés pour l'apprentissage (l'ensemble de ces points est appelé : ensemble d'apprentissage), mais également qu'il soit capable de généraliser de manière satisfaisante. Si le réseau de neurones possède un nombre de paramètres excessifs, en raison d'un

nombre excessif de neurones cachés, sa sortie peut passer avec une très grande précision par tous les points d'apprentissage, mais fournit des résultats dépourvus de signification entre ces points ; s'il possède un nombre de paramètres trop restreint, le modèle n'est pas suffisamment riche pour rendre compte de la complexité de la fonction de régression inconnue. Ce dilemme est le problème essentiel que doit affronter le concepteur de modèles [55,63].

La figure III.7 présente le résultat de l'apprentissage à partir des mêmes points de mesure, de deux réseaux à une couche cachée de neurones à fonction d'activation sigmoïde. A l'évidence, le réseau le plus parcimonieux en neurones cachés (A) est celui qui présente les meilleures capacités de généralisation, tandis que le réseau qui a trop de paramètres (B) conduit au phénomène classique de surajustement (en anglais *overfitting*). En pratique on s'efforcera toujours de faire en sorte que le nombre de paramètres ajustables soit petit devant le nombre d'exemples [63].

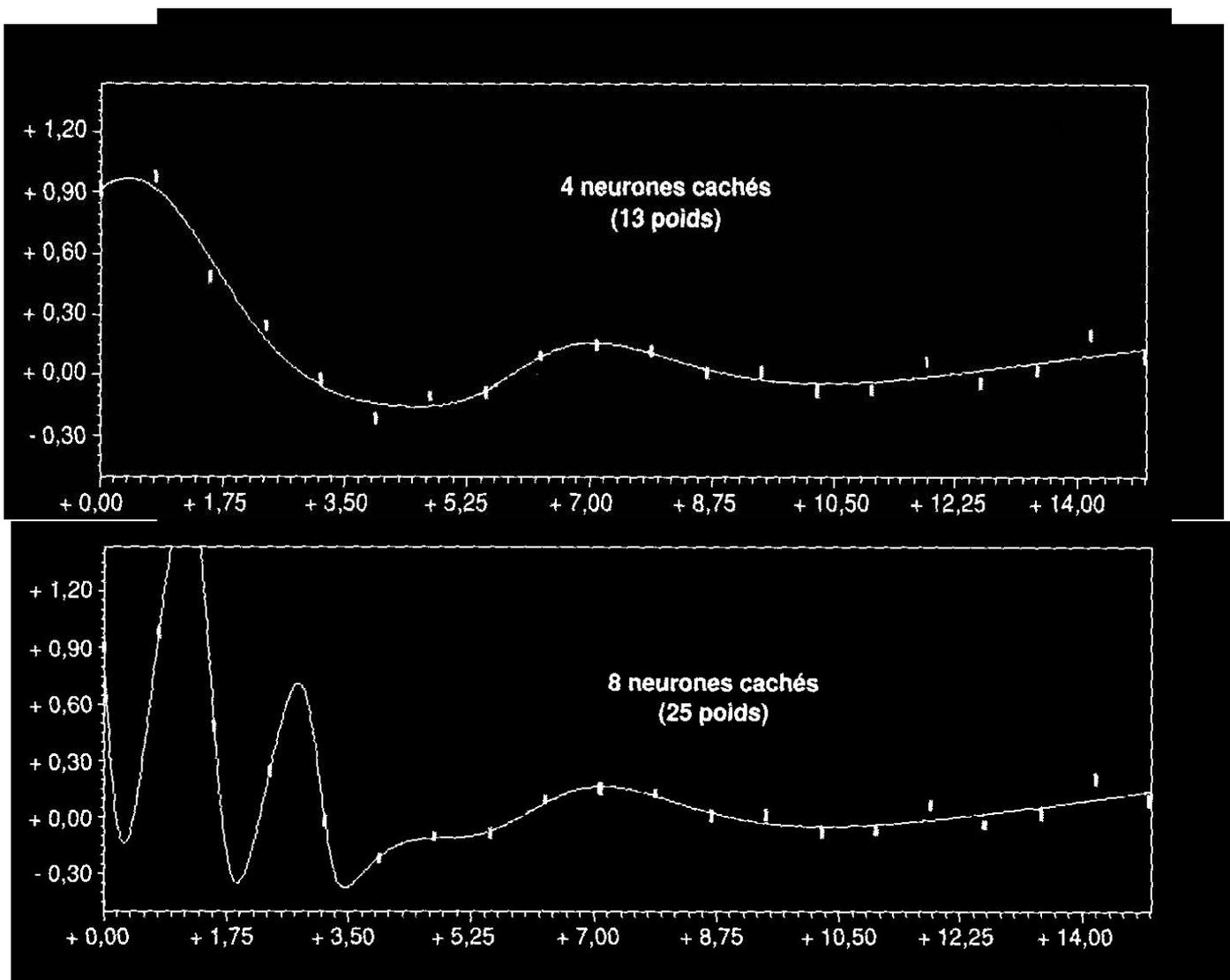


Figure III.7 : Exemple d'un surapprentissage d'un réseau de neurones [63].

Si, au vu des résultats portés sur la figure III.7, le choix entre les deux réseaux est évident, il n'en va pas de même lorsque le modèle possède plusieurs entrées, car il n'est alors pas possible de représenter graphiquement le résultat de manière aussi simple. Il faut donc trouver un critère global de jugement des performances du réseau. Pour ce faire, il convient de constituer, outre l'ensemble d'apprentissage utilisé pour le calcul des poids du réseau, un ensemble de test, constitué de données différentes de celles de l'ensemble d'apprentissage, à partir duquel on estime les performances du réseau après un apprentissage. L'indice de performance le plus fréquemment utilisé est l'erreur quadratique moyenne commise sur l'ensemble de test (ou ensemble d'estimation des performances),

désignée par EQMT [55,63,72] :

$$EQMT = \sqrt{\frac{1}{N_T} \sum_{k=1}^{N_T} [y_p^k - g(x^k, w)]^2} \dots\dots\dots(III.5)$$

Où  $N_T$  est le nombre d'éléments de l'ensemble de test.

Nous nous plaçons ici, pour simplifier, dans le cas très fréquent, d'un réseau de neurones à une sortie. Si le réseau a plusieurs sorties, l'EQMT est la somme des écarts quadratiques moyens relatifs à chacune des sorties.

Cette grandeur doit être comparée à l'erreur quadratique moyenne commise sur l'ensemble d'apprentissage (EQMA).

$$EQMA = \sqrt{\frac{1}{N_A} \sum_{k=1}^{N_A} [y_p^k - g(x^k, w)]^2} \dots\dots\dots(III.6)$$

Où  $N_A$  est le nombre d'exemples de l'ensemble d'apprentissage.

En effet, on voit sur l'exemple présenté sur la figure III.7, que l'EQMA du second réseau (B) est certainement inférieure à celle du premier (A), mais que son EQMT est certainement supérieure à celle du premier. Estimer les performances d'un modèle sur la base de sa seule EQMA serait donc erroné et conduirait systématiquement à concevoir des modèles possédant trop de paramètres, donc souffrant de surajustement.

Remarquons que si la modélisation était parfaite, c'est-à-dire si la sortie du modèle  $g(x, w)$  était identique à la fonction de régression, et si le nombre d'éléments des ensembles d'apprentissage et de test était très grand, l'EQMA et l'EQMT seraient égales, et de l'ordre de grandeur de l'écart-type du bruit de mesure. L'objectif de la modélisation peut donc s'exprimer de la manière suivante : trouver le modèle le plus parcimonieux possible, tel que l'erreur quadratique moyenne sur l'ensemble d'apprentissage, et l'erreur quadratique moyenne sur l'ensemble de test soient de même ordre de grandeur, et aussi petite que possible, c'est-à-dire de l'ordre de grandeur de l'écart type du bruit [63].

### III.9.5 Apprentissage du réseau

Une fois que le nombre de neurones cachés est défini, on doit procéder à l'apprentissage, c'est-à-dire à l'estimation des paramètres du réseau de neurones qui, avec le nombre de neurones cachés dont il dispose, permet d'obtenir une erreur quadratique minimum sur les points de l'ensemble d'apprentissage. L'apprentissage est donc un problème numérique d'optimisation.

Considérons pour simplifier, un réseau de neurones à une sortie  $g(x, w)$ . On dispose d'un ensemble d'apprentissage comprenant  $N$  exemples. Nous avons défini plus haut la fonction de coût des moindres carrés (III.4) :

$$J(w) = \frac{1}{2} \sum_{k=1}^N [y_p^k - g(x^k, w)]^2$$

L'apprentissage consiste à trouver l'ensemble des paramètres  $w$  qui rendent  $J(w)$  minimum.

Si, on met en œuvre des réseaux de neurones tels que les perceptrons multicouches (parcimonieux) dont la sortie n'est pas linéaire par rapport aux paramètres, on doit résoudre un problème d'optimisation non linéaire multivariable. Les méthodes utilisées à cet effet sont toutes

des méthodes de gradient. Ce sont des techniques itératives, qui, à partir d'un réseau muni de poids dont les valeurs sont aléatoires, modifient ces paramètres jusqu'à ce qu'un minimum de la fonction de coût soit atteint, ou qu'un critère d'arrêt soit satisfait. Néanmoins, il faut savoir que dans ce cas la fonction de coût des moindres carrés présente plusieurs minima, parmi lesquels il convient de faire un choix ; cela rend le problème de modélisation plus compliqué, c'est justement le prix que l'on paie pour bénéficier de la propriété de parcimonie [57,63,67,72].

Deux types de méthodes peuvent être mises en œuvre pour minimiser la fonction de coût [63,74] :

- Des méthodes non adaptatives d'apprentissage (off-line training) : elles consistent à estimer les paramètres du modèle par minimisation de la fonction de coût des moindres carrées, qui tient compte simultanément de tous les exemples de l'ensemble d'apprentissage ; l'utilisation d'une telle méthode nécessite évidemment que les N exemples soient disponibles dès le début de l'apprentissage ;
- Des méthodes adaptatives d'apprentissage (on-line training) : elles consistent à modifier les paramètres du modèle successivement en fonction du coût partiel relatif à chaque exemple k :

$$J^k(w) = \frac{1}{2} (y_p^k - g(x^k, w))^2 \quad (III.7)$$

Cette technique est la seule qui puisse être utilisée lorsqu'on désire effectuer l'apprentissage au fur et à mesure de la disponibilité des exemples.

Nous abordons dans ce qui suit l'apprentissage non adaptatif des réseaux de neurones non bouclés, dont la sortie n'est pas linéaire par rapport aux paramètres. La minimisation de la fonction de coût s'effectue donc par des méthodes itératives, chaque itération nécessite la mise en œuvre de deux ingrédients bien distincts :

- L'évaluation du gradient de la fonction de coût.
- La modification des paramètres en fonction de ce gradient.

Nous allons aborder successivement ces deux points, en commençant par le préalable indispensable que constitue la normalisation des grandeurs d'entrée et de sortie.

### **III.9.5.1 Normalisation des entrées et des sorties**

Avant tout apprentissage, il est indispensable de normaliser et de centrer toutes les variables d'entrée. En effet, si des entrées ont des grandeurs très différentes, celles qui sont petites n'ont pas d'influence sur l'apprentissage [63,72,74].

En pratique, il est donc recommandé, pour chaque vecteur d'entrée  $\xi_i$ , de calculer la moyenne  $\mu_i$  et l'écart-type  $\sigma_i$  de ses composantes, et d'effectuer le changement de variables :

$$\xi_i' = (\xi_i - \mu_i) / \sigma_i \quad (III.8)$$

Ou tout autre changement de variables ayant un effet analogue : c'est-à-dire de faire en sorte que la moyenne de chaque variable soit voisine de zéro, et son écart-type de l'ordre de 1 [63].

Il peut également être avantageux, de normaliser les sorties selon la même technique afin d'améliorer la performance des réseaux. En général, on normalise les données de telle sorte qu'elles se trouvent dans l'intervalle [-1 +1] [63,72,74].

### III.9.5.2 Evaluation du gradient de la fonction de coût par rétropropagation

L'évaluation du gradient de la fonction de coût peut être effectuée d'une façon économique à l'aide d'un algorithme appelé algorithme de rétropropagation, devenu tellement populaire qu'il apparaît parfois comme synonyme d'apprentissage de réseaux de neurones [55,57,59,60,67,72,73,74]. En réalité, l'algorithme de rétropropagation n'est pas un algorithme d'apprentissage, mais un ingrédient dans une procédure d'apprentissage [63,65].

Nous considérons ici un réseau de neurones non bouclé avec neurones cachés et un neurone de sortie. L'extension à un réseau qui possède plusieurs neurones de sorties est triviale. Rappelons que le neurone  $i$  calcule une grandeur  $y_i$  qui est une fonction non linéaire de son potentiel  $v_i$ , le potentiel  $v_i$  est une somme pondérée des entrées  $x_j$ , la valeur de l'entrée  $x_j$  étant pondérée par un paramètre  $w_{ij}$ .

$$y_i = f\left(\sum_{j=1}^n w_{ij} x_j\right) = f(v_i) \tag{III.9}$$

Les  $n_i$  entrées du neurone  $i$  peuvent être soit les sorties d'autres neurones, soit les entrées du réseau. Dans toute la suite,  $x_j$  désignera donc indifféremment soit la sortie  $y_j$  du neurone  $j$ , soit l'entrée  $j$  du réseau.

La fonction de coût dont on cherche à évaluer le gradient est de la forme :

$$J(w) = \sum_{k=1}^M [y_p^k - g(x^k, w)]^2 = \sum_{k=1}^N J^k(w) \tag{III.10}$$

Pour évaluer son gradient, il suffit donc d'évaluer le gradient du coût partiel  $J(w)$  relatif à l'observation  $k$ , et de faire ensuite la somme sur tous les exemples.

L'algorithme de rétropropagation consiste essentiellement en l'application répétée de la règle des dérivées composées [55,57,59,60,63,65,67,72,73,74]. On remarque tout d'abord que la fonction de coût partiel ne dépend du paramètre  $w_{ij}$  que par l'intermédiaire de la valeur de la sortie du neurone  $i$ , qui est elle-même fonction uniquement du potentiel du neurone  $i$ , on peut donc écrire :

$$\left(\frac{\partial J^k}{\partial w_{ij}}\right)_k = \left(\frac{\partial J^k}{\partial v_i}\right)_k \left(\frac{\partial v_i}{\partial w_{ij}}\right)_k = \delta_i^k x_j^k \tag{III.11}$$

Où

$\left(\frac{\partial J^k}{\partial v_i}\right)_k$  désigne la valeur du gradient du coût partiel par rapport au potentiel du neurone  $i$  lorsque les entrées du réseau sont celles qui correspondent à l'exemple  $k$ ,

-  $\left(\frac{\partial v_i}{\partial w_{ij}}\right)_k$  désigne la valeur de la dérivée partielle du potentiel du neurone  $i$  par rapport au paramètre  $w_{ij}$  lorsque les entrées du réseau sont celles qui correspondent à l'exemple  $k$ ,

-  $x_j^k$  est la valeur de l'entrée  $j$  du neurone  $i$  lorsque les entrées du réseau sont celles qui correspondent à l'exemple  $k$ .

Il reste donc à évaluer les quantités  $\delta_i^k$  présentes dans le membre de droite de l'équation. Nous allons voir que ces quantités peuvent être avantageusement calculées d'une manière récursive

en menant les calculs depuis la (ou les) sortie(s) du réseau vers ses entrées. En effet :

- pour le neurone de sortie  $i$  :

$$\delta_i^k = \left( \frac{\partial J}{\partial v_i} \right)_k = \left( \frac{\partial}{\partial v_i} \left[ \frac{1}{2} (y_i^k - g(x, w))^2 \right] \right)_k = -2 g(x^k, w) \left( \frac{\partial g(x, w)}{\partial v_i} \right)_k \dots \dots \dots (III.12)$$

Or, la sortie du modèle est la sortie  $y_i$  du neurone de sortie ; cette relation s'écrit donc :

$$\delta_i^k = -2 g(x^k, w) f'(v_i^k) \dots \dots \dots (III.13)$$

Où  $f'(v_i^k)$  désigne la dérivée de la fonction d'activation du neurone de sortie lorsque les entrées du réseau sont celles de l'exemple  $k$ .

Si, comme c'est le cas lorsque le réseau est utilisé en régression, le neurone de sortie est linéaire, l'expression se réduit à :

$$\delta_i^k = -2 g(x^k, w) \dots \dots \dots (III.14)$$

- pour un neurone caché  $i$  : la fonction de coût ne dépend du potentiel du neurone  $i$  que par l'intermédiaire des potentiels des neurones  $m$  qui reçoivent la valeur de la sortie du neurone  $i$ , c'est-à-dire de tous les neurones qui, dans le graphe des connexions du réseau, sont adjacents au neurone  $i$ , entre ce neurone et la sortie :

$$\delta_i^k = \left( \frac{\partial J}{\partial v_i} \right)_k = \sum_m \left( \frac{\partial J}{\partial v_m} \right)_k \left( \frac{\partial v_m}{\partial v_i} \right)_k = \sum_m \delta_m^k \left( \frac{\partial v_m}{\partial v_i} \right)_k \dots \dots \dots (III.15)$$

Or  $v_m^k = \sum_i w_{mi} x_i^k = \sum_i w_{mi} f(v_i^k)$  .....(III.16) d'où  $\left( \frac{\partial v_m}{\partial v_i} \right)_k = w_{mi} f'(v_i^k)$  .....(III.17)

Finalement on obtient la relation :

$$\delta_i^k = \sum_m \delta_m^k w_{mi} f'(v_i^k) = \sum_m \delta_m^k w_{mi} \dots \dots \dots (III.18)$$

Ainsi les quantités  $\delta_i^k$  peuvent être calculées récursivement, en parcourant le graphe des connexions à l'envers, depuis la (les) sorties(s) vers les entrées du réseau (ce qui explique le terme rétropropagation).

Une fois que les gradients des coûts partiels ont été calculés, il suffit d'en faire la somme pour obtenir le gradient de la fonction de coût total.

En résumé, l'algorithme de rétropropagation comporte deux phases pour chaque exemple  $k$  :

- une phase de propagation, au cours de laquelle les entrées correspondant à l'exemple  $k$  sont utilisées pour calculer les sorties et les potentiels de tous les neurones,
- une phase de rétropropagation, au cours de laquelle sont calculées les quantités  $\delta_i^k$ .

Une fois que ces quantités sont disponibles, on calcule les gradients des coûts partiels puis le gradient du coût total, respectivement, par les relations :

$$\left( \frac{\partial J}{\partial w_{ij}} \right)_k = \delta_i^k x_j^k \dots \dots \dots (III.19) \quad \text{et} \quad \left( \frac{\partial J}{\partial w_{ij}} \right)_k = \sum_k \left( \frac{\partial J}{\partial w_{ij}} \right)_k \dots \dots \dots (III.20)$$

Une mythologie s'est développée, selon laquelle l'apprentissage des réseaux possédant des neurones cachés a été rendu possible par l'invention de la rétropropagation. En réalité, il est tout à

fait possible d'évaluer le gradient de la fonction de coût par un calcul plus simple dans son principe. Ce calcul s'effectue dans le sens direct, en évaluant les gradients à partir des entrées, vers les sorties. Mais cette technique, par rapport à la rétropropagation, est bien plus coûteuse en temps de calcul ; c'est la raison pour laquelle, elle est peu utilisée [63].

**III.9.5.3 Modification des paramètres en fonction du gradient de la fonction de coût**

Dans la section précédente, nous avons vu comment on évalue le gradient de la fonction de coût par rapport aux paramètres du modèle, à chaque itération du processus d'apprentissage. Une fois que l'on dispose de cette évaluation, on effectue une modification des poids, afin d'approcher d'un minimum de la fonction de coût. Nous allons examiner quelques algorithmes itératifs de modification des paramètres du modèle mais avant cela nous allons présenter la manière d'initialiser ces paramètres.

**a) Initialisation des paramètres**

L'initialisation des poids se fait d'une manière aléatoire. Ils sont initialisés aléatoirement, à de petites valeurs mais dans un intervalle prédéfini, les valeurs des poids et des biais sont comprises entre [-1, +1]. Cette technique est utilisée afin de permettre une convergence rapide au réseau [57,72,73].

**b) La méthode du gradient simple**

La méthode du gradient simple consiste à modifier les paramètres par la formule suivante, à l'itération  $i$  de l'apprentissage [63] :

$$w(i) = w(i-1) - \mu_i \nabla J(w(i-1)) \quad \text{avec } \mu_i > 0 \quad \dots\dots\dots(III.21)$$

La direction de descente est donc simplement opposée à celle du gradient : c'est en effet la direction suivant laquelle la fonction de coût diminue le plus rapidement. La quantité  $\mu_i$  est appelée pas du gradient ou pas d'apprentissage.

Cette méthode est simple, mais elle présente de nombreux inconvénients [63] :

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

Pour pallier le premier inconvénient, de très nombreuses heuristiques ont été proposées, avec des succès divers. Les méthodes de recherche unidimensionnelle, fondées sur des principes solides, sont recommandées [63].

Pour faire face aux deux autres problèmes, on utilise des méthodes du second ordre qui, au

lieu de modifier les coefficients uniquement en fonction du gradient de la fonction de coût, utilisent les dérivées secondes de cette dernière. Certaines de ces méthodes font également intervenir un paramètre  $\mu$  susceptible d'être choisi à l'aide de méthodes de recherche unidimensionnelle [63].

**c) Les méthodes du gradient du second ordre**

Toutes les méthodes du second ordre sont dérivées de la méthode de Newton, dont nous présentons le principe ci après [59,63,72,73].

Le développement de Taylor d'une fonction  $J(w)$  d'une seule variable  $w$  au voisinage d'un minimum  $w^*$  est donné par la relation (III.22) car le gradient de la fonction de coût est nul au minimum.

$$J(w) = J(w^*) + \frac{1}{2}(w - w^*)^2 \left( \frac{d^2 J}{dw^2} \right)_{w=w^*} + o(w^3) \dots\dots\dots(III.22)$$

Une approximation du gradient de la fonction de coût au voisinage du minimum est obtenue aisément en dérivant la relation précédente par rapport à  $w$  :

$$\frac{dJ}{dw} = (w - w^*) \left( \frac{d^2 J}{dw^2} \right)_{w=w^*} \dots\dots\dots(III.23)$$

Par conséquent, lorsque la variable  $w$  est au voisinage de  $w^*$ , on pourrait atteindre ce minimum en une seule itération si l'on connaît la dérivée seconde de la fonction à son minimum : il

suffirait pour cela de modifier la variable  $w$  de la quantité  $\Delta w = \frac{(dJ/dw)}{\left( \frac{d^2 J}{dw^2} \right)_{w=w^*}} \dots\dots\dots(III.24)$

Le même raisonnement s'applique à une fonction de plusieurs variables, la dérivée seconde étant remplacée par la matrice hessienne  $H(w)$  de la fonction à optimiser, de terme général

$\frac{\partial^2 J}{\partial w_i \partial w_j}$  : pour atteindre le minimum de la fonction de coût en une itération, il suffirait d'appliquer

au vecteur des poids la modification suivante (sous réserve que la matrice hessienne soit inversible):

$$\Delta w = -H(w^*)^{-1} \nabla J(w) \dots\dots\dots(III.25)$$

Cette dernière formule n'est évidemment pas applicable en pratique, puisque le vecteur  $w$  n'est pas connu. Néanmoins, elle suggère plusieurs techniques qui mettent en œuvre une approximation itérative de la matrice hessienne (ou de son inverse). Parmi ces méthodes, on distingue la méthode de Broyden-Fletcher-Goldfarb-Shanno (algorithme BFGS) et l'algorithme de Levenberg-Marquardt (LM). C'est deux méthodes sont, en effet, les plus performantes et les plus utilisées mais elle ne sont pas spécifiques aux réseaux de neurones : ce sont des méthodes très générales d'optimisation, qui existaient bien avant l'introduction des réseaux de neurones [63,72].

**d) Choix de la méthode**

En premier lieu, il ne faut pas utiliser la méthode du gradient simple et ses variantes, dont les temps de convergence (en nombre d'itérations) sont supérieurs de plusieurs ordres de grandeur à ceux des méthodes du second ordre. L'utilisation de la méthode du gradient simple ne peut se justifier que dans le cas où l'on utilise de très gros réseau (plusieurs milliers de poids), ce qui peut se produire pour des problèmes de classification dans lesquels les objets à classer ont de très nombreux descripteurs (typiquement, des images représentées par des descripteurs de bas niveau)

[63].

Le choix entre les méthodes de BFGS et de Levenberg-Marquardt repose essentiellement sur des considérations relatives au temps de calcul et à la taille de la mémoire. La méthode de BFGS nécessite de choisir le moment où l'on passe du gradient simple à BFGS ; il n'y a pas à cet effet, de règle fondée sur des arguments théoriques ; quelques tâtonnements sont parfois nécessaires pour trouver le bon nombre d'itérations (typiquement, une centaine), mais celui-ci n'est pas critique. La méthode de Levenberg-Marquardt ne présente pas cet inconvénient, mais elle devient lourde pour des gros réseaux (une centaine de paramètres), en raison de l'inversion de matrice nécessaire à chaque itération. Globalement, on a donc intérêt à choisir la méthode de Levenberg-Marquardt si le réseau est petit, et celle de BFGS dans le cas contraire [63,72,73].

**e) L'algorithme de Levenberg-Marquardt**

L'algorithme de Levenberg-Marquardt (LM) consiste à modifier les paramètres par la formule [63,72,73] :

$$w(i) = w(i-1) - [H(w(i-1) + \mu_i I)]^{-1} \nabla J(w(i-1)) \dots \dots \dots (III.26)$$

Pour de petites valeurs du pas  $\mu_i$ , la méthode de Levenberg-Marquardt s'approche de celle de Newton. Inversement, pour de grandes valeurs de  $\mu_i$ , l'algorithme de Levenberg-Marquardt est équivalent à l'application de la règle du gradient simple avec un pas de  $1/\mu_i$  [63,72,73]. L'application de cet algorithme nécessite l'inversion de la matrice  $[H(w(i-1) + \mu_i I)]$ . L'expression exacte de la matrice hessienne de la fonction de coût totale  $J(w)$  est :

$$H(w(i)) = \sum_{k=1}^N \left( \frac{\partial e^k}{\partial w(i)} \right) \left( \frac{\partial e^k}{\partial w(i)} \right)^T + \sum_{k=1}^N \frac{\partial^2 e^k}{\partial w(i)^2} \dots \dots \dots (III.27)$$

Le second terme de cette expression étant proportionnel à l'erreur, on peut le négliger en première approximation, ce qui fournit une expression approchée :

$$\tilde{H}(w(i)) = \sum_{k=1}^N \left( \frac{\partial e^k}{\partial w(i)} \right) \left( \frac{\partial e^k}{\partial w(i)} \right)^T = \sum_{k=1}^N \left( \frac{\partial y^k}{\partial w(i)} \right) \left( \frac{\partial y^k}{\partial w(i)} \right)^T \dots \dots \dots (III.28)$$

Plusieurs techniques sont envisageables pour l'inversion de la matrice  $H + \mu_i I$  [[63] :

**Inversion indirecte**

Un lemme d'inversion permet de calculer récursivement la matrice inverse. En effet, soient  $A, B, C$  et  $D$  quatre matrices. On a la relation suivante :

$$(A + B C D)^{-1} = A^{-1} + A^{-1} B C (I + D A^{-1} B)^{-1} D A^{-1} \dots \dots \dots (III.29)$$

Par ailleurs, en posant :  $\zeta^k = \frac{\partial y^k}{\partial w(i)} \dots \dots \dots (III.30)$

On peut construire récursivement la matrice  $\tilde{H}$  en définissant des matrices partielles  $\tilde{H}^k$ , de dimension  $(k, k)$  par :  $\tilde{H}^k = \tilde{H}^{k-1} + z^k (z^k)^T$  avec  $k = 1, \dots, N$  .....(III.31)

On a bien :  $\tilde{H}^N = \tilde{H} \dots \dots \dots (III.32)$

Si l'on applique le lemme d'inversion à la relation précédente en choisissant  $A = H$ ,  $B = \zeta^k$ ,  $\tilde{C} = I$  et  $D = (\zeta^k)^T$ , on obtient la relation suivante :

$$(\tilde{H}^k)^{-1} = (\tilde{H}^{k-1})^{-1} - \frac{(\tilde{H}^{k-1})^{-1} \zeta^k (\zeta^k)^T (\tilde{H}^{k-1})^{-1}}{1 + (\zeta^k)^T (\tilde{H}^{k-1})^{-1} \zeta^k} \dots \dots \dots (III.33)$$

En prenant, à la première étape  $k = 1$ ,  $H = \mu_1 I$  on obtient, à l'étape  $N$  :

$$(\tilde{H}^N)^{-1} = [\tilde{H} + \mu_N I]^{-1} \dots \dots \dots (III.34)$$

**Inversion directe**

Plusieurs méthodes directes d'inversion existent. Comme l'algorithme est itératif, et que la procédure de recherche du pas nécessite souvent plusieurs inversions de matrice, on a intérêt à utiliser une méthode qui n'engage pas trop de calculs. Comme l'approximation de la matrice hessienne augmentée de  $\mu_i I$  reste une matrice symétrique définie comme positive, il est avantageux d'utiliser la méthode de Cholesky.

Comme pour l'algorithme du gradient simple, le pas  $\mu_i$  doit être ajusté à chaque itération. Une méthode de recherche unidimensionnelle peut être utilisée à cet effet [63].

Voici une stratégie très simple, couramment utilisée pour la méthode de Levenberg-Marquardt [63,72]. Soit  $r > 1$  (généralement égale à 10) un facteur d'échelle. Au début de l'algorithme, on initialise  $\mu_0$  (généralement à 0,001). À l'itération  $i$  de l'algorithme :

1. Calculer  $J(w(i))$  avec  $\mu_i$  déterminé à l'étape précédente.
2. Si  $J(w(i)) < J(w(i-1))$ , alors accepter le changement de paramètres et multiplier  $\mu_i$  par  $r$ .
3. Sinon, récupérer  $w(i-1)$  et diviser  $\mu_i$  par  $r$ . Répéter cette dernière étape jusqu'à ce qu'une valeur de  $\mu_i$  correspondant à une décroissance de  $J$  soit trouvée.

Cette procédure présente l'avantage de nécessiter peu d'inversions de matrice à chaque itération de l'algorithme. En revanche, le choix du pas initial a une influence sur la vitesse de convergence de l'algorithme.

**III.9.5.4 Le Momentum**

Actuellement, l'apprentissage des réseaux est pratiquement toujours utilisé avec le momentum. Cette méthode consiste à ajouter à la modification des paramètres  $\Delta w$ , en plus du terme du gradient, un terme proportionnel à la variation des poids à l'itération précédente  $\lambda[w(i-1) - w(i-2)]$ , ce terme est appelé momentum [59,63,72,73]. La constante du momentum  $\lambda$  est utilisée pour pondérer l'effet de ce terme ; sa valeur est comprise entre 0 et 1 [72].

Le momentum est un moyen efficace pour éviter les oscillations du gradient, il contribue à sa stabilisation en ralentissant les changements de direction. Il permet également l'accélération du processus d'apprentissage [59,63,72,73].

**III 9.5.5 Arrêt prématuré**

L'arrêt prématuré ou "*early stopping*" en anglais, est une méthode qui permet l'amélioration

de la généralisation du modèle. Cette technique consiste à diviser aléatoirement les données en trois ensembles [63,72].

Le premier est bien évidemment l'ensemble d'apprentissage, qui sert à calculer le gradient et à mettre à jours les paramètres du réseau.

Le second est appelé ensemble de validation ; l'erreur sur cet ensemble est suivie tout au long du processus d'apprentissage. L'erreur de validation, comme celle de l'apprentissage, décroît au cours de la phase initiale de l'apprentissage. Cependant, lorsque le réseau commence à s'ajuster trop finement aux données d'apprentissage, l'erreur sur l'ensemble de validation augmente. Quand l'erreur de validation augmente pendant un nombre déterminé d'itérations, on arrête prématurément l'apprentissage, c'est-à-dire avant convergence complète de l'algorithme, ainsi le surajustement est limité. Les poids et les biais retenus sont ceux qui correspondent au minimum de l'erreur de validation [63,72].

Le troisième ensemble est l'ensemble de test, il n'est pas utilisé au cours de l'apprentissage, il sert à comparer entre les différents réseaux et à choisir la structure optimale du modèle.

En général, l'ensemble d'apprentissage est constitué de 60% des données, tandis que les ensembles de validation et de test sont constitués chacun de 20% [72].

### **III.10 Conclusion**

Les réseaux de neurones artificiels, tels que nous les avons définis, possèdent une propriété remarquable qui est à l'origine de leur intérêt pratique dans des domaines très divers : ce sont des approximateurs universels parcimonieux.

L'apprentissage est vraisemblablement la propriété la plus intéressante des réseaux neuronaux, cependant il existe plusieurs algorithmes utilisés pour faire l'apprentissage des réseaux multicouches, la méthode retenue est celle de la rétropropagation du gradient avec momentum. La modification des poids s'effectue à l'aide des méthodes de second ordre, l'algorithme de Levenberg-Marquardt est le plus rapide et assure la meilleure convergence vers un minimum de l'erreur quadratique pour les problèmes d'approximation de fonctions où le nombre de poids du réseau est inférieur à cent. La technique de l'arrêt prématuré permet d'éviter le phénomène de surajustement et permet ainsi au modèle une meilleure généralisation.

Il serait donc très intéressant d'exploiter les réseaux de neurones multicouches pour la modélisation de type boîte noire des différents procédés de raffinage. Cette approche est justement choisie pour la modélisation du procédé d'hydrotraitement du gasoil.

## ***CHAPITRE IV***

---

### ***APPLICATION DES RESEAUX DE NEURONES POUR LA MODELISATION DU PROCEDE D'HYDROTRAITEMENT DU GASOIL***

---

## **IV.1 Introduction**

Le procédé d'hydrotraitement du gasoil, vu son importance dans le domaine du raffinage, a été modélisé à plusieurs reprises pour différentes charges [75,76,77,78,79,80,81,82]. Ces modèles sont tous des modèles de connaissance, ils sont basés surtout sur les lois de la cinétique chimique et les équations de transfert de matière entre les phases existantes dans le réacteur d'hydrotraitement.

Ils permettent de prédire les caractéristiques des produits (teneurs en soufre, aromatiques, oléfines...etc.) en fonction des différentes variables opératoires du procédé telles que la température, la VVH et la pression partielle d'hydrogène. Ces modèles sont souvent utilisés pour la simulation d'unités industrielles d'hydrotraitement. Ils présentent néanmoins un grand inconvénient. En effet, chaque modèle est spécifique au gasoil pour lequel il a été établi. Lors d'un changement dans la composition de la charge le modèle en question fournit des résultats dépourvus de tous sens. Ceci est essentiellement dû aux hypothèses établies par les auteurs pour simplifier les équations mathématiques qui décrivent le procédé.

Il serait donc très intéressant, de mettre en œuvre un modèle qui pourrait prédire les caractéristiques des produits pour tous les gasoils hydrotraités. Une telle approche est difficilement réalisable à l'aide des lois fondamentales qui régissent le génie des procédés. On propose donc, pour atteindre cet objectif, un modèle de type boîte noire basé uniquement sur des données expérimentales. La technique adoptée pour l'élaboration de ce modèle est bien entendue la modélisation par réseaux neuronaux. Les réseaux de neurones artificiels, comme nous l'avons constaté dans le chapitre III, présentent certaines propriétés qui rendent leur utilisation plus avantageuse que celle des autres méthodes.

Le logiciel MATLAB (version 7.3) présente une interface graphique spéciale pour les réseaux de neurones artificiels, qui rend leur mise en œuvre très facile. D'ailleurs, aucune programmation n'est nécessaire, lorsqu'il s'agit de modèles classiques tels que les perceptrons multicouches.

Nous allons dans ce chapitre présenter, les différentes étapes nécessaires pour la modélisation de type boîte noire du procédé d'hydrotraitement du gasoil, à l'aide d'un perceptron multicouche.

## **IV.2 Présentation des données**

### **IV.2.1 Déroulement des expériences**

Les données que nous allons utiliser pour la modélisation du procédé d'hydrotraitement nous ont été fournies par la raffinerie de Skikda (RA<sub>1</sub>/K). Les expériences se sont déroulées dans une unité pilote d'un grand laboratoire et présentées à la raffinerie au cours d'un appel d'offre. Un exemple d'une unité pilote d'hydrotraitement est représenté sur la figure IV.1.

Trois gasoils légers de distillation atmosphérique, ayants différentes teneurs en soufre, ont été testés sous des conditions opératoires variables (température, pression, VVH), le rapport hydrogène/charge est maintenu égale à 500ml.ml<sup>-1</sup> toutes les expériences. Certaines expériences sont effectuées en un seul étage et d'autres en deux. Les caractéristiques de ces gasoils sont représentées sur le tableau IV.1.

La quantité de catalyseur employé pour l'hydrotraitement de ces gasoils est de 50cm<sup>3</sup>. Le catalyseur se trouve dans le réacteur pilote entre deux couches de 20cm<sup>3</sup> de silice. La silice permet une meilleure distribution des phases liquide et gazeuse sur le lit catalytique. Le catalyseur utilisé est de type CoMo/alumine, ses caractéristiques sont représentées sur le tableau IV.2.

Les teneurs en soufre des charges et des produits sont déterminées par fluorescence X.

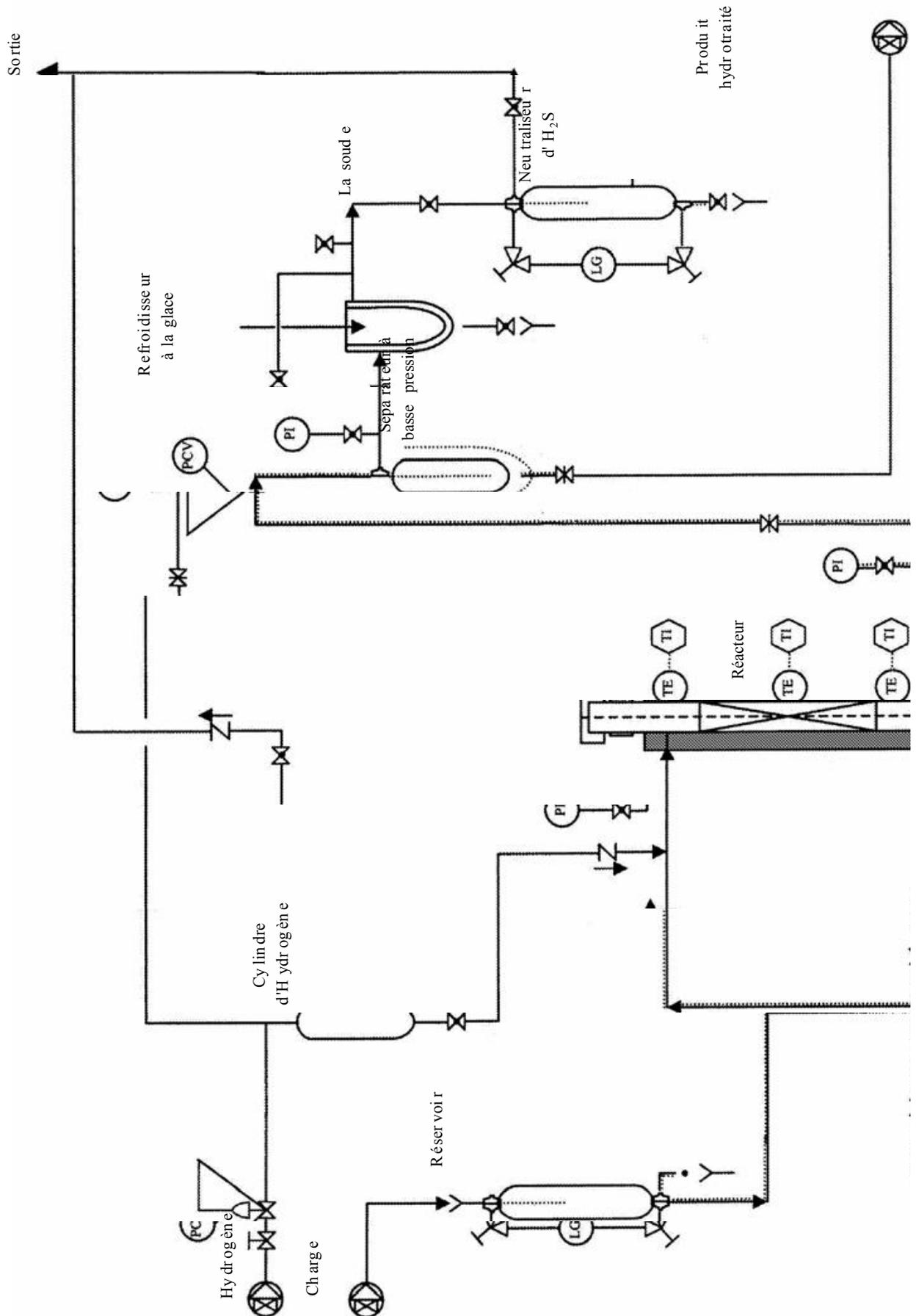


Figure IV.1 : Exemple d'une unité pilote d'hydrotraitement [30].

Gasoil	A	B	C
Densité	0,864	0,846	0,838
Teneur en soufre (ppm)	12000	10000	7500
Distillation ASTM (°C)			
10% vol	259	255	251
50% vol	292	289	286
90% vol	341	337	333
PF	357	354	351

Tableau IV.1 : Caractéristiques des charges.

---

Composition chimique (%)	
Co	4.2
Mo	16.7
S	4.6
Co/Mo rapport atomique	0.4
Masse volumique (g/cm <sup>3</sup> )	0.87
Surface spécifique (m <sup>2</sup> /g)	130
Porosité (cm <sup>3</sup> /g)	0.19
Diámetro moyen des pores (nm)	4.5

---

Tableau IV.2 : Composition et caractéristiques d'un catalyseur commercial sulfurisé de type CoMo/alumine.

## IV.2.2 Résultats expérimentaux et commentaires

Au total 25 expériences ont été réalisées, pour étudier la variation de la teneur en soufre du gasoil hydrotraité en fonction des conditions opératoires du procédé, de la teneur en soufre de la charge et du nombre d'étage d'hydrotraitement effectué. Le tableau IV.3 regroupe les caractéristiques de chaque expérience ainsi que les résultats obtenus.

D'après la représentation graphique des expériences N° 1 à 8 (figure IV.2), la teneur en soufre des produits diminue avec l'augmentation de la température.

Les expériences N° 9 à 12, représentées sur la figure IV.3, indiquent que l'augmentation de la pression diminue la teneur en soufre du produit, en particulier lorsque celle-ci varie entre 800 et 1200 psi ; entre 1200 et 2000 psi la diminution de la teneur en soufre est moins importante.

La figure IV.4 qui représente les expériences N° 13 à 18, N° 19 à 20, N° 21 à 22 et N° 23 à 25 indique que l'augmentation de la vitesse spatiale horaire augmente la teneur en soufre des gasoils hydrotraités.

Ces résultats confirment ce qui a été dit dans la section variables opératoires du chapitre I.

<i>N°exp</i>	<i>T (°C)</i>	<i>P (psi)</i>	<i>VVH (h<sup>-1</sup>)</i>	<i>Teneur en soufre de la charge (ppm)</i>	<i>N° d'étage</i>	<i>Teneur en soufre du produit (ppm)</i>
1	348	800	2.6	10000	1	1260
2	357	800	2.6	10000	1	990
3	362	800	2.6	10000	1	850
4	368	800	2.6	10000	1	700
5	374	800	2.6	10000	1	550
6	386	800	2.6	10000	1	300
7	397	800	2.6	10000	1	110
8	403	800	2.6	10000	1	80
9	360	800	2.8	10000	1	910
10	360	1200	2.8	10000	1	780
11	360	1600	2.8	10000	1	720
12	360	2000	2.8	10000	1	700
13	360	800	0.4	12000	1	350
14	360	800	0.47	12000	1	390
15	360	800	0.67	12000	1	470
16	360	800	1.14	12000	1	640
17	360	800	2.37	12000	1	890
18	360	800	2.6	12000	1	920
19	360	800	6.0	7500	1	1870
20	360	800	3.0	7500	1	1250
21	360	800	6.0	1870	2	1140
22	360	800	1.5	1870	2	350
23	360	800	2.1	990	2	170
24	360	800	1.0	990	2	50
25	360	800	0.4	990	2	10

Tableau IV.3 : Caractéristiques et résultats des expériences d'hydrotraitement de différents gasoils de distillation atmosphérique.

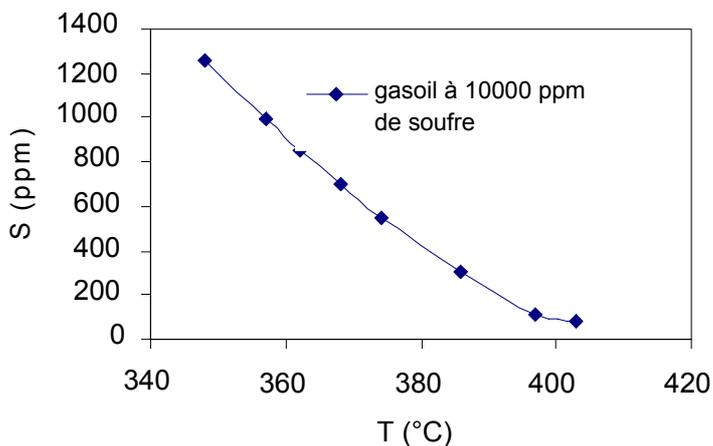


Figure IV.2 : Influence de la température d'hydrotraitement sur la teneur en soufre du produit (P=800 psi, VVH=2,6 h ).

-1

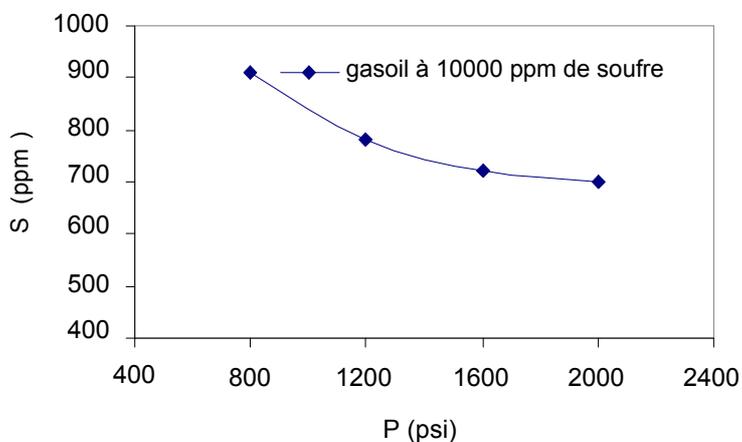


Figure IV.3 : Influence de la pression d'hydrotraitement sur la teneur en soufre du produit (T=360 °C, VVH=2,8 h ).

-1

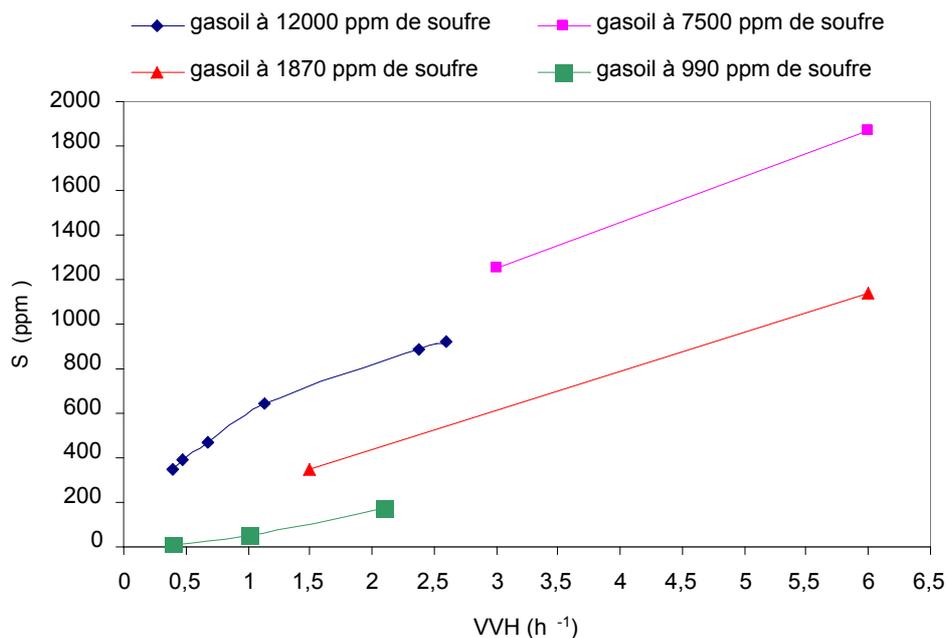


Figure IV.4 : Influence de la vitesse spatiale horaire d'hydrotraitement sur la teneur en soufre du produit (T=360 °C, P=800 psi).

Le gasoil hydrotraité de l'expérience N° 2 est utilisé comme charge pour les essais N° 23,24 et 25 ; tandis que le gasoil obtenu de l'expérience N°19 est utilisé comme charge pour les essais N°21 et 22. Un deuxième étage d'hydrotraitement permet l'obtention de résultats nettement plus satisfaisants, par exemple l'expérience N° 2 en un étage fournit un gasoil avec une teneur en soufre de 990 ppm, en subissant un deuxième étage (expérience N° 25) la teneur en soufre diminue jusqu'à 10 ppm.

Les variables de ce procédé c'est-à-dire la température, la pression, la vitesse spatiale horaire, la teneur en soufre de la charge ainsi que le nombre d'étage constituent les entrées du modèle ou du réseau ; tandis que la teneur en soufre du gasoil hydrotraité constitue sa sortie. Le perceptron multicouche comportera donc trois couches :

- La première est constituée de cinq variables d'entrées.
- La seconde contiendra un certain nombre de neurones cachés.
- La troisième est constituée d'un seul neurone de sortie.

L'objectif que l'on se fixe est la détermination de la structure optimale du réseau, autrement dit du nombre de neurones cachés optimal, qui permet la meilleure approximation de la teneur en soufre des gasoils à la sortie du procédé, tout en garantissant une bonne généralisation. Mais avant cela, il est nécessaire comme expliqué dans le chapitre III de normaliser ces données.

### **IV.2.3 Normalisation des données**

Les données du procédé d'hydrotraitement sont tous positives, il est donc plus logique de les centrer entre 0 et 1. En pratique, il est recommandé de normaliser les données d'entrée et de sortie du modèle dans l'intervalle [0.1, 0.9], cette technique permet une meilleure convergence du réseau [83,84]. La fonction utilisée pour la normalisation des différentes variables est donnée par la relation suivante [72] :

$$y = (y_{\max} - y_{\min}) * (x - x_{\min}) / (x_{\max} - x_{\min}) + y_{\min} \quad \dots\dots\dots(IV.1)$$

Où :

$x$  : est un élément du vecteur (d'entrée ou de sortie) à normaliser.

$x_{\max}$  : est la valeur de l'élément le plus grand du vecteur à normaliser.

$x_{\min}$  : est la valeur de l'élément le plus petit du vecteur à normaliser.

$y$  : est la valeur normalisée de  $x$ .

$y_{\max}$  : est la valeur maximale de l'intervalle de normalisation c'est-à-dire 0.9.

$y_{\min}$  : est la valeur minimale de l'intervalle de normalisation c'est-à-dire 0.1.

La commande qui réalise cette opération, sous MATLAB, est désignée par « mapminmax ». Les données normalisées sont regroupées dans le tableau IV.4.

N°exp	Entrées du réseau					Sortie du réseau
	T (°C)	P (psi)	VVH (h <sup>-1</sup> )	Teneur en soufre de la charge (ppm)	N° d'étage	Teneur en soufre du produit (ppm)
1	0.1000	0.1000	0.4143	0.7547	0.1000	0.6376
2	0.2309	0.1000	0.4143	0.7547	0.1000	0.5215
3	0.3036	0.1000	0.4143	0.7547	0.1000	0.4613
4	0.3909	0.1000	0.4143	0.7547	0.1000	0.3968
5	0.4782	0.1000	0.4143	0.7547	0.1000	0.3323
6	0.6527	0.1000	0.4143	0.7547	0.1000	0.2247
7	0.8127	0.1000	0.4143	0.7547	0.1000	0.1430
8	0.9000	0.1000	0.4143	0.7547	0.1000	0.1301
9	0.2745	0.1000	0.4429	0.7547	0.1000	0.4871
10	0.2745	0.3667	0.4429	0.7547	0.1000	0.4312
11	0.2745	0.6333	0.4429	0.7547	0.1000	0.4054
12	0.2745	0.9000	0.4429	0.7547	0.1000	0.3968
13	0.2745	0.1000	0.1000	0.9000	0.1000	0.2462
14	0.2745	0.1000	0.1100	0.9000	0.1000	0.2634
15	0.2745	0.1000	0.1386	0.9000	0.1000	0.2978
16	0.2745	0.1000	0.2057	0.9000	0.1000	0.3710
17	0.2745	0.1000	0.3814	0.9000	0.1000	0.4785
18	0.2745	0.1000	0.4143	0.9000	0.1000	0.4914
19	0.2745	0.1000	0.9000	0.5730	0.1000	0.9000
20	0.2745	0.1000	0.4714	0.5730	0.1000	0.6333
21	0.2745	0.1000	0.9000	0.1639	0.9000	0.5860
22	0.2745	0.1000	0.2571	0.1639	0.9000	0.2462
23	0.2745	0.1000	0.3429	0.1000	0.9000	0.1688
24	0.2745	0.1000	0.1857	0.1000	0.9000	0.1172
25	0.2745	0.1000	0.1000	0.1000	0.9000	0.1000

Tableau IV.4 : Données normalisées du procédé (apprentissage, validation et test).

### IV.3 Détermination de la structure optimale du réseau

#### IV.3.1 Démarche à suivre

La détermination de la structure optimale du modèle, comme expliquée dans le chapitre III, se fait d'une manière itérative. Pour cela, les données sont partagées aléatoirement en trois ensembles :

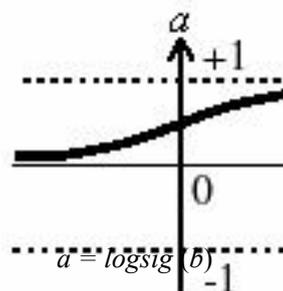
- Un ensemble d'apprentissage constitué de 17 exemples, les données de cet ensemble sont représentées sur les tableaux en noir.
- Un ensemble de validation comprenant 3 exemples, les données de cet ensemble sont représentées sur les tableaux en vert.
- Un ensemble de test contenant 5 exemples, les données de cet ensemble sont représentées sur les tableaux en rouge.

Pour garantir la propriété d'approximation universelle du réseau, on a choisi comme fonction d'activation des neurones cachés une fonction sigmoïde. La fonction la plus utilisée en pratique est la fonction logistique [72] qui vaut :

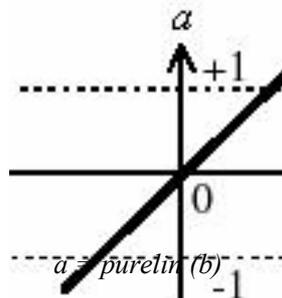
$$f(v) = \frac{1}{1 + \exp(-v)} \dots\dots\dots(IV.2)$$

Où  $v = w_0 + \sum_{i=1}^{n-1} w_i x_i \dots\dots\dots(IV.3)$

Cette fonction permet de limiter l'amplitude des sorties entre 0 et 1, elle est désignée sous MATLAB par *logsig*.



Suivant le théorème d'approximation universelle, la fonction d'activation du neurone de sortie doit être linéaire :  $f(v) = v$ . Elle est désignée sous MATLAB par *purelin*.



L'apprentissage du perceptron multicouche considéré, s'effectue à l'aide de l'algorithme de rétropropagation du gradient avec momentum, la modification des poids du réseau se fait par la méthode du second ordre de Levenberg-Marquardt. La technique de l'arrêt prématuré est utilisée pour éviter le surajustement du modèle.

En pratique, il est recommandé d'initialiser le pas du gradient  $\mu$  à 0,001, ceci permet une meilleure convergence du réseau. Le facteur d'échelle  $r$ , utilisé pour l'ajustement du pas  $\mu$ , est pris égale à 10. La constante du momentum  $\lambda$  est généralement prise égale à 0,9 [72].

Lorsque le nombre d'itérations (ou *epochs* en anglais) dépasse les 150, ou bien l'erreur quadratique moyenne devient égale à 0, ou le gradient atteint une valeur minimale ( $1e-01$ ), ou le pas  $\mu$  atteint une valeur maximale ( $1e10$ ) ou encore l'erreur sur l'ensemble de validation croît durant 6 itérations successives : l'algorithme s'arrête et n'évolue plus [72].

Notre stratégie consiste à faire varier le nombre de neurones de la couche cachée. Plusieurs réseaux de différentes structures vont être testés. Nous choisissons par la suite le réseau optimal qui donne une erreur de test la plus petite possible.

L'initialisation des poids et biais s'effectue aléatoirement, entre -1 et +1. Il est donc nécessaire, pour un nombre déterminé de neurones cachés, de faire plusieurs tentatives d'initialisation de paramètres afin d'avoir plus de chance de trouver le modèle optimal.

### IV.3.2 Résultats et commentaires

Nous avons fait varier le nombre de neurones de la couche cachée entre 1 et 10. Pour chaque nombre de neurones cachés, plusieurs essais d'initialisation de paramètres ont été effectués (plus d'une trentaine). Pour certains de ces essais, les erreurs quadratiques moyennes des trois ensembles apprentissage (EQMA), validation (EQMV) et test (EQMT) sont représentées sur les tableaux suivants :

<i>N° de l'essai</i>	1	2	3	4	5
<b><i>EQMA</i></b>	9.3788e-003	4.6268e-004	1.2930e-003	5.9598e-002	4.7983e-003
<b><i>EQMV</i></b>	8.1430e-004	2.5127e-003	1.0494e-003	1.2419e-002	6.5600e-004
<b><i>EQMT</i></b>	1.2263e-003	5.3157e-003	3.9113e-003	1.4136e-002	8.4966e-004

Tableau IV.5 : Valeurs des EQMA, EQMV et EQMT du réseau à un neurone caché pour différents essais d'initialisation de paramètres.

<i>N° de l'essai</i>	1	2	3	4	5
<b><i>EQMA</i></b>	6.8671e-005	7.6967e-004	1.8890e-004	2.2794e-002	6.7009e-003
<b><i>EQMV</i></b>	2.8272e-004	1.9525e-003	1.2022e-004	1.3669e-003	1.5839e-003
<b><i>EQMT</i></b>	2.8078e-003	3.7695e-003	1.6995e-003	2.8674e-003	3.1484e-003

Tableau IV.6 : Valeurs des EQMA, EQMV et EQMT du réseau à deux neurones cachés pour différents essais d'initialisation de paramètres.

<i>N° de l'essai</i>	1	2	3	4	5
<b><i>EQMA</i></b>	1.6284e-004	5.1442e-003	3.9618e-005	6.0318e-002	2.9334e-005
<b><i>EQMV</i></b>	6.7497e-005	4.5272e-004	1.5708e-003	1.3671e-002	1.0807e-003
<b><i>EQMT</i></b>	2.8356e-003	3.4727e-003	5.3081e-003	1.7739e-003	3.5475e-003

Tableau IV.7 : Valeurs des EQMA, EQMV et EQMT du réseau à trois neurones cachés pour différents essais d'initialisation de paramètres.

<i>N° de l'essai</i>	1	2	3	4	5
<b><i>EQMA</i></b>	5.3073e-003	6.4500e-006	4.4921e-004	7.6166e-007	3.0114e-005
<b><i>EQMV</i></b>	2.0824e-003	7.7387e-003	4.5863e-004	2.9521e-003	1.5228e-003
<b><i>EQMT</i></b>	4.7878e-003	3.8450e-003	3.0157e-003	5.1933e-003	4.4919e-003

Tableau IV.8 : Valeurs des EQMA, EQMV et EQMT du réseau à quatre neurones cachés pour différents essais d'initialisation de paramètres.

<i>N° de l'essai</i>	1	2	3	4	5
<b><i>EQMA</i></b>	2.5779e-004	7.3485e-002	1.4664e-005	2.3826e-002	3.2057e-006
<b><i>EQMV</i></b>	2.0461e-003	7.3554e-003	2.0587e-003	1.0102e-003	2.6726e-003
<b><i>EQMT</i></b>	4.3839e-003	7.6453e-003	6.3401e-003	1.3693e-002	4.8961e-003

Tableau IV.9 : Valeurs des EQMA, EQMV et EQMT du réseau à cinq neurones cachés pour différents essais d'initialisation de paramètres.

<i>N° de l'essai</i>	1	2	3	4	5
<b><i>EQMA</i></b>	2.2121e-006	2.5309e-011	3.3909e-002	5.1351e-005	6.8727e-004
<b><i>EQMV</i></b>	9.5913e-004	5.3055e-004	1.7251e-003	2.7808e-003	1.1263e-003
<b><i>EQMT</i></b>	3.3704e-003	4.7559e-003	2.0861e-002	7.1309e-003	3.3742e-003

Tableau IV.10 : Valeurs des EQMA, EQMV et EQMT du réseau à six neurones cachés pour différents essais d'initialisation de paramètres.

<i>N° de l'essai</i>	1	2	3	4	5
<i>EQMA</i>	5.8702e-005	9.5640e-004	2.7498e-003	4.0027e-006	<b>1.6952e-005</b>
<i>EQMV</i>	1.6307e-003	2.5868e-003	2.3895e-004	2.2681e-004	<b>8.7257e-005</b>
<i>EQMT</i>	2.4134e-003	4.8288e-003	2.4668e-003	2.4668e-003	<b>2.9462e-004</b>

Tableau IV.11 : Valeurs des EQMA, EQMV et EQMT du réseau à sept neurones cachés pour différents essais d'initialisation de paramètres.

<i>N° de l'essai</i>	1	2	3	4	5
<i>EQMA</i>	1.8364e-004	1.1688e-002	1.7042e-005	2.5250e-002	3.0461e-006
<i>EQMV</i>	3.2452e-003	3.6110e-003	4.0852e-003	4.3283e-003	1.9109e-003
<i>EQMT</i>	7.4355e-003	4.1479e-003	4.2741e-003	1.3096e-002	4.6458e-003

Tableau IV.12 : Valeurs des EQMA, EQMV et EQMT du réseau à huit neurones cachés pour différents essais d'initialisation de paramètres.

<i>N° de l'essai</i>	1	2	3	4	5
<i>EQMA</i>	2.0861e-003	2.4534e-005	8.7344e-007	2.0885e-006	2.3622e-003
<i>EQMV</i>	1.3904e-002	2.2605e-004	1.5971e-003	1.8127e-003	2.2744e-003
<i>EQMT</i>	1.0426e-002	3.1622e-003	2.6872e-003	3.3880e-003	4.0871e-003

Tableau IV.13 : Valeurs des EQMA, EQMV et EQMT du réseau à neuf neurones cachés pour différents essais d'initialisation de paramètres.

<i>N° de l'essai</i>	1	2	3	4	5
<i>EQMA</i>	3.1335e-003	5.4414e-005	2.9568e-008	3.3384e-004	7.0841e-007
<i>EQMV</i>	1.0676e-003	5.0998e-004	1.0803e-004	4.2865e-004	1.7542e-004
<i>EQMT</i>	3.0400e-003	4.0030e-003	2.6541e-003	2.0968e-003	1.0196e-003

Tableau IV.14 : Valeurs des EQMA, EQMV et EQMT du réseau à dix neurones cachés pour différents essais d'initialisation de paramètres.

Les évolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage de tous ces essais sont représentées par les courbes de l'annexe. La figure IV.5 correspond à l'essai N°1 du réseau à un neurone caché. On remarque que l'apprentissage s'arrête lorsque l'EQMV augmente au cours de 6 itérations successives, l'augmentation de l'EQMV est représentée par une ligne droite. Les poids et biais retenus par le réseau sont ceux qui correspondent au minimum de l'erreur de validation. Ce minimum est atteint au bout de 15 itérations en même temps que celui de l'EQMT. Tout ceci confirme l'efficacité de la technique de l'arrêt prématuré pour garantir une bonne EQMT et donc une meilleure généralisation du modèle.

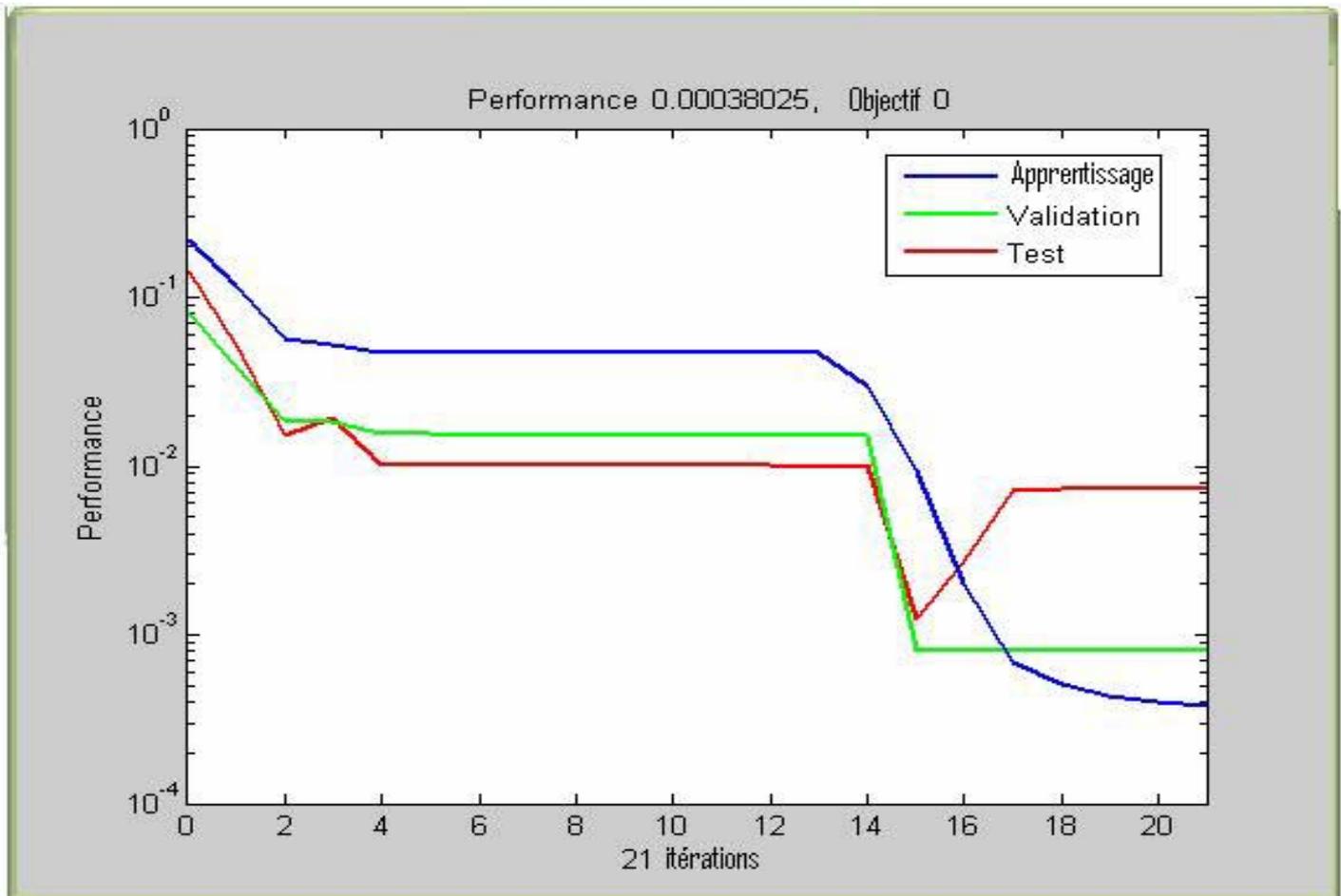


Figure IV.5 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à un neurone caché, essai N°1.

Après analyse des résultats obtenus, on remarque l'effet important de l'initialisation des paramètres. En effet, pour un certain nombre de neurones dans la couche cachée, on peut avoir des résultats complètement différents en changeant tout simplement l'initialisation des poids et biais. Ceci est dû au fait que la fonction de coût des moindres carrés, qu'on cherche à minimiser, n'est pas une fonction uniforme, elle peut présenter plusieurs minima locaux avec des pentes et des piques ainsi que des plateaux. D'autre part, l'utilisation de la technique de l'arrêt prématuré de l'apprentissage, comme son nom l'indique arrête le processus plus tôt et ne le laisse pas évoluer pour sortir de ses obstacles. Il est donc nécessaire pour trouver le modèle optimal, de tâtonner, comme nous l'avons fait, sur l'initialisation des paramètres.

Le critère de choix entre ces différents modèles est bien évidemment l'erreur quadratique moyenne sur l'ensemble de test. On remarque que le réseau à sept neurones cachés avec des paramètres initiaux correspondant à l'essai N°5, présente l'EQMT la plus petite ( $2.9462e-004$ ) avec de très bonnes erreurs quadratiques moyennes sur les ensembles d'apprentissage ( $1.6952e-005$ ) et de validation ( $8.7257e-005$ ). Ce modèle est donc le modèle optimal qui permet la meilleure approximation de la teneur en soufre des gasoils hydrotraités. Les évolutions des EQMA, EQMV et EQMT au cours du processus d'apprentissage du réseau optimal sont représentées sur la figure IV.6.

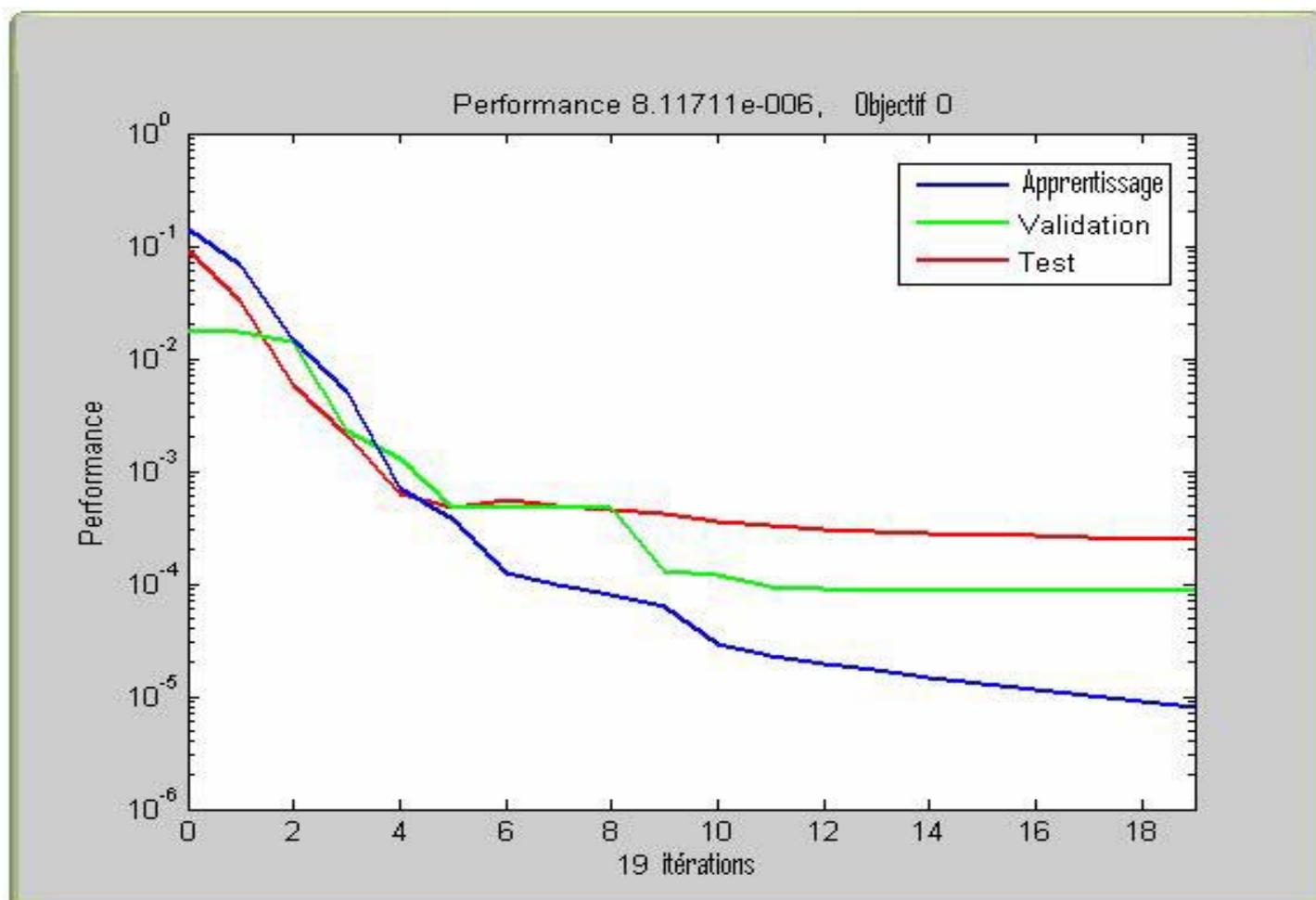


Figure IV.6 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à sept neurones cachés, essai N°5.

## IV.4 Evaluation des performances du modèle

### IV.4.1 Dénormalisation des valeurs de la sortie du réseau

Pour pouvoir faire la comparaison entre les valeurs mesurées (S) et les valeurs prédites ( $S^*$ ) de la teneur en soufre des gasoils hydrotraités, il est nécessaire de dénormaliser les valeurs de la sortie du réseau. La commande qui réalise cette opération, sous MATLAB, est désignée par « mapminmax, 'reverse' ».

Les points sur la figure IV.7, représentent la teneur en soufre prédite (après dénormalisation) en fonction de la teneur en soufre mesurée du produit de chaque expérience. On remarque que ces points sont distribués autour de l'axe  $y=x$ , donc les valeurs de  $S^*$  sont très proches de celles de S,

elles sont presque égales. On peut d'ores et déjà dire que le modèle fournit de bons résultats, mais un calcul d'erreur sur chaque point nous permettrait de tirer une meilleure conclusion.

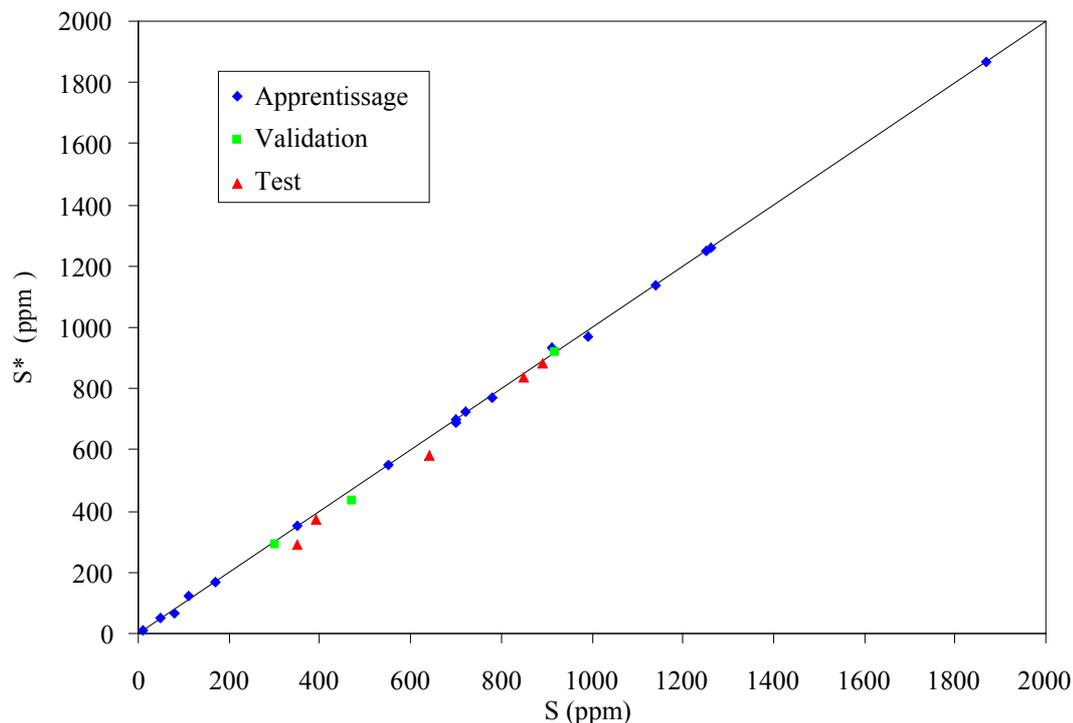


Figure IV.7 : Les teneurs en soufre prédites en fonction des teneurs en soufre mesurées des gasoils hydrotraités.

#### IV.4.2 Calcul d'erreurs

Afin d'évaluer les performances de notre modèle, on suppose que les valeurs mesurées de la teneur en soufre des produits sont les valeurs réelles (en vérité il y a des erreurs sur ces mesures).

L'erreur absolue  $\Delta S$  est, comme son nom l'indique, la valeur absolue de la différence entre la valeur réelle et la valeur approchée  $S - S^*$  [85].

L'erreur relative est le rapport entre l'erreur absolue et la valeur réelle  $\frac{\Delta S}{S}$  [85]. Les résultats de calcul de l'erreur relative, pour chaque exemple, sont présentés sur le tableau IV.15.

L'erreur relative moyenne sur l'ensemble d'apprentissage est de **2.31%**, celle sur l'ensemble de validation est de **3.62%** et pour l'ensemble de test elle est de **6.74%**. Ces résultats sont très satisfaisants, le réseau a bien appris et il généralise aussi bien.

$N^{\circ}exp$	$S$ (ppm)	$S^*$ (ppm)	$\Delta S / S .100$
1	1260	1260	0.00
2	990	970	2.02
3	850	835	1.76
4	700	689	1.57
5	550	549	0.18
6	300	291	3.00
7	110	123	11.82
8	80	66	17.50
9	910	934	2.64
10	780	772	1.03
11	720	726	0.83
12	700	700	0.00
13	350	353	0.86
14	390	373	4.36
15	470	433	7.87
16	640	580	9.38
17	890	883	0.79
18	920	920	0.00
19	1870	1868	0.11
20	1250	1249	0.08
21	1140	1140	0.00
22	350	289	17.43
23	170	169	0.59
24	50	50	0.00
25	10	10	0.00

Tableau IV.15 : Résultats de calcul de l'erreur relative pour chaque exemple.

## **IV.5 Conclusion**

Pour modéliser le procédé d'hydrotraitement du gasoil à l'aide des réseaux de neurones artificiels, on a opté pour un réseau de type perceptron multicouche. La détermination du nombre de neurones de la couche cachée ainsi que le choix des paramètres initiaux du réseau s'effectuent par tâtonnement. L'utilisation de l'interface graphique du logiciel MATLAB nous a énormément facilité ces tâches. Les résultats obtenus sont très bons, l'erreur de test est inférieure à 7%. La technique des réseaux de neurones artificiels est donc très bien adaptée pour la modélisation de ce type de processus.

---

## ***CONCLUSION ET PERSPECTIVES***

---

## **Conclusion et perspectives**

Les spécifications requises pour le gasoil et les normes de pollution appliquées sur les moteurs diesel ont considérablement évoluées ces dernières années. Le procédé d'hydrotraitement en deux étages permet l'obtention d'un gasoil de bonne qualité avec une teneur en soufre faible (jusqu'à 10 ppm) répondant ainsi aux exigences du marché.

Dans ce présent travail, nous avons réussi à modéliser le procédé d'hydrotraitement du gasoil en deux étages à l'échelle pilote. Le modèle élaboré relie la teneur en soufre du gasoil hydrotraité avec les différentes variables opératoires (température, pression, vitesse spatiale) ainsi que la teneur en soufre de la charge et le nombre d'étages d'hydrotraitement effectué. Ce modèle permet de prévoir le comportement du procédé sous diverses conditions et avec différents gasoils et ceci sans aucune expérience. Le modèle permet donc de gagner du temps et surtout de l'énergie.

La réalisation d'un tel modèle est passée, en vérité, par une succession de choix et de décisions :

- Le premier consiste à choisir le type de modèle le mieux adapté pour décrire le procédé. Les modèles qui existent déjà dans la littérature sont tous basés sur les lois fondamentales du génie des procédés. Ces derniers, comme expliqué précédemment, présente un inconvénient majeur, en effet, ils sont spécifiques à un seul type de gasoil, ceci est dû aux hypothèses énoncées pour simplifier les équations mathématiques du modèle en question. Notre choix s'est donc posé sur la modélisation de type boîte noire qui malgré son coût élevé en terme de nombre d'expériences, permet la construction d'un modèle plus simple et surtout plus général.
- Le deuxième choix va porter sur l'approche adoptée pour la réalisation de ce modèle. On peut soit utiliser une fonction mathématique telle que les polynômes, soit une technique d'intelligence artificielle telle que les réseaux de neurones, les systèmes experts, la logique floue ou la physique qualitative. Les trois dernières techniques sont complètement écartées car contrairement aux réseaux de neurones qui fournissent une approche quantitative des interactions entrées/sorties du procédé, elles donnent plutôt des réponses de type décisionnel. Il reste donc à choisir entre les fonctions mathématiques et les réseaux de neurones, c'est là que la propriété d'approximation universelle parcimonieuse des réseaux de neurones intervient. Les réseaux de neurones artificiels possèdent la particularité de pouvoir approcher n'importe quelle fonction suffisamment régulière avec moins de paramètres ou avec plus de précision que les approximateurs usuels. Nous avons par conséquent opté pour une modélisation de type boîte noire à l'aide des réseaux de neurones artificiels.
- Il existe différents types de réseaux neuronaux qui trouvent leurs applications dans divers domaines pour la réalisation de différents objectifs. Après une étude théorique on a constaté que les réseaux de neurones non bouclés (statique) à apprentissage supervisé de type Multi Layer Perceptron sont les mieux adaptés pour la modélisation d'un processus tel que l'hydrotraitement du gasoil.
- Il reste maintenant à choisir la manière d'effectuer l'apprentissage de ce MLP. La méthode la plus rapide et la plus efficace consiste, dans ce cas, à évaluer le gradient de la fonction de coût par rétropropagation et à modifier les paramètres du réseau par la méthode du second ordre de Levenberg-Marquardt. Le terme du momentum est ajouté pour améliorer cette procédure tandis que la technique de l'arrêt prématuré est utilisée pour garantir une bonne généralisation du modèle.
- En fin, nous devons choisir la structure optimale du MLP, c'est-à-dire de déterminer le nombre de neurones cachés et les paramètres initiaux du réseau qui permettent la meilleure approximation de la teneur en soufre des gasoils hydrotraités tout en garantissant une bonne généralisation. Après

tâtonnement, on a constaté que le réseau à sept neurones cachés présente la plus faible EQMT, c'est donc le modèle recherché.

Le modèle du procédé d'hydrotraitement du gasoil en deux étages, ainsi élaboré, a fourni de très bons résultats, ceci prouve l'aptitude des réseaux de neurones à résoudre ce genre de problème. En perspective il serait donc intéressant :

- D'enrichir ce type de modèle, en prenant en considération plus de variables d'entrées, par exemple les caractéristiques des charges (teneurs en aromatiques, teneurs en oléfines, les indices de cétones...etc) ou bien les caractéristiques du procédé (le type de catalyseur, le taux d'hydrogène...etc). Le modèle deviendrait plus complet et plus général, il permettrait la prédiction de toutes les caractéristiques des produits avec une bonne précision.
- L'utilisation d'un tel modèle à l'échelle industrielle pourrait rendre d'énormes services, il permettrait d'attirer l'attention des opérateurs lorsqu'il y a des anomalies dans le fonctionnement du procédé, par exemple lors du vieillissement du catalyseur. Il est important de noter que la mise en œuvre de ce type de modèle dans une raffinerie est beaucoup plus simple qu'à l'échelle pilote car les données sont disponibles et en grandes quantités, en effet le suivi du procédé s'effectue d'une manière rigoureuse et ceci plusieurs fois par jour.
- L'application de ce genre d'approche pour d'autres procédés de raffinage ou de pétrochimie serait très bénéfique.
- En fin, ce type de modèle pourrait servir à l'optimisation des procédés. L'utilisation des algorithmes génétiques pour l'optimisation du procédé de craquage catalytique fluidisé à partir d'un modèle neuronal a fourni de très bon résultats.

---

***REFERENCES BIBLIOGRAPHIQUES***

---

## *Références bibliographiques*

- [1] P. Leprince, "Le Raffinage du Pétrole: Procédés de transformation", Tome 3, Editions Technip, (1998), France.
- [2] "Hydrotraitement", Site du groupe Total  
[http://www.total.com/portail/webzine/gloss/main.php?lang=fr&id\\_gloss=15](http://www.total.com/portail/webzine/gloss/main.php?lang=fr&id_gloss=15)
- [3] "Hydrotreating Process", Site de la compagnie UOP  
<http://www.uop.com/refining/1060.html>
- [4] Paul Grange, Xavier Vanhaeren, "Hydrotreating catalysts, an old story with new challenges", Catalysis Today, Vol. 36, (1997), PP. 375-391.
- [5] P. Courty, J.F. Gruson, "Refining clean fuels for the future", Oil & Gas Science and Technology (Rev.IFP), Vol. 56, (2001), PP. 551-524.
- [6] Bernard Bensaïd et Valérie Saint-Antonin, "Le Diesel aux Etats-Unis", Panorama 2004, Site de l'Institut Français du Pétrole.  
[http://ogst.ifp.fr/IFP/fr/fichiers/cinfo/IFP-Panorama04\\_06-DieselVF.pdf](http://ogst.ifp.fr/IFP/fr/fichiers/cinfo/IFP-Panorama04_06-DieselVF.pdf)
- [7] E.F. Kaminskii, V.A. Khavkin, V.M. Kurganov, L.N. Osipov, D.M. Bychkova, I.N. Loshchenk, "Production of ecological clean diesel fuels", Chemistry and Technology of Fuels and Oils, Vol. 32, (1996), PP. 68-70.
- [8] J.C. Guibet, "Carburants liquides", Techniques de l'ingénieur, BE 8545, (1998).
- [9] A. Douaud, C. Girard, "Les nouvelles technologies de moteurs et de carburants pour les transports routiers", Site de l'organisation WEC (The World Energy Council).  
[http://www.worldenergy.org/wec-geis/publications/default/tech\\_papers/17th\\_congress/3\\_4\\_13.asp](http://www.worldenergy.org/wec-geis/publications/default/tech_papers/17th_congress/3_4_13.asp)
- [10] "Le Dioxyde de Soufre", Site de l'université de Manchester (Manchester Metropolitan university).  
[http://www.ace.mmu.ac.uk/eae/french/Acid\\_Rain/Older/Sulphur\\_Dioxide.html](http://www.ace.mmu.ac.uk/eae/french/Acid_Rain/Older/Sulphur_Dioxide.html)
- [11] J.P. Wauquier, "Le Raffinage du Pétrole: Pétrole brut, Produits pétroliers, Schémas de fabrication", Tome 1, Editions Technip, (1994), France.
- [12] J.C. Guibet, "Carburants et Moteurs", Tome 1, Editions Technip, (1997), France.
- [13] "Hydrocarbures aromatiques polycycliques", Article de l'encyclopédie libre Wikipédia, (2007).  
<http://fr.wikipedia.org/wiki/HAP>
- [14] "Les HAP", Site de l'institut national de la recherche agronomique (INRA) de Montpellier.  
<http://www.montpellier.inra.fr/narbonne/francais/objets-thematiques/ot-cmm/ot-cmm-11-hap.html>
- [15] J.C. Guibet, "Carburants et Moteurs", Tome 2, Editions Technip, (1997), France.

- [16] "Norme européenne d'émission Euro", Article de l'encyclopédie libre Wikipédia, (2006).  
[http://fr.wikipedia.org/wiki/Norme\\_europ%C3%A9enne\\_d'%C3%A9mission\\_Euro](http://fr.wikipedia.org/wiki/Norme_europ%C3%A9enne_d'%C3%A9mission_Euro)
- [17] "Normes d'émissions pour les véhicules", Site de EurActiv.com (le principal portail Internet entièrement consacré aux affaires européennes).  
<http://www.euractiv.com/fr/transport/normes-emissions-euro-5-vehicules/article-137719>
- [18] "Normes antipollution des moteurs", Site Autotech : la technique automobile.  
<http://autotech.free.fr/autotech/norm/index.php>
- [19] Jean Cosyns, German Martino, "Hydrogénation des hydrocarbures", Techniques de l'ingénieur, J5500, (1997).
- [20] Kim G. Knudsen, Barry H. Cooper, Henrik Topsøe, "Catalyst and process technologies for ultra low sulfur diesel", Applied Catalysis A: General, Vol. 189, (1999), PP. 205-215.
- [21] Sonja Eijsbouts, "Life cycle of hydroprocessing catalysts and total catalyst management", Studies in surface science and catalysis, Vol. 127, (1999), PP. 21-36.
- [22] P. Blanchard, "Etude de la sulfuration industrielle des catalyseurs d'HDS", site de l'université de Lille 1.  
[http://www.univ-lille1.fr/edsmre/Master\\_2005/ENV/LCL/Sujet\\_ENV9.pdf](http://www.univ-lille1.fr/edsmre/Master_2005/ENV/LCL/Sujet_ENV9.pdf)
- [23] Shinichi Inoue, Akihiro Muto, Hidehiko Kudou, Takeo Ono, "Preparation of novel titania support by applying the multi-gelation method for ultra-deep HDS of diesel oil", Applied Catalysis A: General, Vol. 269, (2004), PP. 7-12.
- [24] Takashi Fujikawa, Hiroshi Kimura, Kazuyuki Kiriyama, Kazuhiko Hagiwara, "Development of ultra-deep HDS catalyst for production of clean diesel fuels", Catalysis Today, Vol. 111, (2006), PP. 188-193.
- [25] H.M. Pinzon, L.I. Merino, A. Centeno, S.A. Giraldo, "Performance of noble metal-Mo/ $\gamma$ -Al<sub>2</sub>O<sub>3</sub> catalysts : Effect of preparation parameters", Studies in surface science and catalysis, Vol. 127, (1999), PP. 97-104.
- [26] Stéphanie Pessayre, Christophe Geantet, Robert Bacaud, Michel Vrinat, Thanh Son N'Guyen, Yvonne Soldo, Jean Louis Hazemann, Michèle Breysse, "Platinum Doped Hydrotreating Catalysts for Deep Hydrodesulfurization of Diesel Fuels", Industrial and Engineering Chemistry Research, Vol. 46, (2007), PP. 3877-3883.
- [27] Gophal H. Singhal et al, "Hydrotreating process using noble metal supported catalysts", Site de google.com, Brevet d'invention No. 5152885, (1992), USA.  
<http://www.google.com/patents?hl=fr&lr=&vid=USPAT5152885&id=VXckAAAAEBAJ&oi=fnd&dq=++Hydrotreating+Catalysts++++Diesel+++metal>
- [28] Claude Naccache, "Catalyse hétérogène dans les procédés industriels", Techniques de l'ingénieur, J1255, (2005).

- [29] S. Ramirez, C. Cabreara, C. Aguilar, H. Vaca, P. Vega, R. Agueda, A. Garcia, R. Santiago, P. Schacht, "Two stages light gasoil hydrotreating for low sulfur diesel production", *Catalysis Today*, Vol. 98, (2004), PP. 323-332.
- [30] Gustavo Marroquin-Sanchez, Jorge Ancheyta-Juarez, "Catalytic hydrotreating of middle distillates blends in a fixed-bed pilot reactor", *Applied Catalysis A: General*, Vol. 207, (2001), PP. 407-420.
- [31] Jorge Ancheyta-Juarez, Enrique Aguilar-Rodriguez, Daniel Salazar-Sotelo, Gustavo Marroquin-Sanchez, German Quiroz-Sosa, Miguel Leiva-Nuncio, "Effect of hydrogen sulphide on the hydrotreating of middle distillates over Co-Mo/Al<sub>2</sub>O<sub>3</sub> catalyst", *Applied Catalysis A: General*, Vol. 183, (1999), PP. 265-272.
- [32] N. Guernalec, C. Geantet, P. Raybaud, T. Cseri, M. Aouine, M. Virinat, "Dual effect of H<sub>2</sub>S on Volcano Curves in hydrotreating sulphide catalysis", *Oil & Gas Science and Technology (Rev.IFP)*, Vol. 61, (2006), PP. 515-525.
- [33] "Bruts, Raffinage, Produits, Schéma de fabrication", *Cours de l'Institut Français du Pétrole*, (2005).
- [34] "Hydrotreating", *Hydrocarbon Processing: Refining Processes*, (2006), P. 146.
- [35] "Hydrotreating", Site de la compagnie AXENS.  
[http://www.axens.net/html-gb/offer/offer\\_products\\_4.html.php](http://www.axens.net/html-gb/offer/offer_products_4.html.php)
- [36] "Hydrotreating diesel", *Hydrocarbon Processing: Refining Processes*, (2004), P. 143.
- [37] "Hydrodesulfurization, ultra-low-sulfur diesel", *Hydrocarbon Processing*, Vol. 81, No. 11, (2002), P. 121.
- [38] "Hydrodésaromatisation", *Hydrocarbon Processing: Refining Processes*, (2006), PP. 120-121.
- [39] J.P. Gourlia, "Modélisation en génie des procédés", *Techniques de l'ingénieur*, J1022, (1995).
- [40] M. Boumahrat, A. Gourdin, "Méthodes numériques appliquées", OPU, (1993), Algérie.
- [41] Kaddour Najim, Enso Ikonen, "Outils mathématiques pour le génie des procédés", Edition DUNOD, (1999), France.
- [42] William L. Luyben, "Process modeling, simulation and control for chemical engineers", McGraw-Hill International Edition, (1990), USA.
- [43] Guy-Marie Côme, "Cinétique chimique appliquée : Modélisation", *Techniques de l'ingénieur*, J 1134, (1991).
- [44] Jorge Ancheyta, Rogelio Sotelo, "Kinetic modeling of vacuum gasoil catalytic cracking", *Journal of the Mexican Chemical Society*, Vol. 46, (2002), PP. 38-42.

[45] Jasvinder Singh, M.M. Kumar, Alok K. Saxena, Surendra Kumar, "Reaction pathways and product yields in mild thermal cracking of vacuum residues : A multi-lump kinetic model", Chemical Engineering Journal, Vol. 108, (2005), PP. 239-248.

[46] Jorge Ancheyta, Sergio Sanchez, Miguel A. Rodriguez, "Kinetic modeling of hydrocracking of heavy oil fractions : A review", Catalysis Today, Vol. 109, (2005), PP 76-92.

[47] Jean Leybros, "Extraction liquide-liquide : Modélisation des colonnes", Techniques de l'ingénieur, J 2765, (2004).

[48] Mohammad Reza, Jafari Nasr, Mehdi Majidi Givi, "Modeling of crude oil fouling in preheat exchangers of refinery distillation units", Applied Thermal Engineering, Vol. 26, (2006), PP. 1572-1577.

[49] Z. Wang, B. Yang, C. Chen, J. Yuan, L. Wang, "modeling and optimization for the secondary reaction of FCC gasoline based on the fuzzy neural network and genetic algorithm", Chemical Engineering and Processing, Vol. 46, (2007), PP. 175-180.

[50] Farouk S. Mjalli, S. Al-Asheh, H.E. Alfadala, "Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance", Journal of Environmental Management, Vol. 83, (2007), PP. 329-338.

[51] Dr Khadir Mohamed Tarek, "Commande prédictive 5 MPC des systèmes industriels : Modélisation du procédé de pasteurisation", Université Badji Mokhtar Annaba.  
[http://www.lri-annaba.net/khadir\\_tarek/khadir/commande\\_MPC.htm](http://www.lri-annaba.net/khadir_tarek/khadir/commande_MPC.htm)

[52] Manish C. Tayalat, Yan Fut, "Optimal design of heat exchangers : A genetic algorithm framework", Industrial and Engineering Chemistry Research, Vol. 38, (1999), PP. 456-467.

[53] Pascal Lorenzini, Michel Pons, Jacques Villiermaux, "Free-radical polymerization engineering-III. Modelling homogeneous polymerization of éthylène : mathematical model and new method for obtaining molecular-weight distribution", Chemical Engineering Science, Vol. 47, (1992), PP. 3969-3980.

[54] "Réseau de neurones", Article de l'encyclopédie libre Wikipédia, (2007).  
[http://fr.wikipedia.org/wiki/R%C3%A9seau\\_de\\_neuroneshttp](http://fr.wikipedia.org/wiki/R%C3%A9seau_de_neuroneshttp)

[55] J.P. Corriou, "Les réseaux de neurones pour la modélisation et la conduite des procédés", Lavoisier Technique et Documentation, (1995), Paris.

[56] Frédérique Perez, "Les réseaux de neurones artificiels", (2004).  
<http://www.techno-science.net/?onglet=articles&article=12> http

[57] Claude Touzet, "Les réseaux de neurones artificiels, introduction au connexionnisme, cours, exercices et travaux pratiques", (1992), France.

[58] Yves Kamp, Martin Hasler, "Réseaux de neurones récursifs pour mémoires associatives", Presses Polytechniques et Universitaires Romandes, (1990), Suisse.

- [59] Marc Parizeau, "Réseaux de neurones", Cours de l'Université LAVAL, (2004), Québec.  
<http://wcours.gel.ulaval.ca/2007/a/21410/default/5notes/titre.pdf> http
- [60] Rolf Ingold, "Reconnaissance des formes : Réseaux de neurones", Cours de l'Université de Fribourg, (2006), Suisse.  
[http://diuf.unifr.ch/courses/05-06/patrec/slides/projections/10\\_NeuralNets.pdf](http://diuf.unifr.ch/courses/05-06/patrec/slides/projections/10_NeuralNets.pdf) http
- [61] "Introduction aux réseaux de neurones", Cours de l'Université de Usherbrooke, Canada.  
<http://www.chimique.usherbrooke.ca/cours/gch445/neurones-intro.html>
- [62] "Neurone formel", Article de l'encyclopédie libre Wikipédia, (2007).  
[http://fr.wikipedia.org/wiki/Neurone\\_formel](http://fr.wikipedia.org/wiki/Neurone_formel) http
- [63] G. Dreyfus, J.M. Martinez, M. Samuelides, M.B. Gordon, F. Badran, S. Thiria, L. Héroult, "Réseaux de neurones : Méthodologie et applications", Editions Eyrolles, (2002), Paris.
- [64] G. Dreyfus, T. Cour, G. Giraud, A. Kodsí, T. Luong, R. Lauranson, C. Marcovici, K. Sadeghi, "Reconnaissance de formes par réseaux de neurones", Ecole polytechniques, (2002), France.  
[http://www.seas.upenn.edu/~timothee/papers/vision\\_system.pdf](http://www.seas.upenn.edu/~timothee/papers/vision_system.pdf)
- [65] Isabelle Rivals, "Les réseaux de neurones formels pour le pilotage de robots mobiles", Revue de l'association amicale des ingénieurs SUPÉLEC, No. 178, septembre-octobre 1996.
- [66] Michel Weinfeld, "Réseaux de neurones", Techniques de l'ingénieur, H 1990, (1995).
- [67] C. Pellegrini, "Neurones artificiels : modèles et réseaux", Cours du Centre Universitaire d'Informatique de l'université de Genève, (2004-2005).  
<http://cui.unige.ch/DI/cours/1815/slides/17-reseaux-neurones.pdf>
- [68] Jean-Pierre Briot, "Applications industrielles des réseaux de neurones : IJCNN'93 à Nagoya", Rapport de la conférence : International Joint Conference on Neural Networks, à Nagoya pour l'Université de Tokyo, (1993).  
<http://www-poleia.lip6.fr/~briot/publications/ijcnn93-japon-ia93.pdf>
- [69] Kuo-Yi Huang, "Application of artificial neural network for detecting phalaenopsis seedling diseases using color and texture features", Computers and Electronics in Agriculture, Vol. 57, (2007), PP. 3-11.
- [70] Anda Ionescu, Ives Canadau, "Air pollutant emissions prediction by process modelling - Application in the iron and steel industry in the case of a re-heating furnace", Environmental Modelling & Software, Vol. 22, (2007), PP. 1362-1371.
- [71] Zou Haofei, Xia Guoping, Yang Fangting, Yang Han, "A neural network model based on the multi-stage optimization approach for short-term food price forecasting in China", Expert Systems with applications, Vol. 33, (2007), PP. 347-356.
- [72] Howard Demuth, Mark Beale, Martin Hagan, "Neural Network Toolbox 5 : User's Guide", Guide d'utilisation des reseaux de neurones sur le logiciel MATLAB, (2007).

- [73] Rafic Younes, "Les réseaux de neurones", Cours de l'Université Libanaise, (2006).  
<http://www.ryounes.net/cours/chapitre%203%20RN.pdf>
- [74] "Introduction aux réseaux de neurones", Cours de l'Université Blaise Pascal, (2005-2006).  
[http://wwwobs.univ-bpclermont.fr/atmos/enseignement/cours-Master-2A/cours\\_RN\\_2006.p](http://wwwobs.univ-bpclermont.fr/atmos/enseignement/cours-Master-2A/cours_RN_2006.p)
- [75] G.F. Froment, "Modeling in the development of hydrotreatment processes", Catalysis Today, Vol. 98, (2004), PP. 43-54.
- [76] C. Murali, R.K. Voolapalli, N. Ravichander, D.T. Gokak, N.V. Choudary, "Trickle bed reactor model to simulate the performance of commercial diesel hydrotreating unit", Fuel, (2006), doi:10.1016/j.fuel.2006.09.019.
- [77] V. Ramesh Kumar, K.S. Balaraman, V.S. Ramachandra Rao, M.S. Ananth, "Modelling of hydrotreatment process in a trickle-bed reactor", Petroleum Science and Technology, Vol. 15, (1997), PP. 283-295.
- [78] R.M. Cotta, M.R. Wolf-Maciel, R. Maciel Filho, "A cape of HDT industrial reactor for middle distillates", Computers & Chemical Engineering, Vol. 24, (2000), PP. 1731-1735.
- [79] S. Inoue, T. Takatsuka, Y. Wada, S. Hirohama, T. Ushida, "Distribution function model for deep desulfurization of diesel fuel", Fuel, Vol. 79, (2000), PP. 843-849.
- [80] Michael C Hu, Zbigniew Ring, Jenny Briker, Mure Te, "Rigorous hydrotreater simulation", PTQ, Spring (2002), PP. 85-91.
- [81] S. Mohanty, D. Kunzru, D.N. Saraf, "Modeling of hydrotreating reactors", Science & Technology, Vol. 44, (1991), PP. 459-465.
- [82] L. Vradman, M.V. Landau, M. Herskowitz, "Deep desulphurization of diesel fuels : kinetic modeling of model compounds in trickle-bed", Catalysis Today, Vol. 48, (1999), PP. 41-48.
- [83] Ozel T, Nadgir A, "Prediction of flank wear by using back propagation neural network modeling when cutting hardened H-13 steel with chamfered and honed CBN tools", International Journal of Machine Tools & Manufacture, Vol. 42, (2002), PP. 287-297.
- [84] Tahmi Raouf, "Conception optimisée d'un réseau de neurones pour la surveillance en ligne de l'usure des outils coupants", Mémoire de magister de l'université M'hamed Bougara Bumerdes, Faculté des Sciences de l'Ingénieur, Département Génie Mécanique, (2006), Algérie.
- [85] Olivier Burri, "Le point sur les calculs d'erreurs", Cours de l'Ecole Polytechnique Fédérale de Lausanne,(2004).  
<http://www2.epfl.ch/webdav/site/coaching/shared/import/migration/Calculserreur.pdf>

# *Annexe*

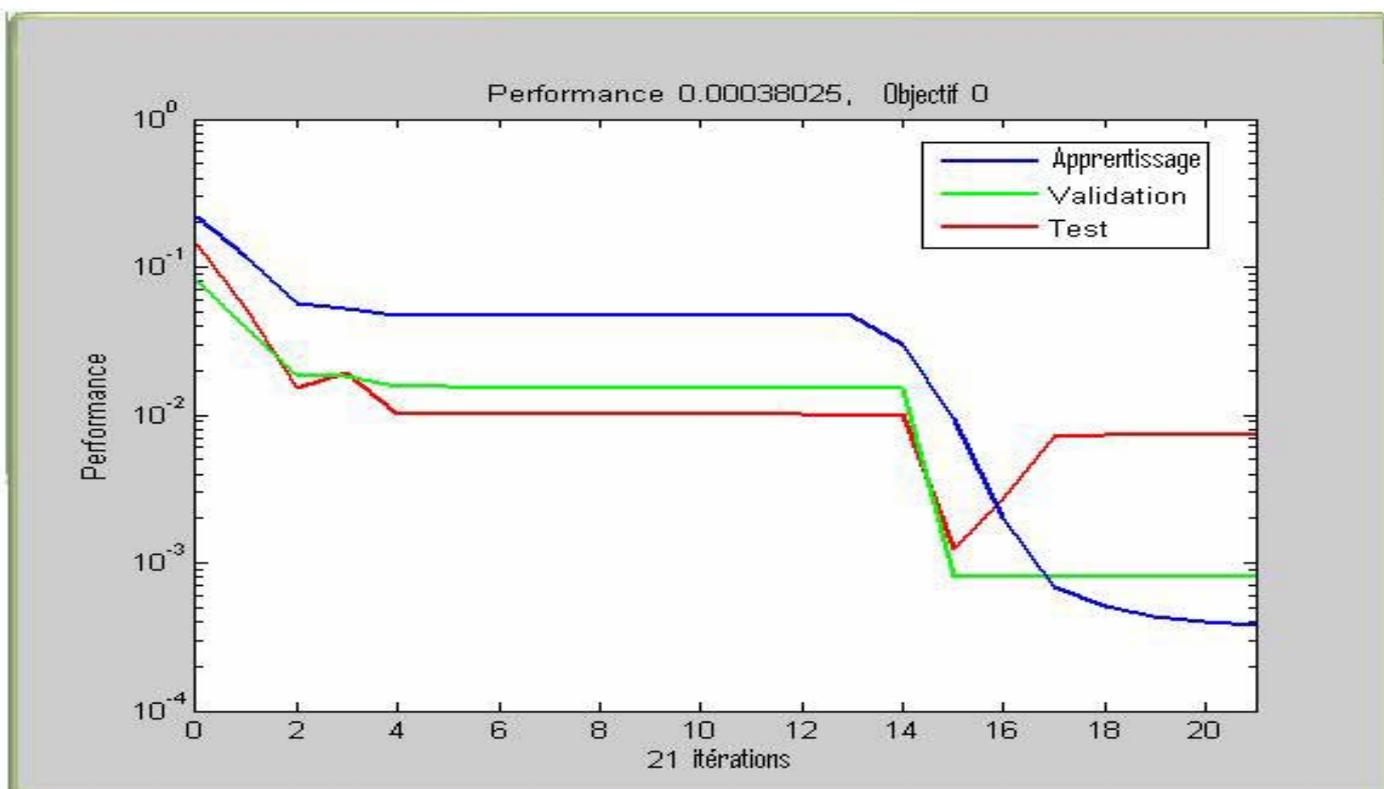


Figure V.1 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à un neurone caché, essai N°1.

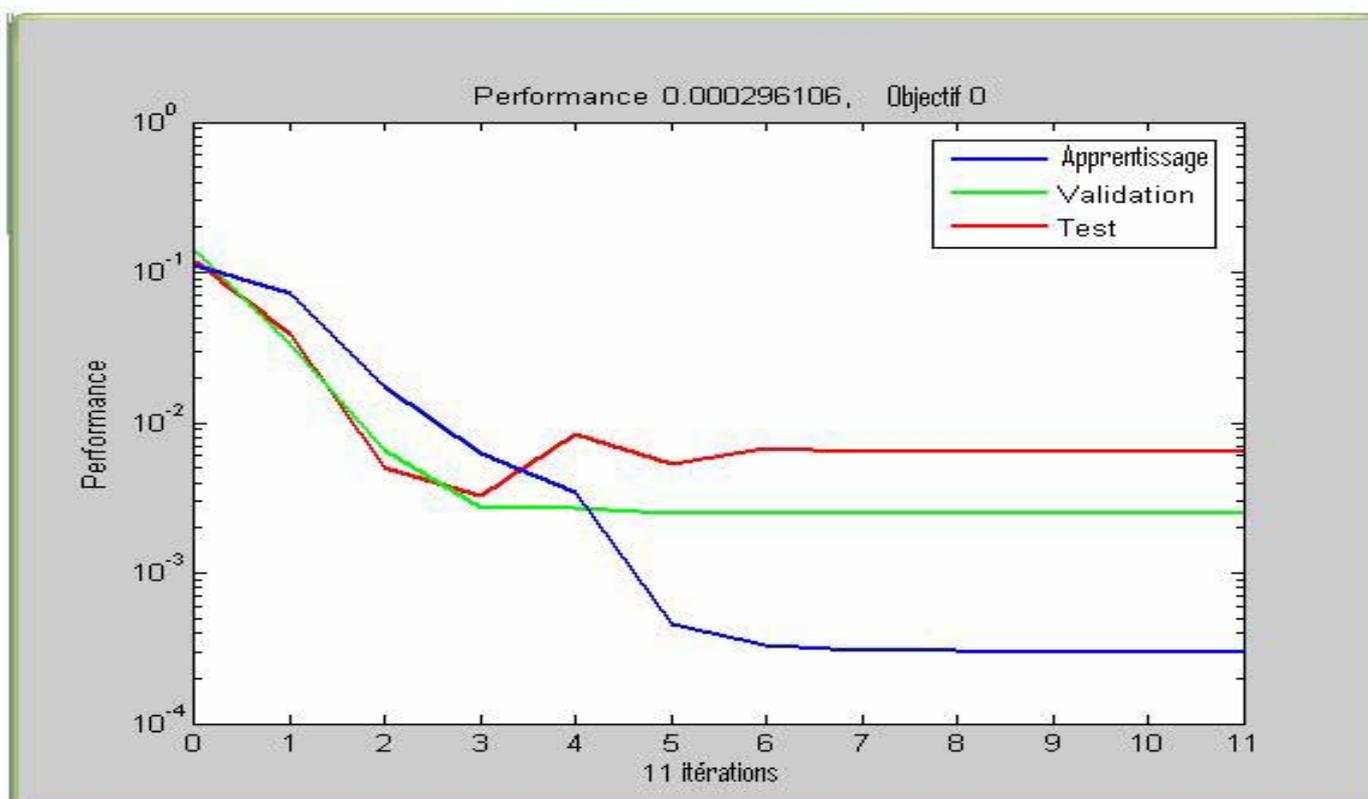


Figure V.2 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à un neurone caché, essai N°2.

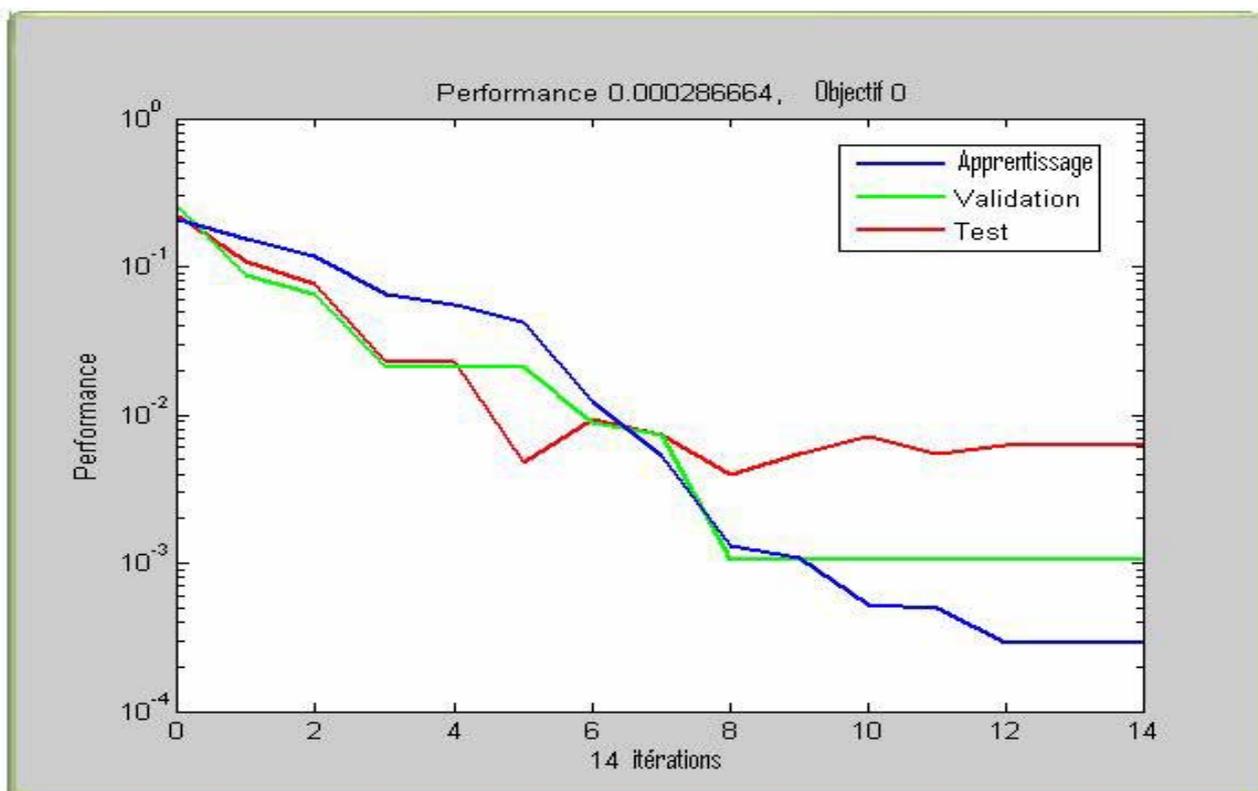


Figure V.3 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à un neurone caché, essai N°3.

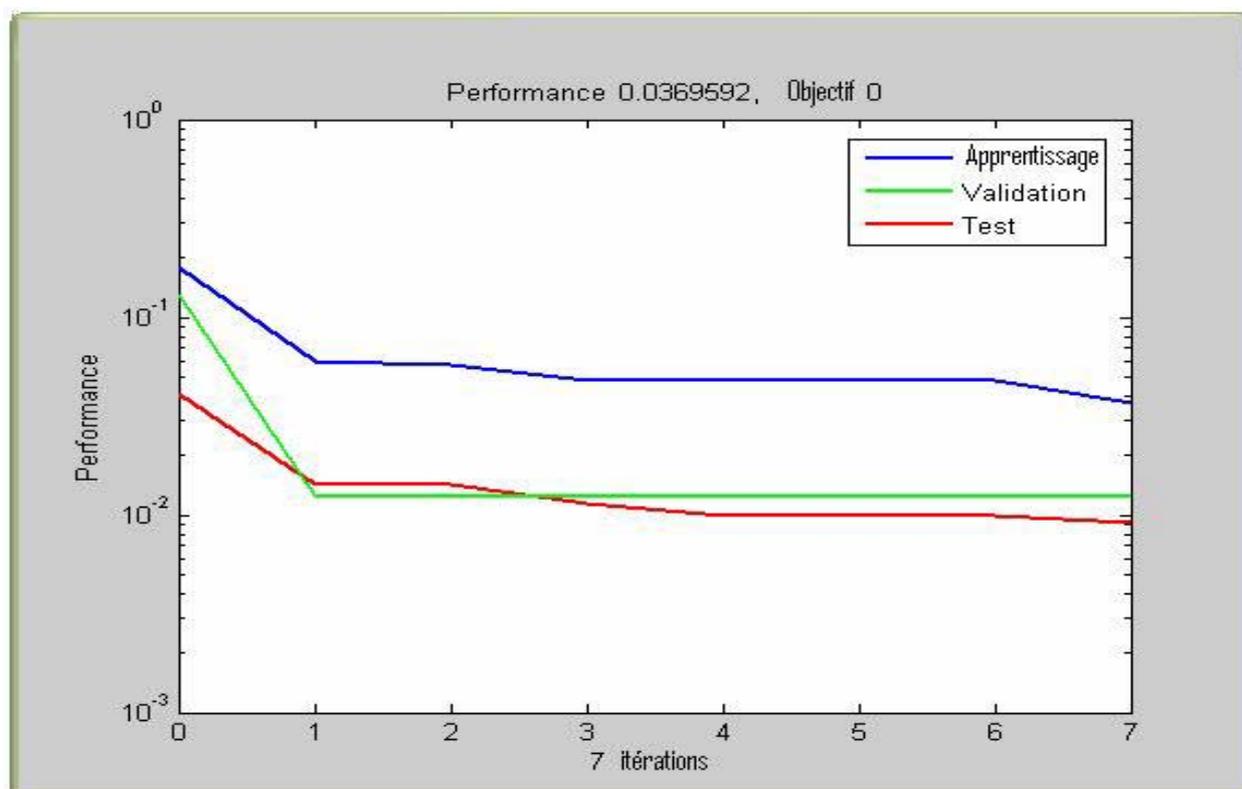


Figure V.4 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à un neurone caché, essai N°4.

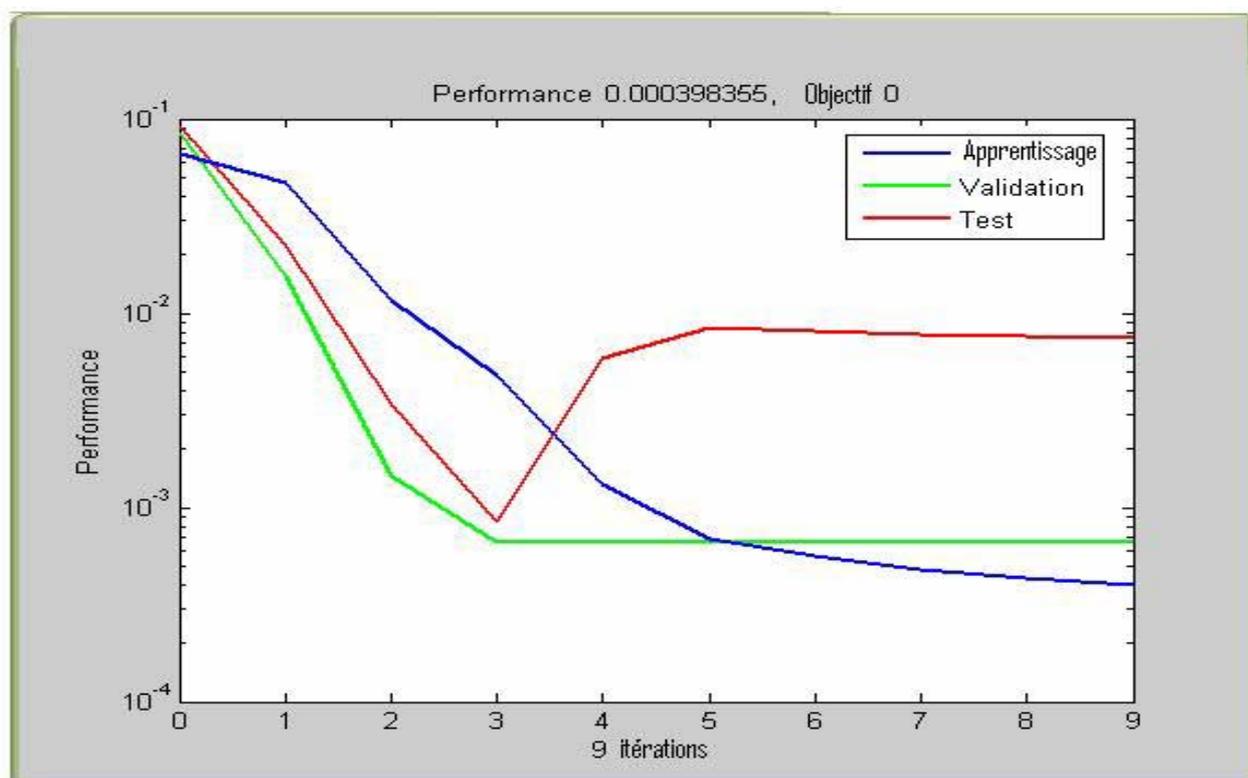


Figure V.5 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à un neurone caché, essai N°5.

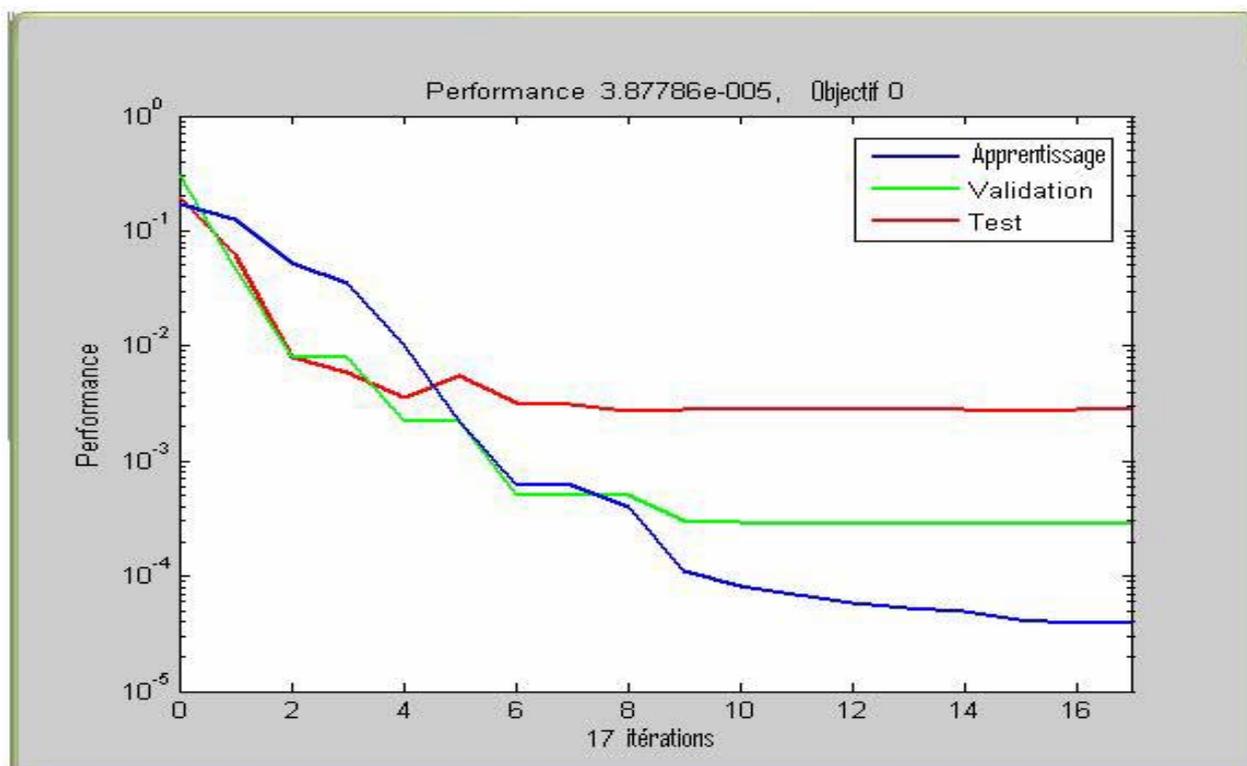


Figure V.6 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à deux neurones cachés, essai N°1.

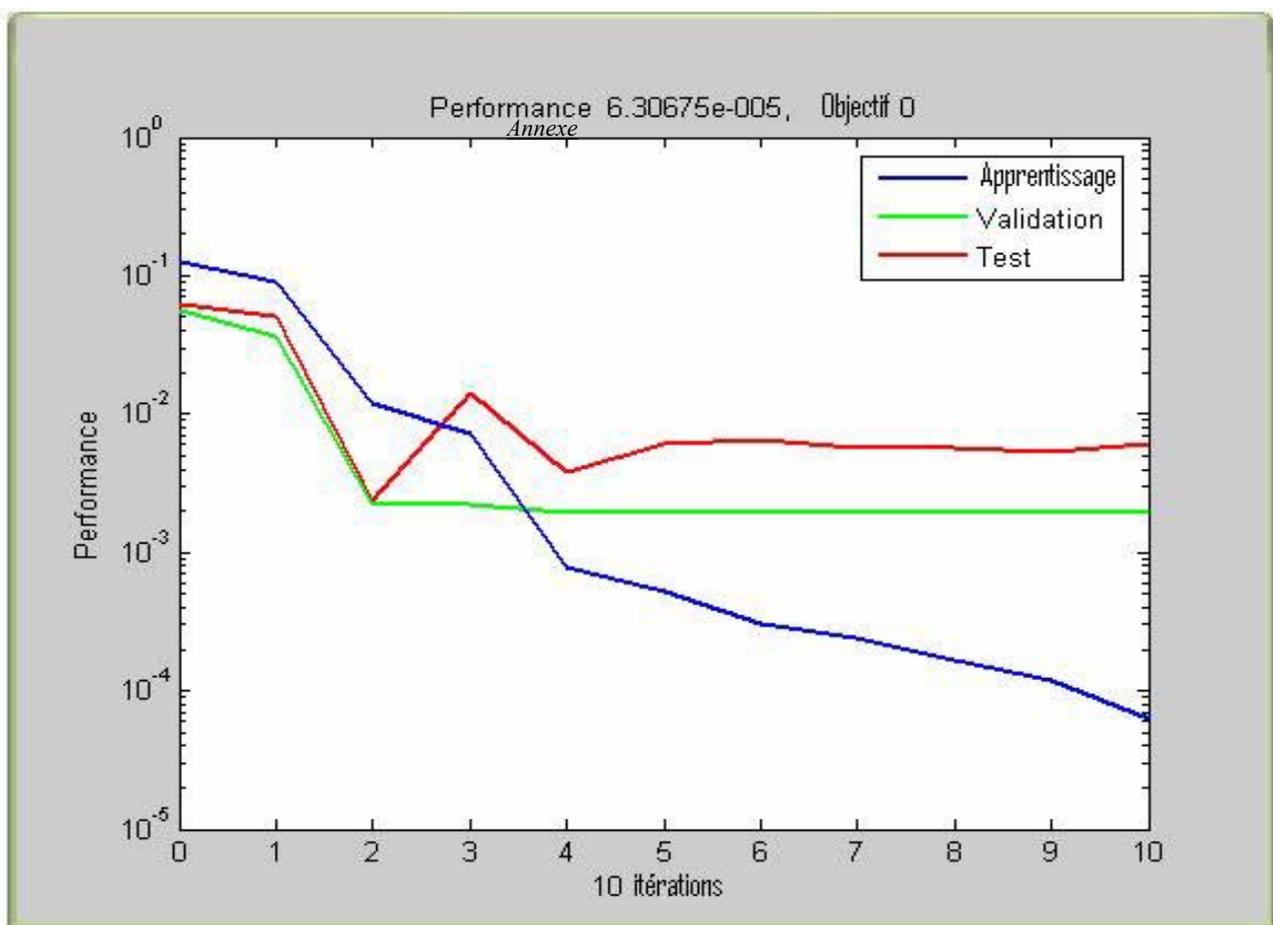


Figure V.7 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à deux neurones cachés, essai N°2.

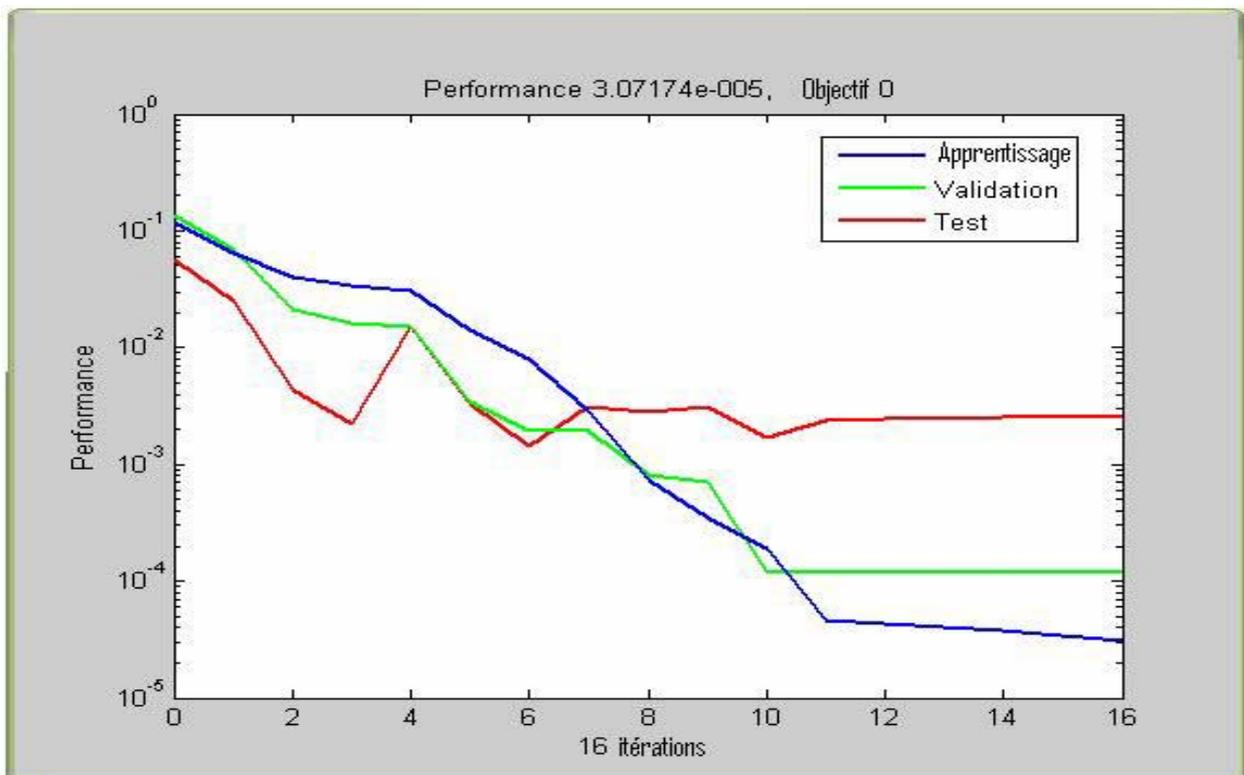


Figure V.8 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à deux neurones cachés, essai N°3.

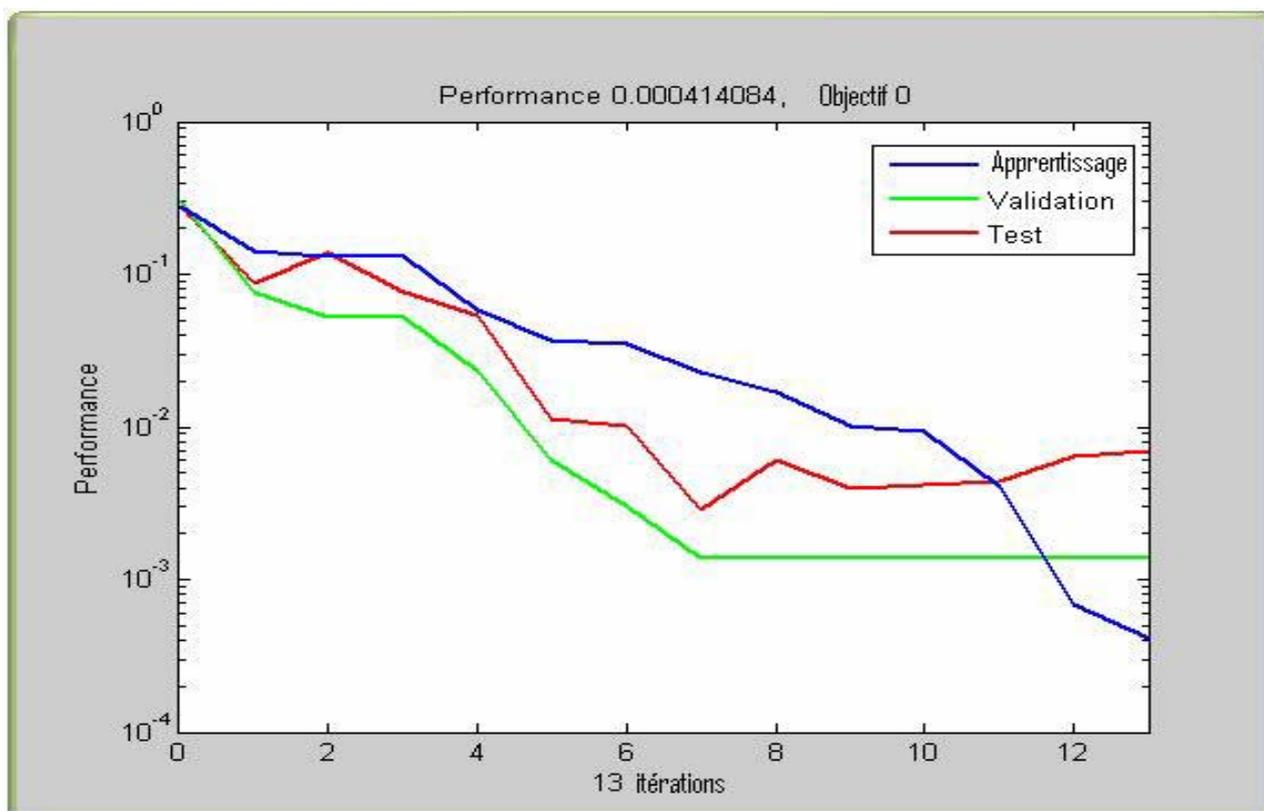


Figure V.9 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à deux neurones cachés, essai N°4.

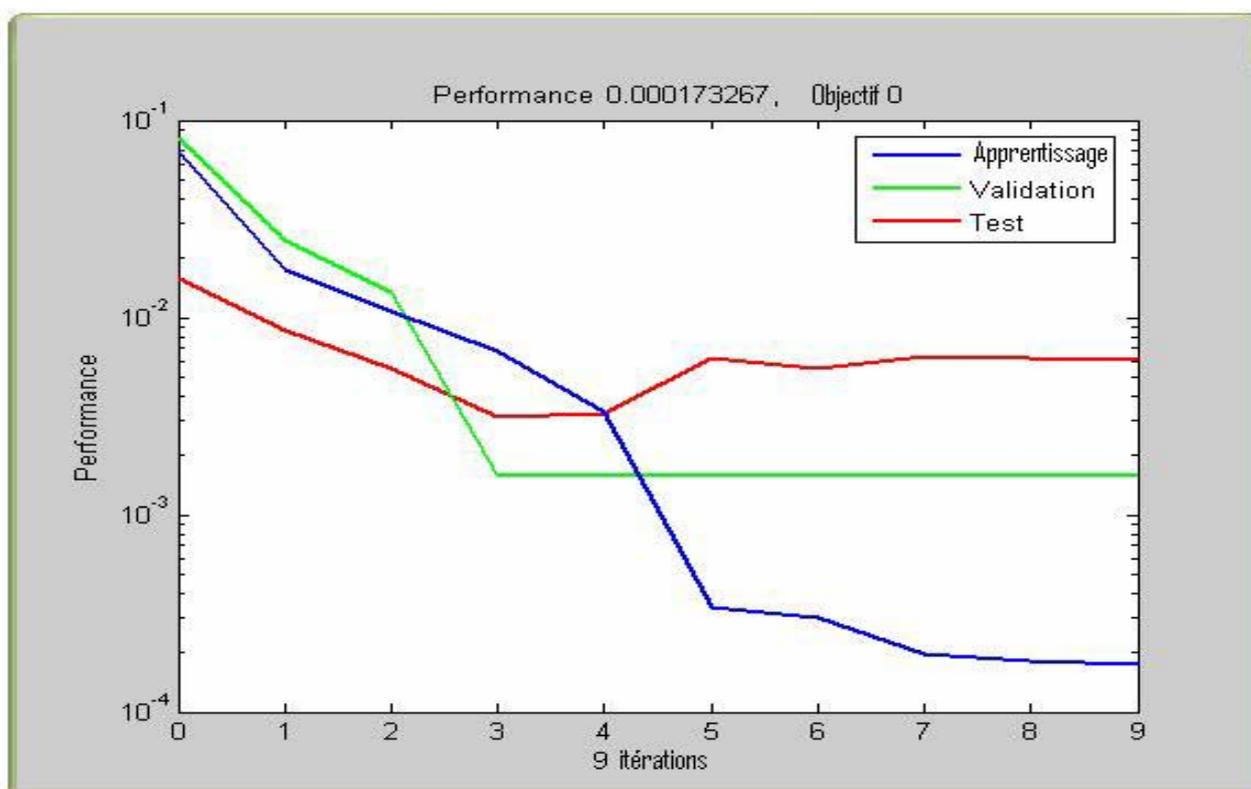


Figure V.10 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à deux neurones cachés, essai N°5.

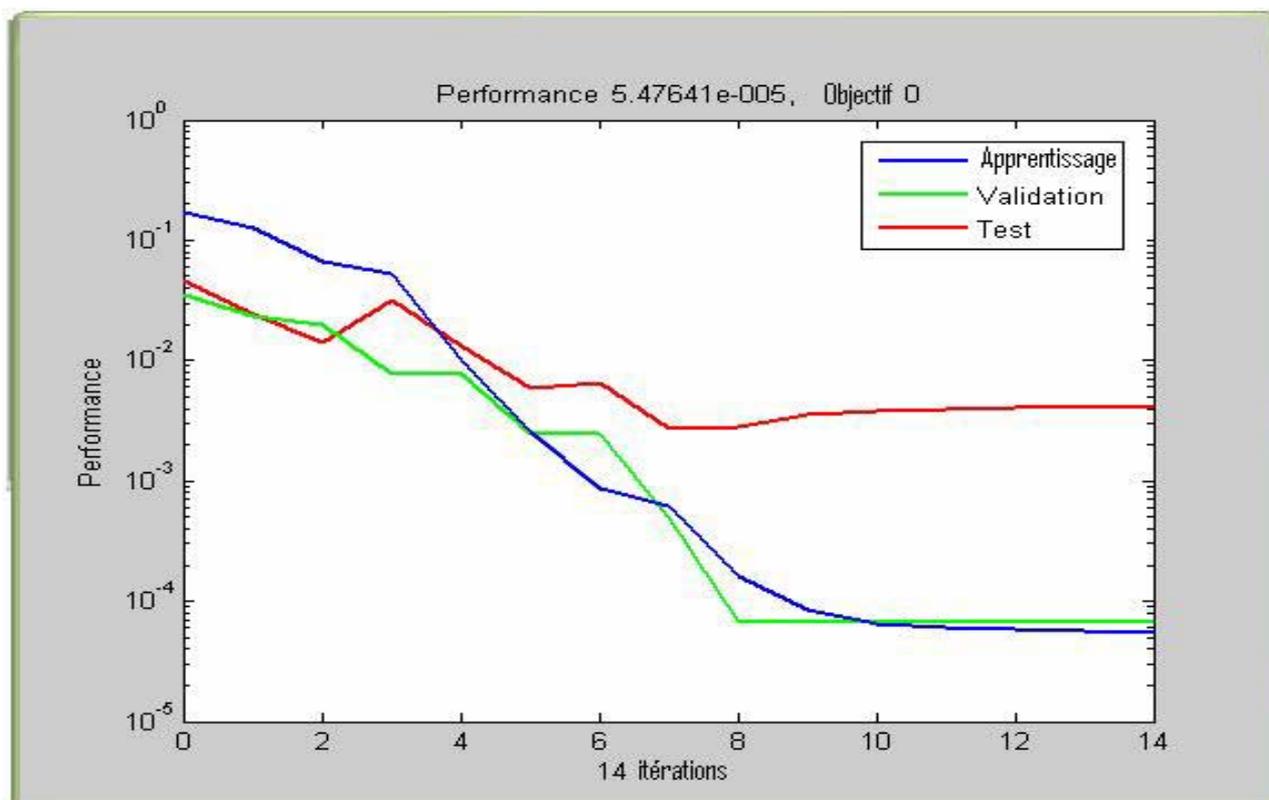


Figure V.11 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à trois neurones cachés, essai N°1.

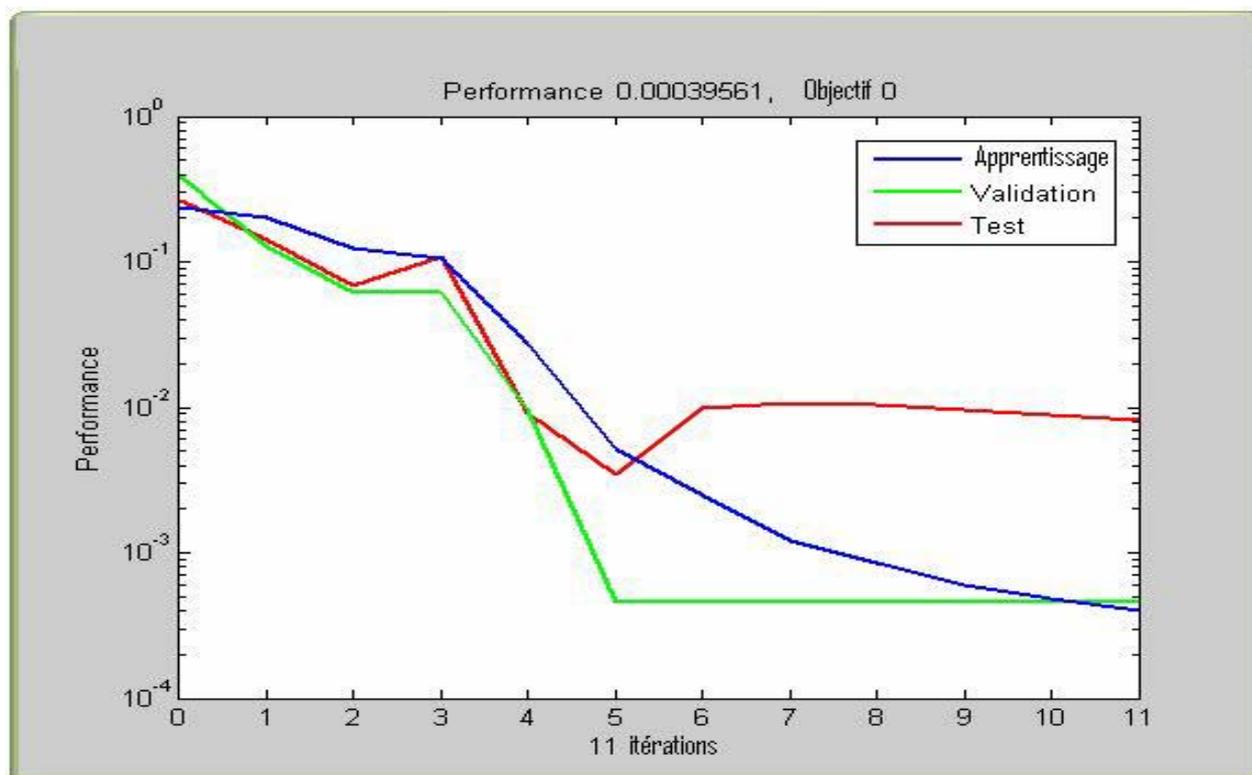


Figure V.12 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à trois neurones cachés, essai N°2.

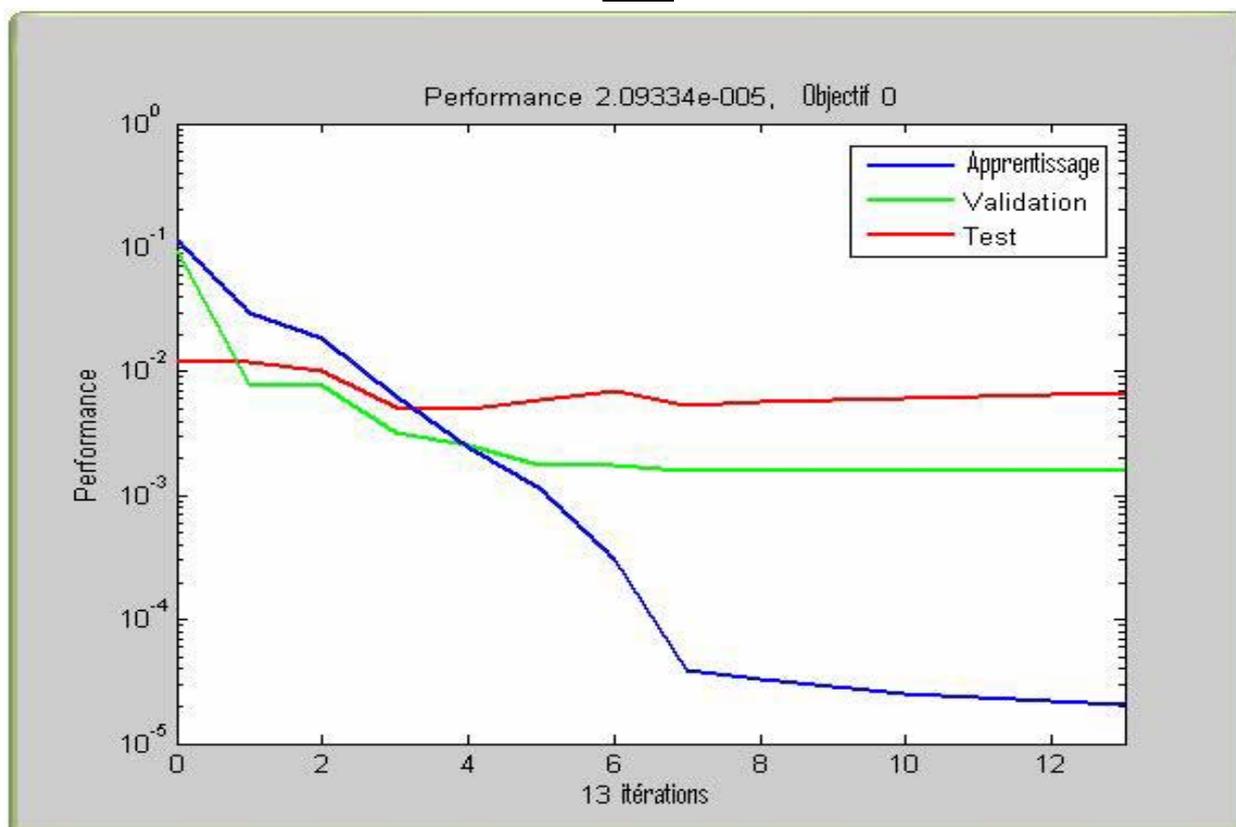


Figure V.13 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à trois neurones cachés, essai N°3.

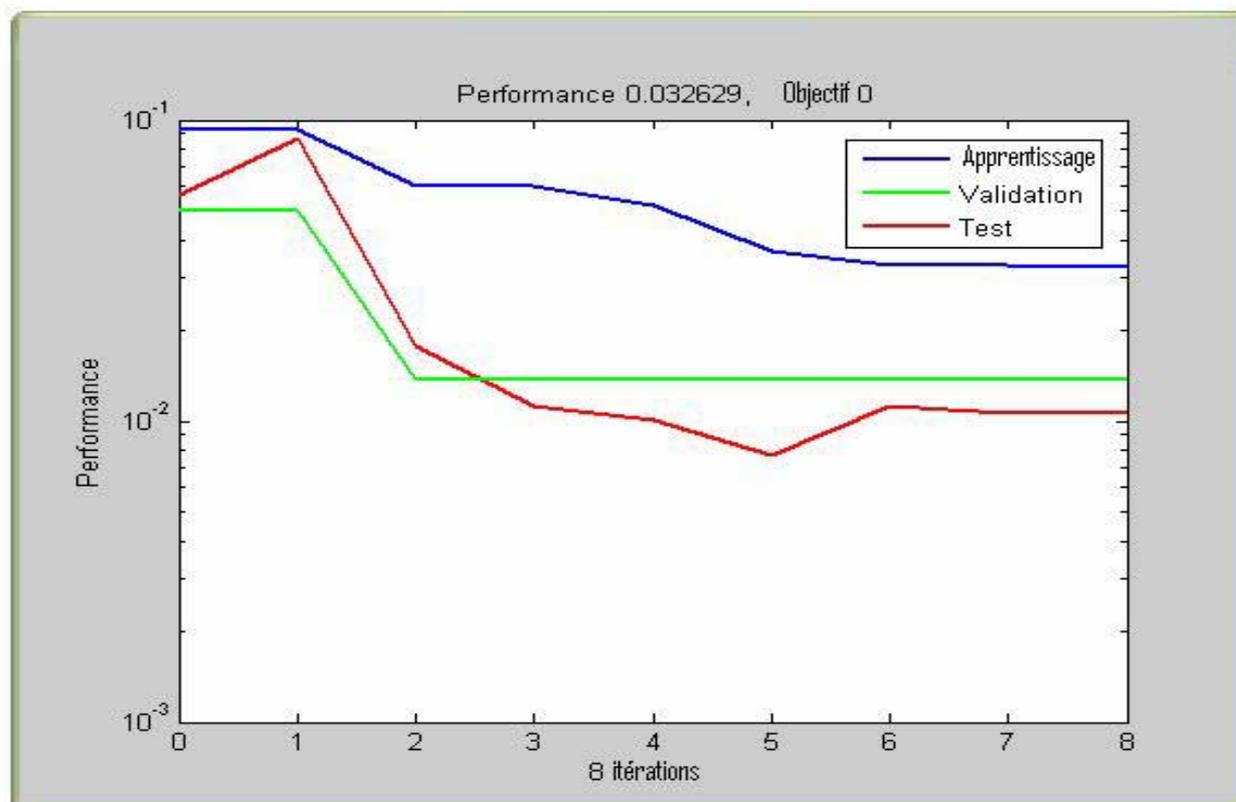


Figure V.14 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à trois neurones cachés, essai N°4.

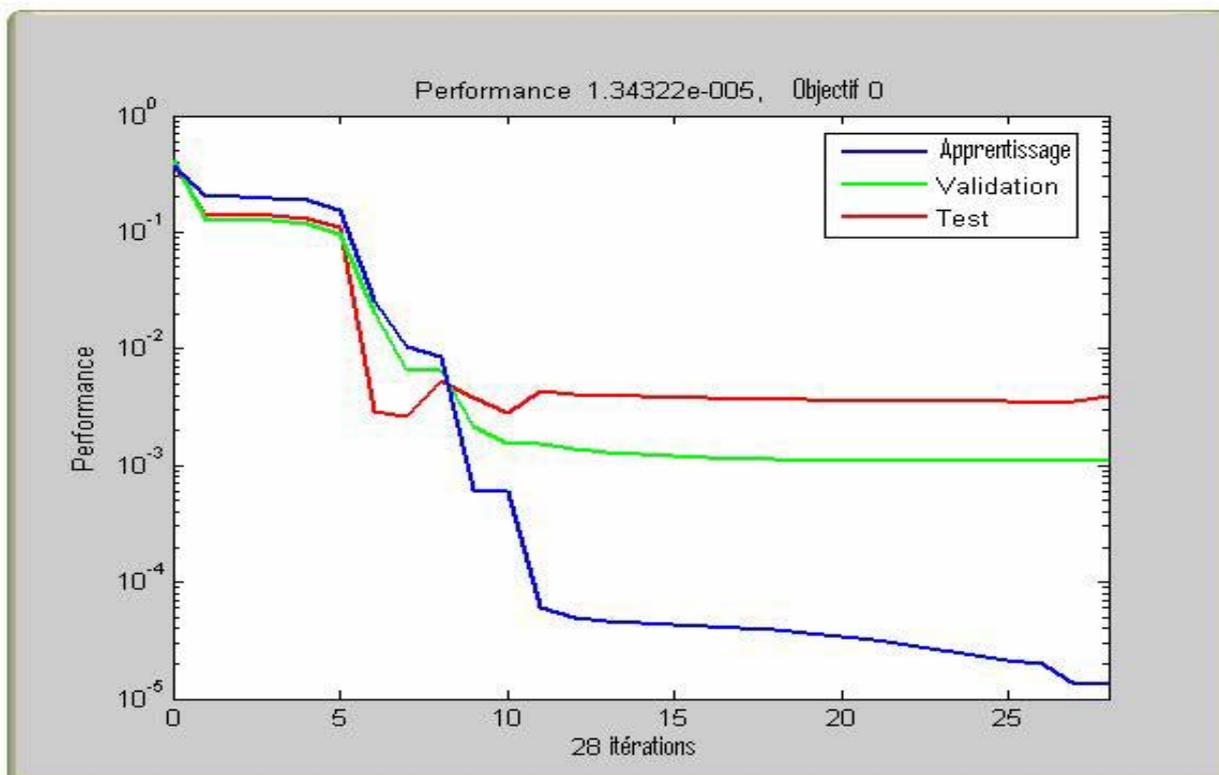


Figure V.15 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à trois neurones cachés, essai N°5.

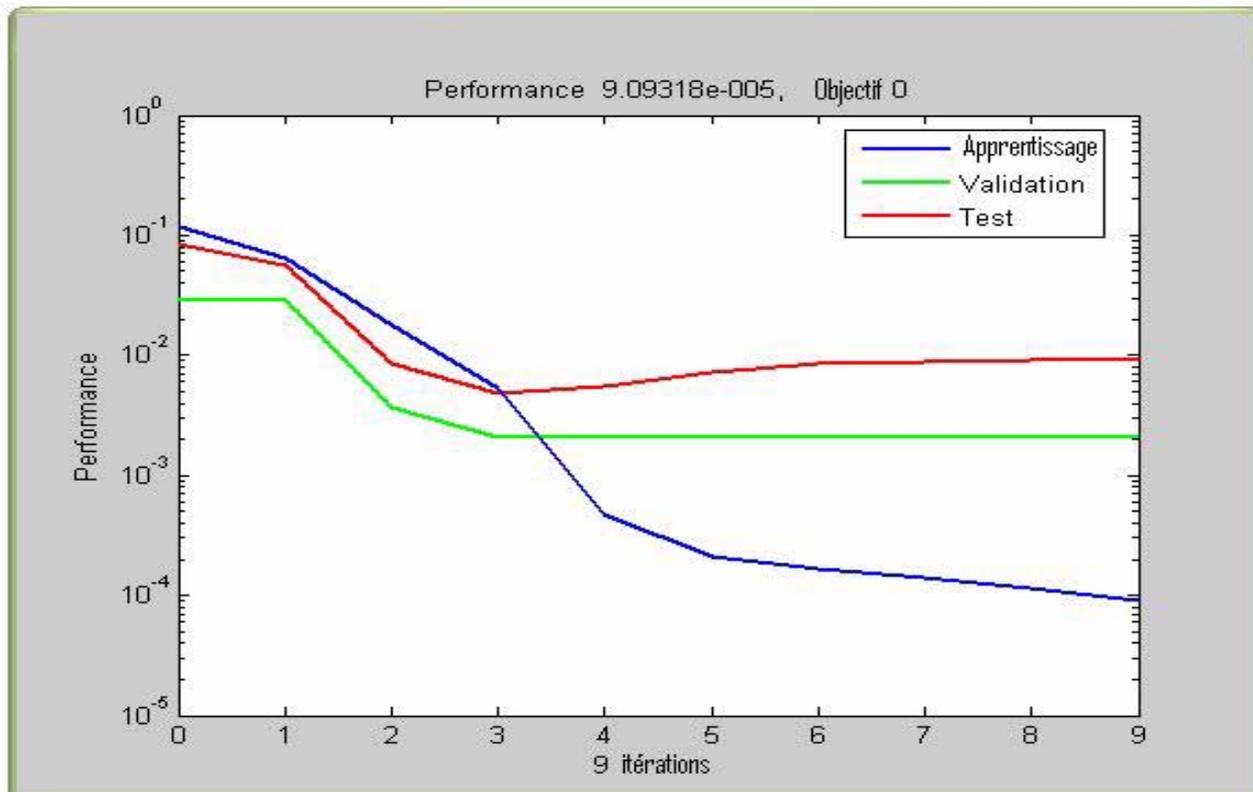


Figure V.16 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à quatre neurones cachés, essai N°1.

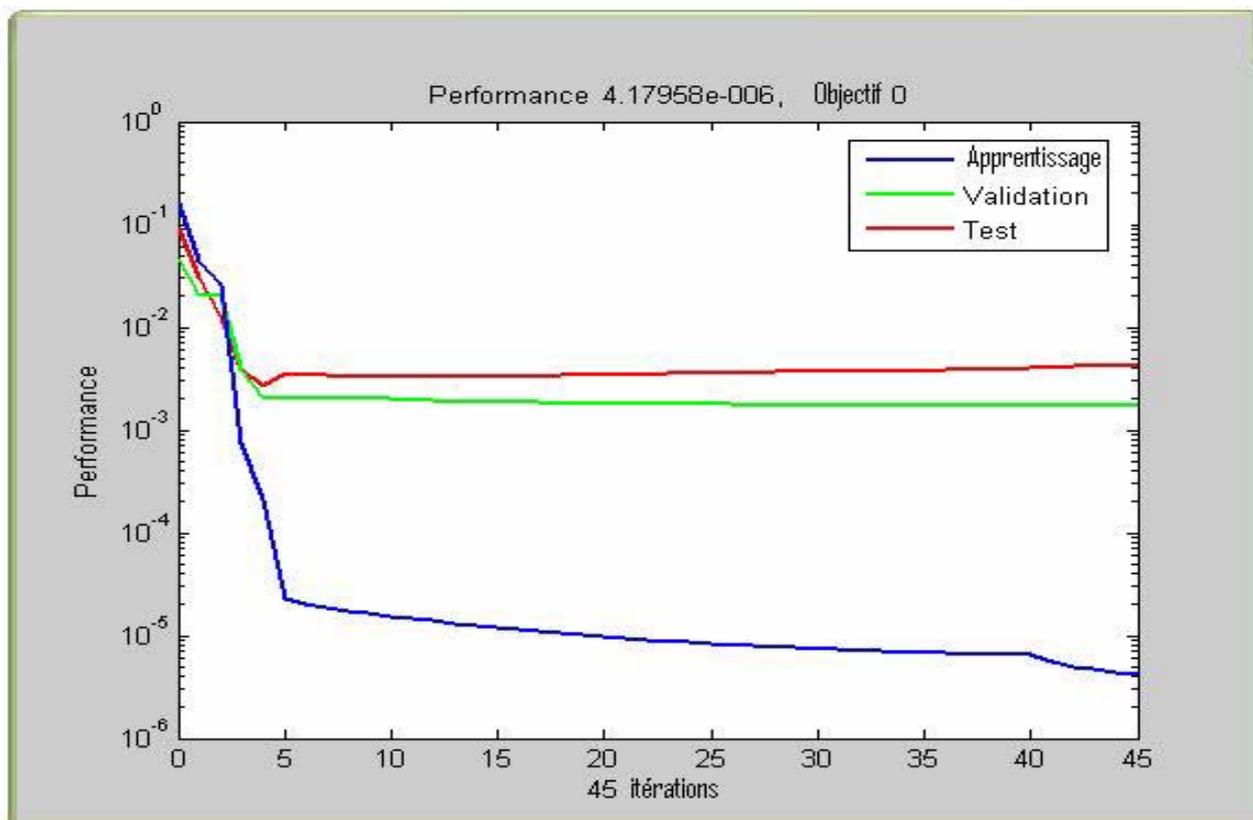


Figure V.17 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à quatre neurones cachés, essai N°2.

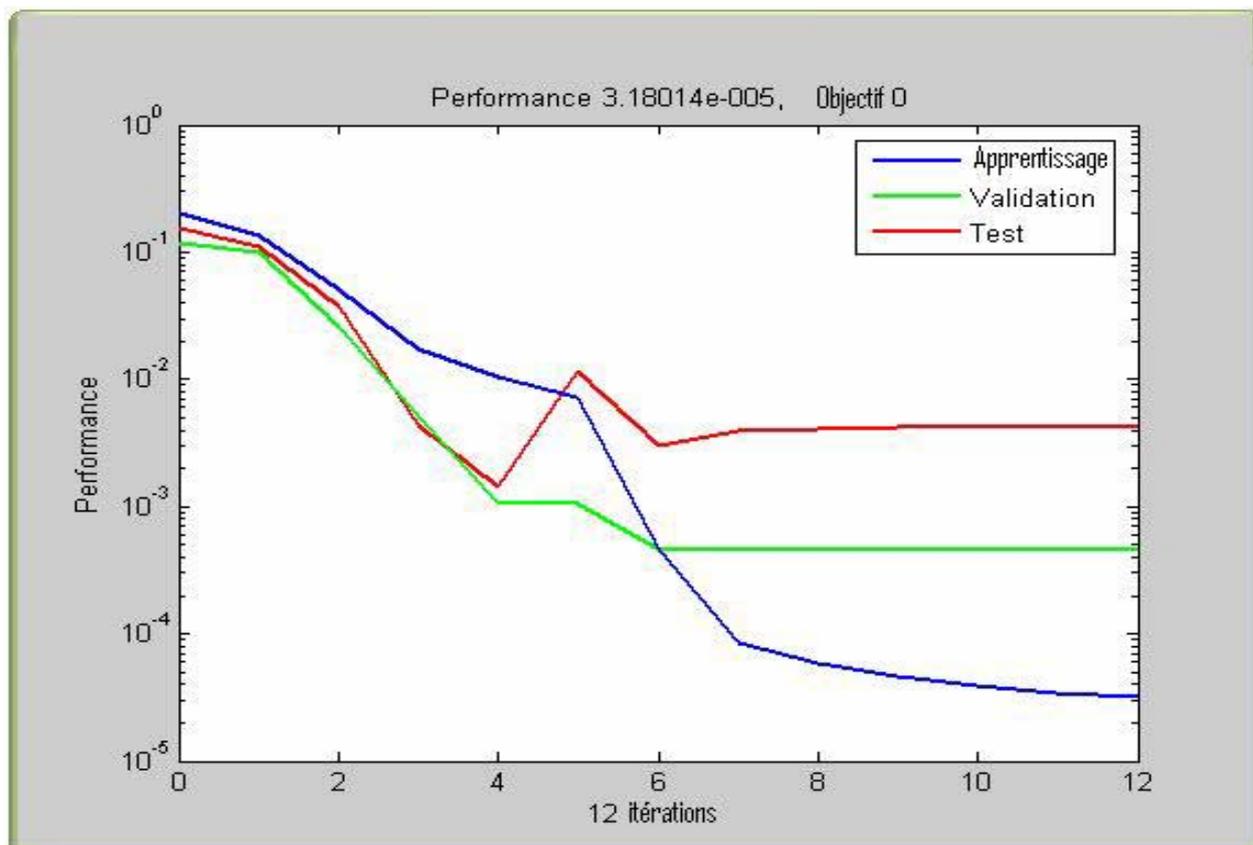


Figure V.18 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à quatre neurones cachés, essai N°3.

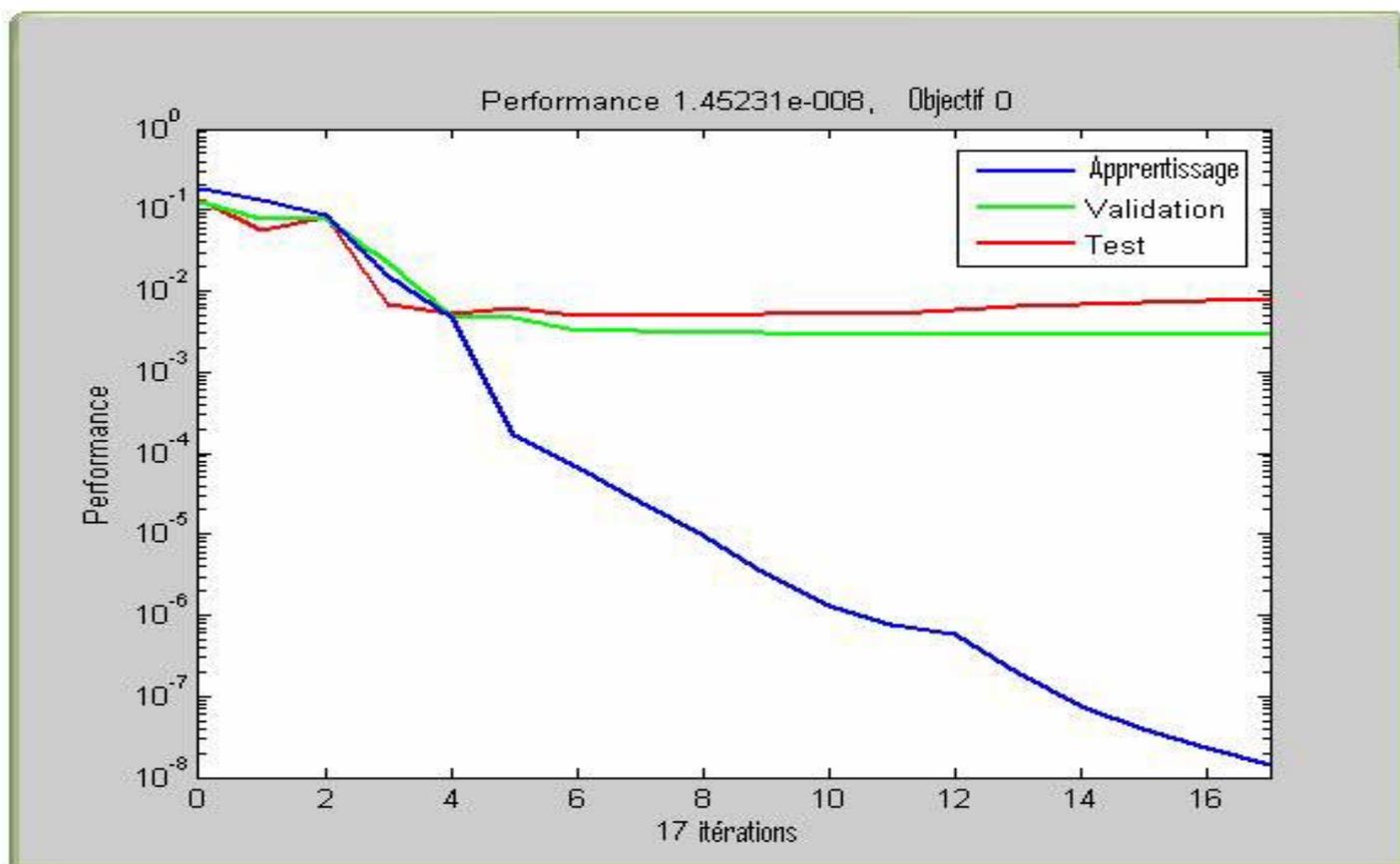


Figure V.19 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à quatre neurones cachés, essai N°4.

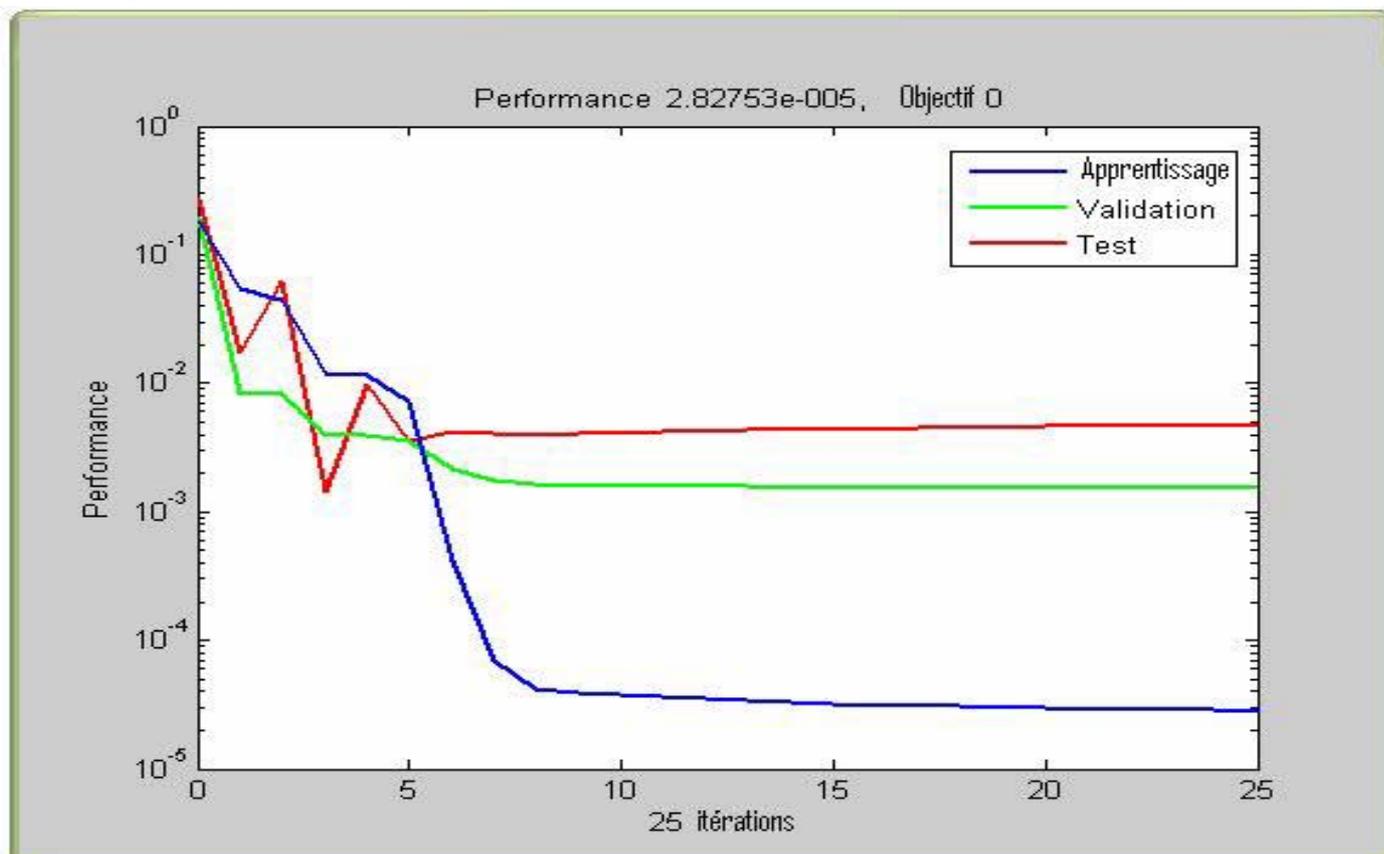


Figure V.20 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à quatre neurones cachés, essai N°5.

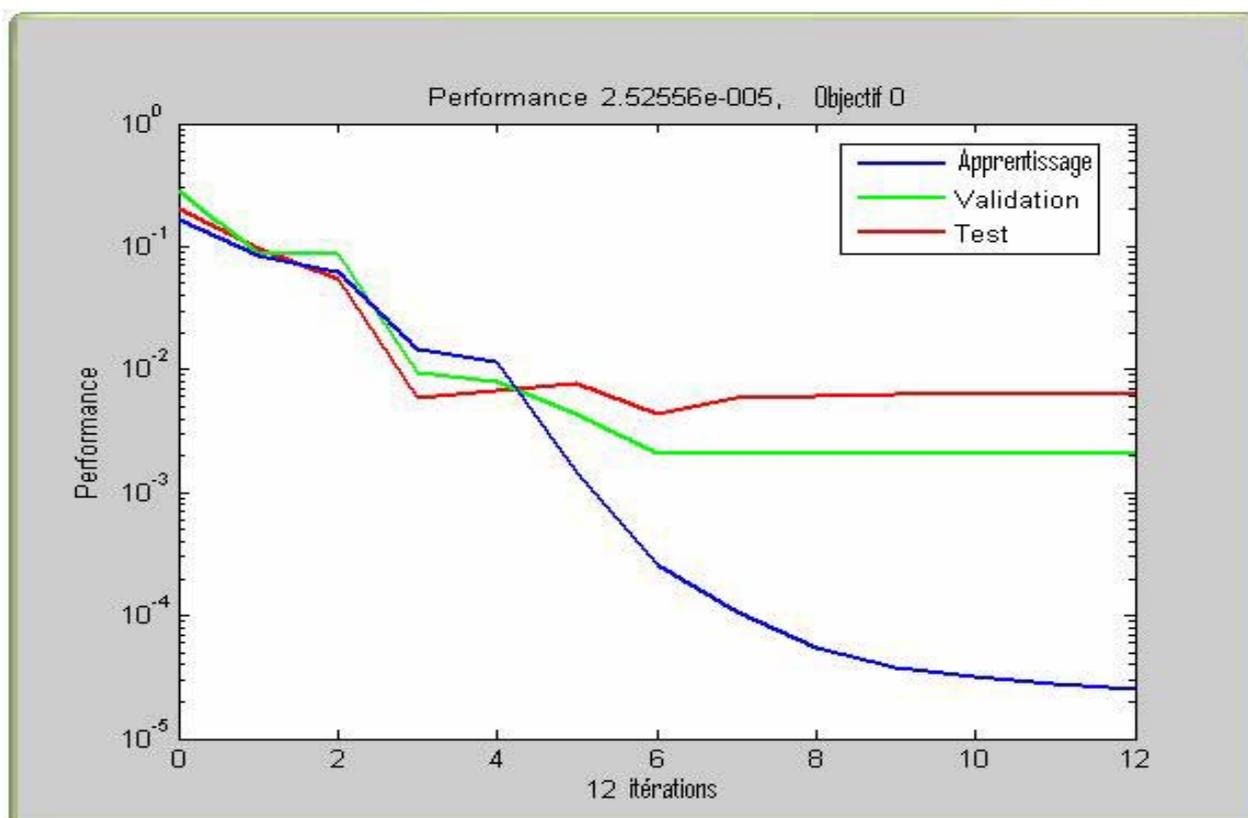


Figure V.21 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à cinq neurones cachés, essai N°1.

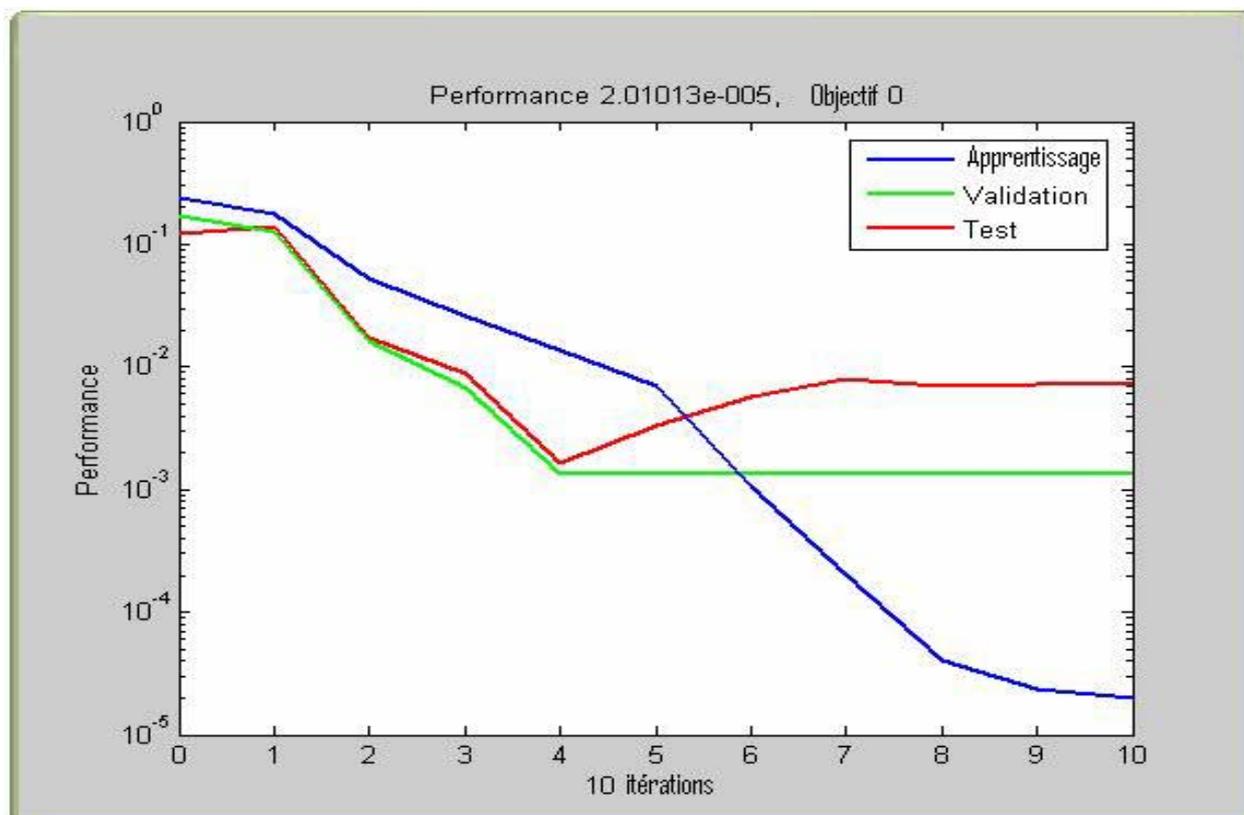


Figure V.22 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à cinq neurones cachés, essai N°2.

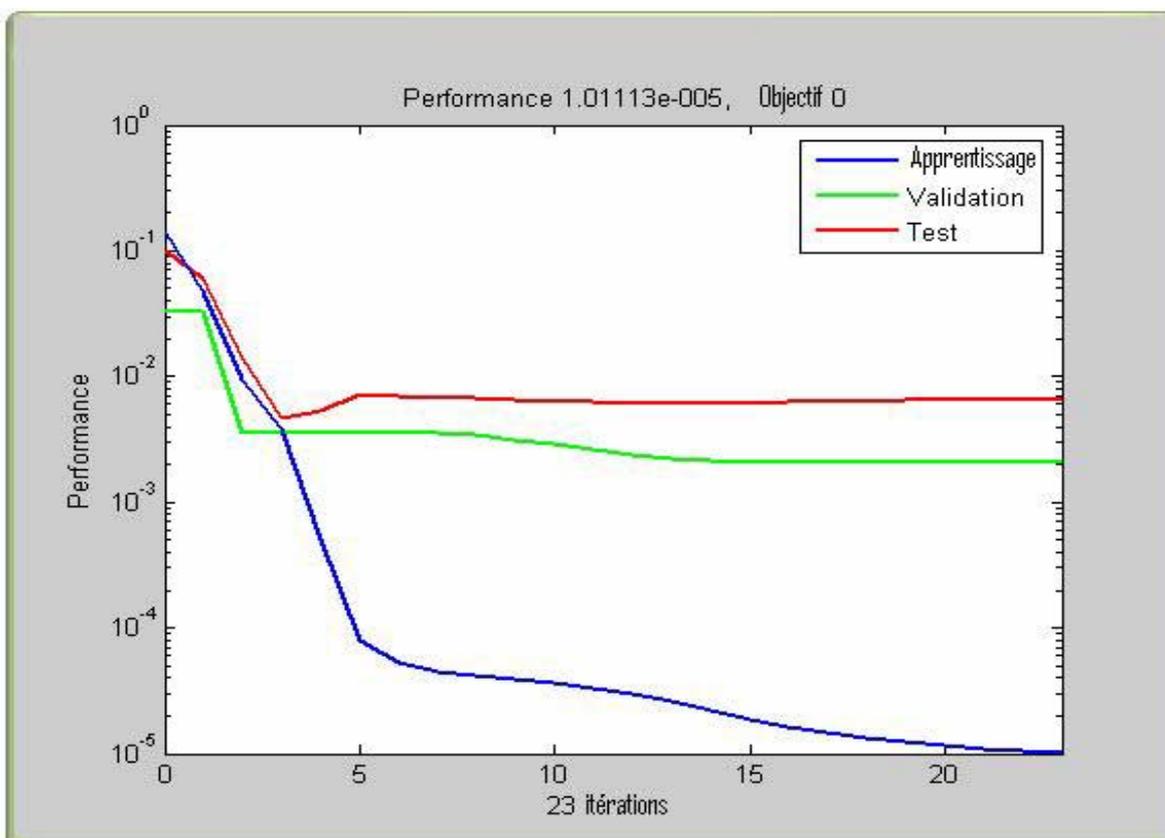


Figure V.23 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à cinq neurones cachés, essai N°3.

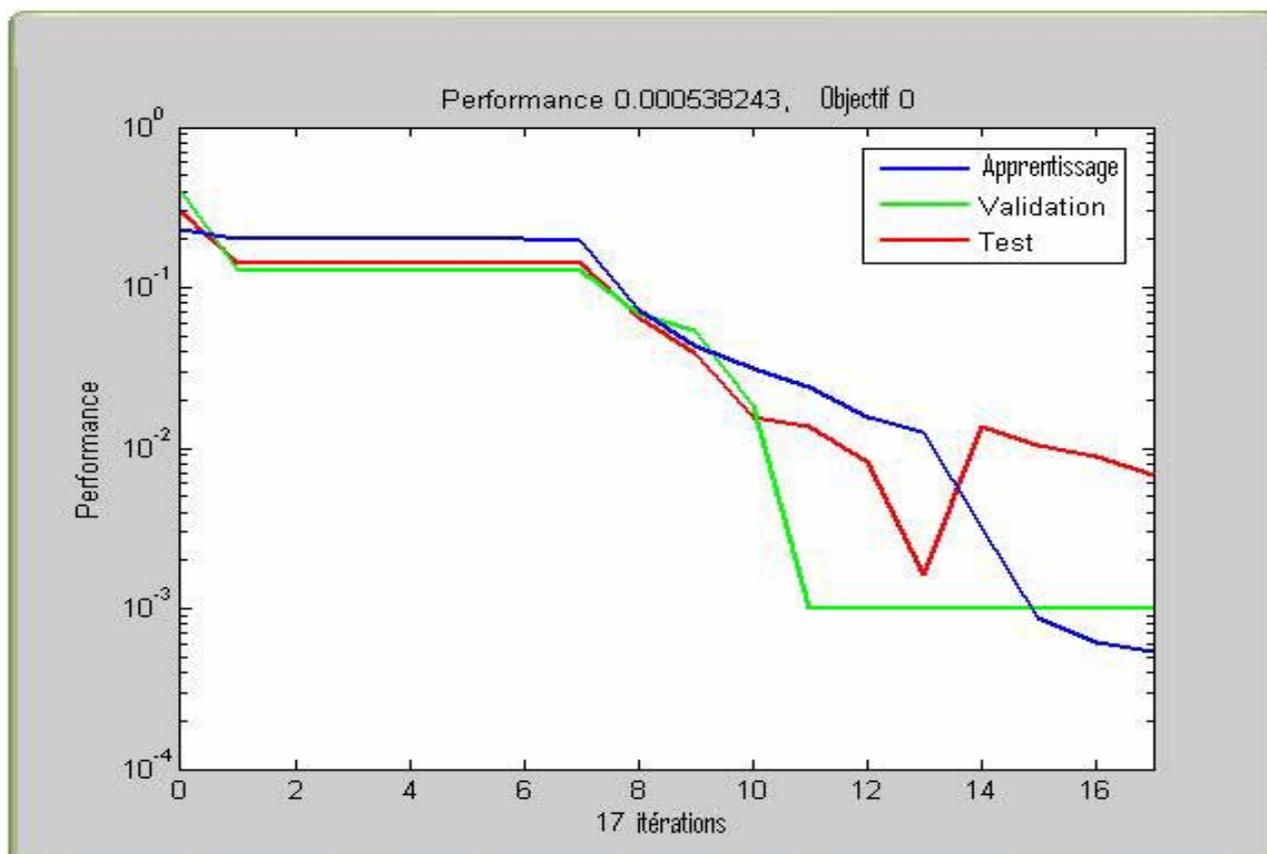


Figure V.24 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à cinq neurones cachés, essai N°4.

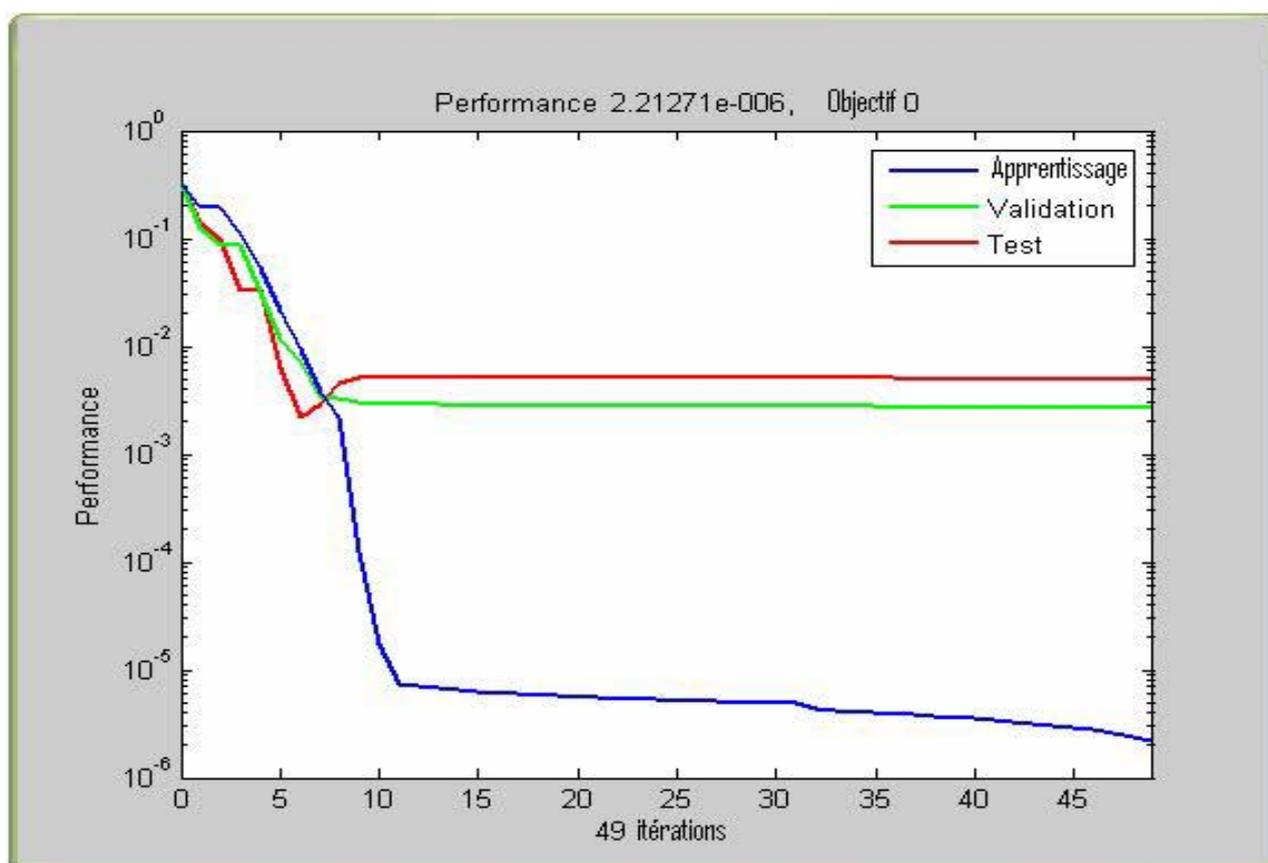


Figure V.25 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à cinq neurones cachés, essai N°5.

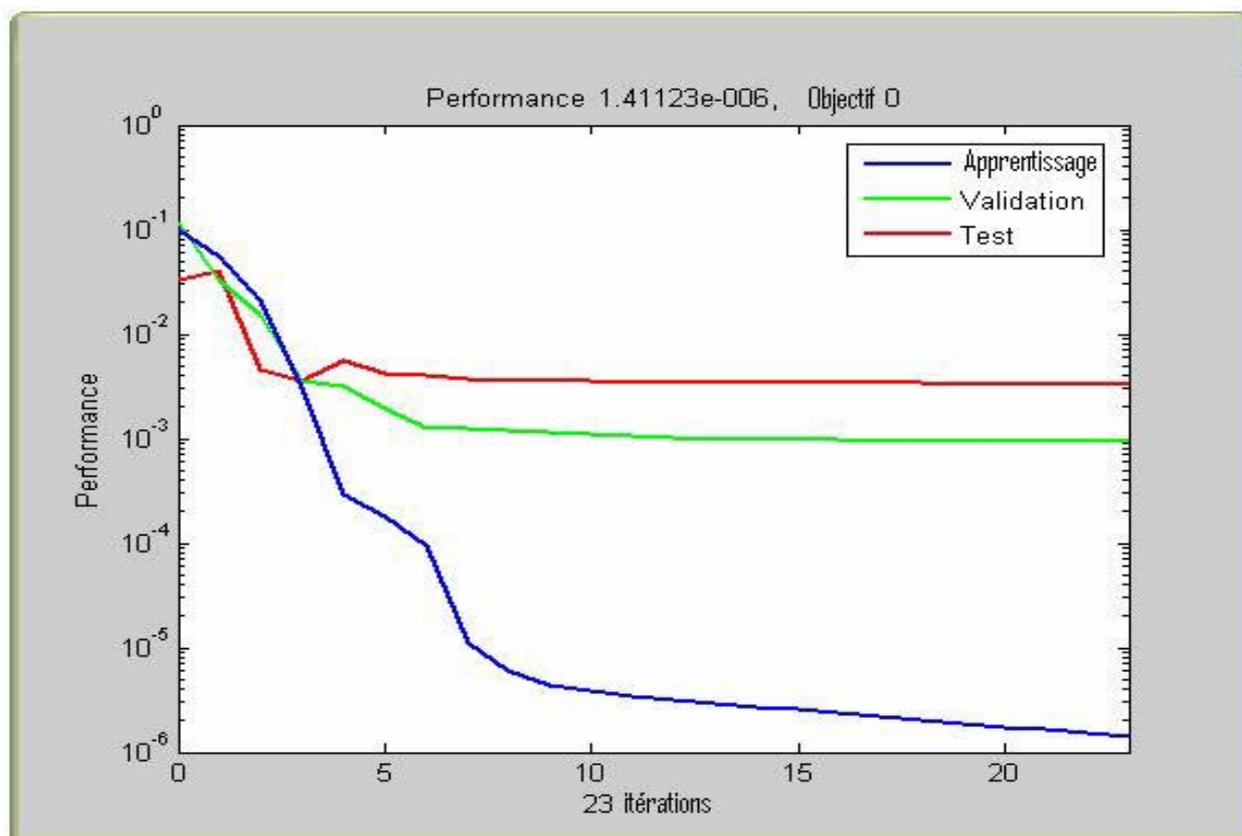


Figure V.26 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à six neurones cachés, essai N°1.

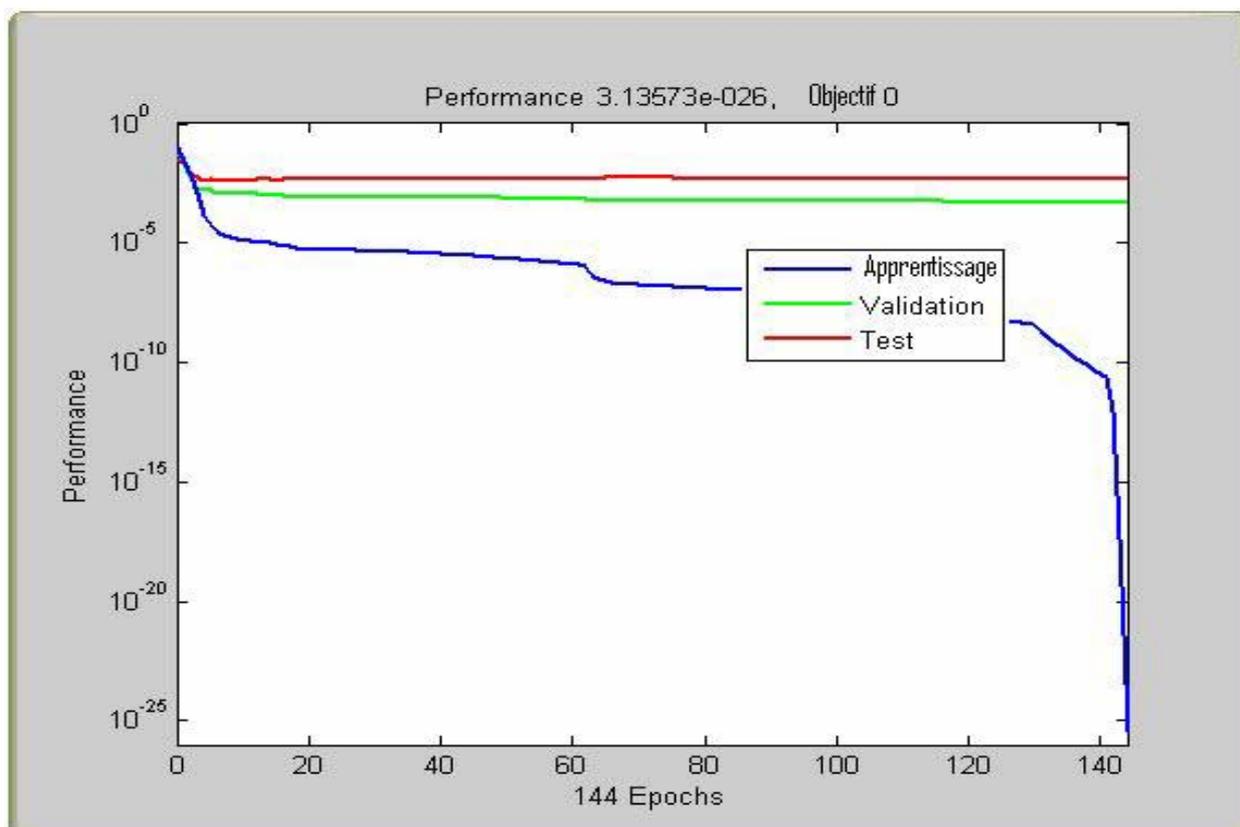


Figure V.27 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à six neurones cachés, essai N°2.

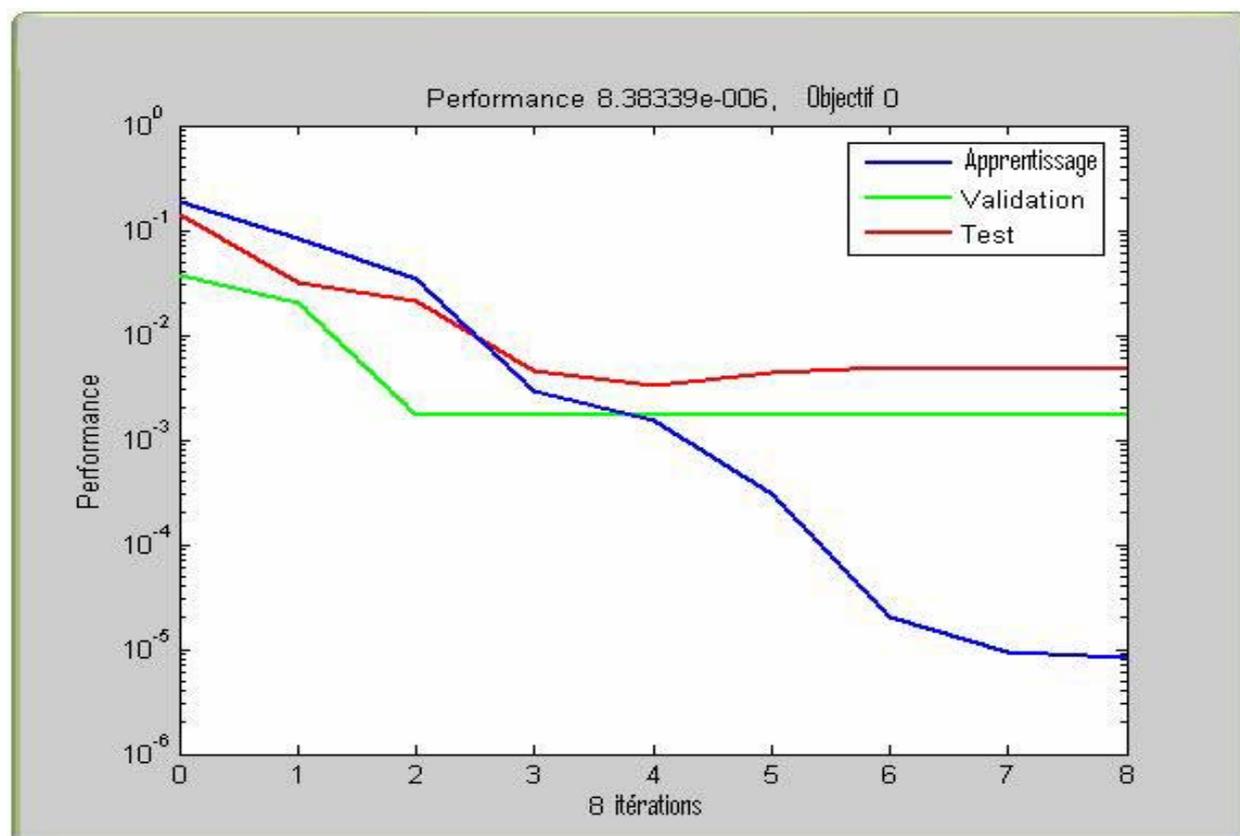


Figure V.28 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à six neurones cachés, essai N°3.

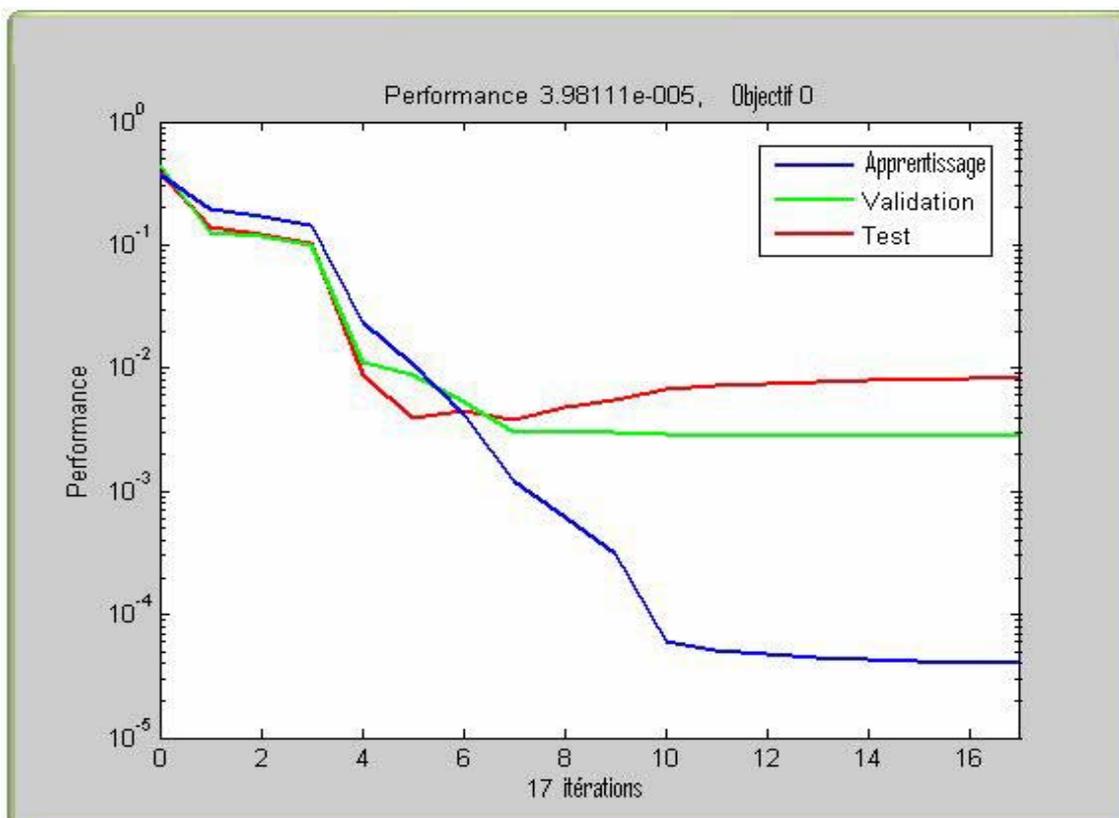


Figure V.29 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à six neurones cachés, essai N°4.

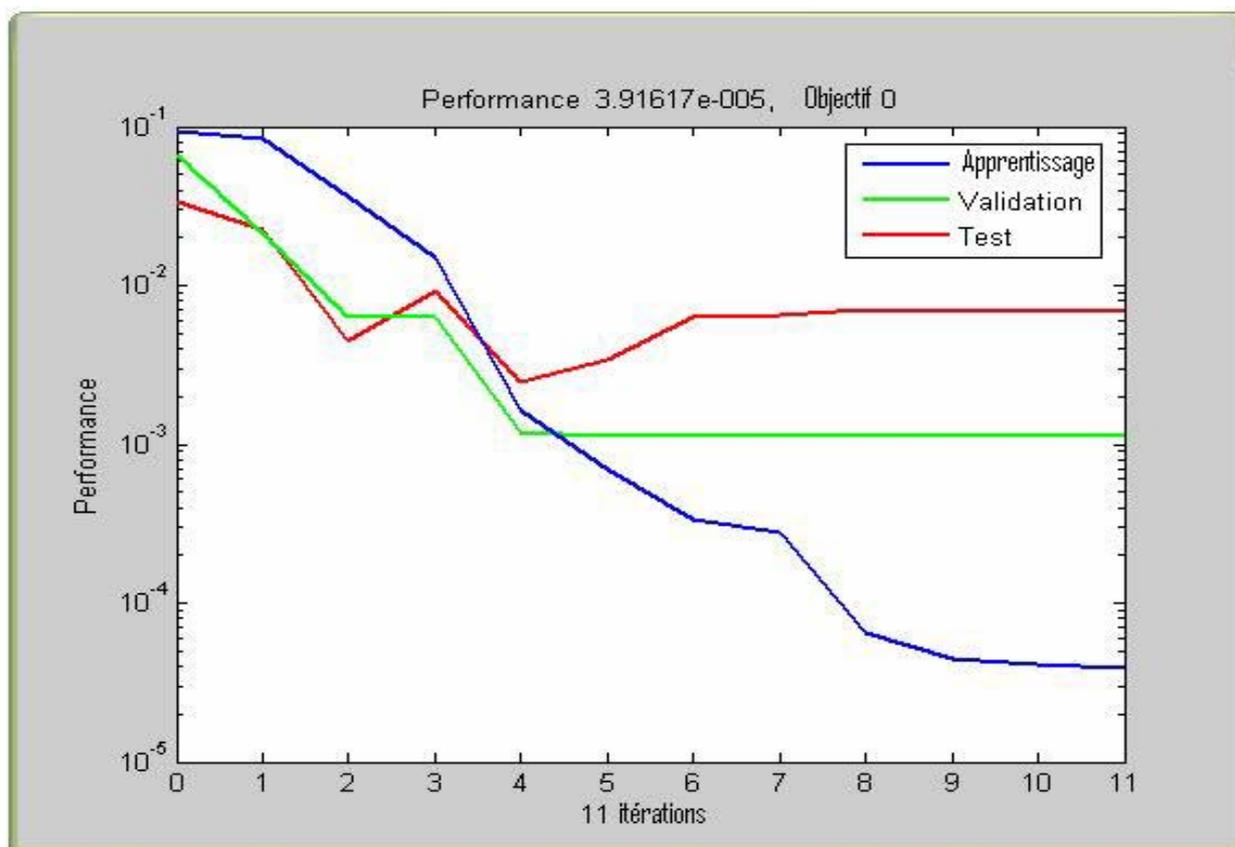


Figure V.30 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à six neurones cachés, essai N°5.

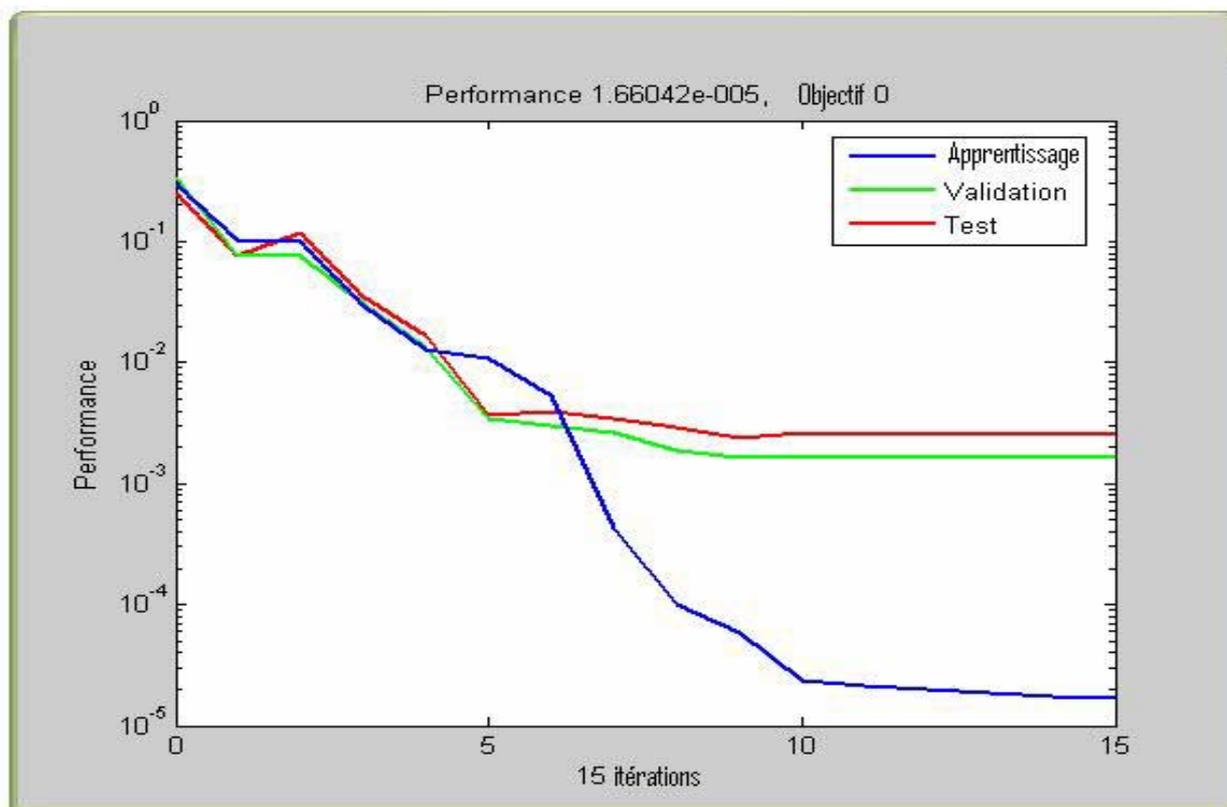


Figure V.31 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à sept neurones cachés, essai N°1.

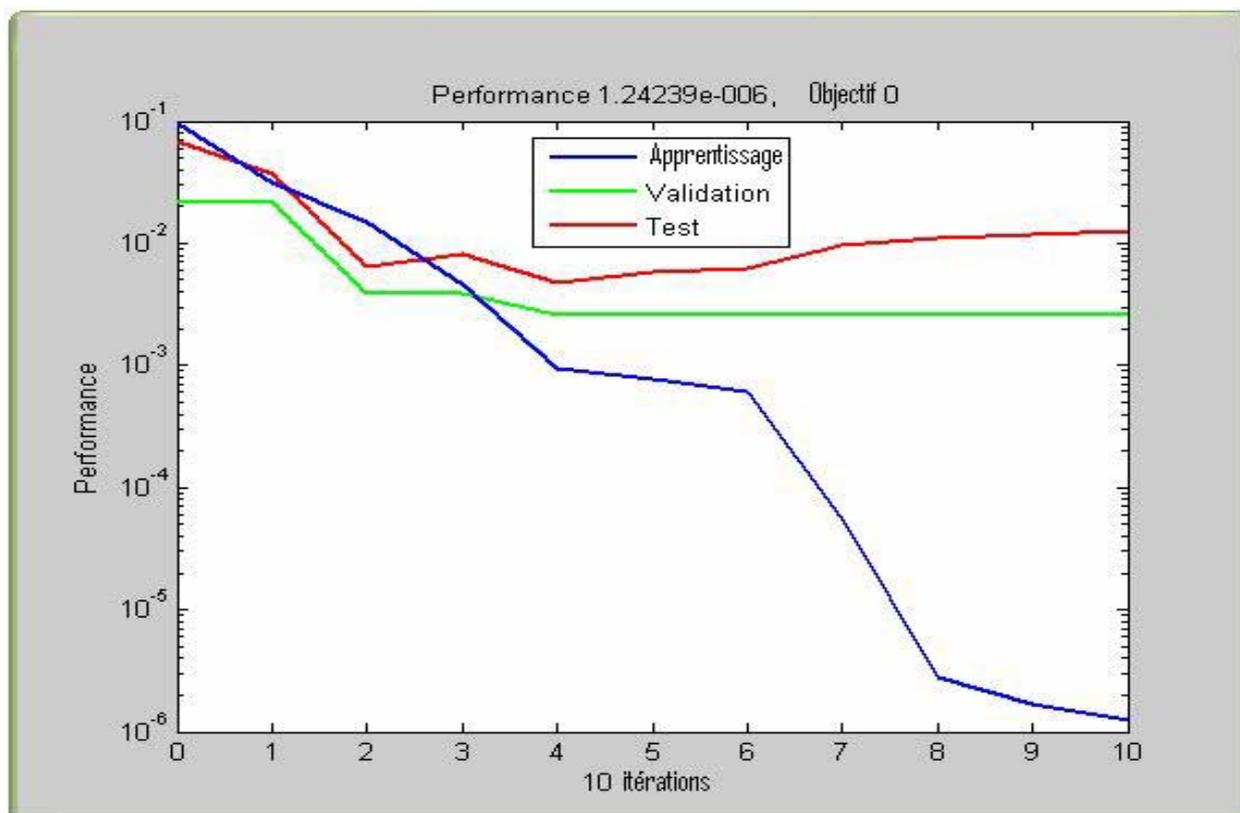


Figure V.32 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à sept neurones cachés, essai N°2.

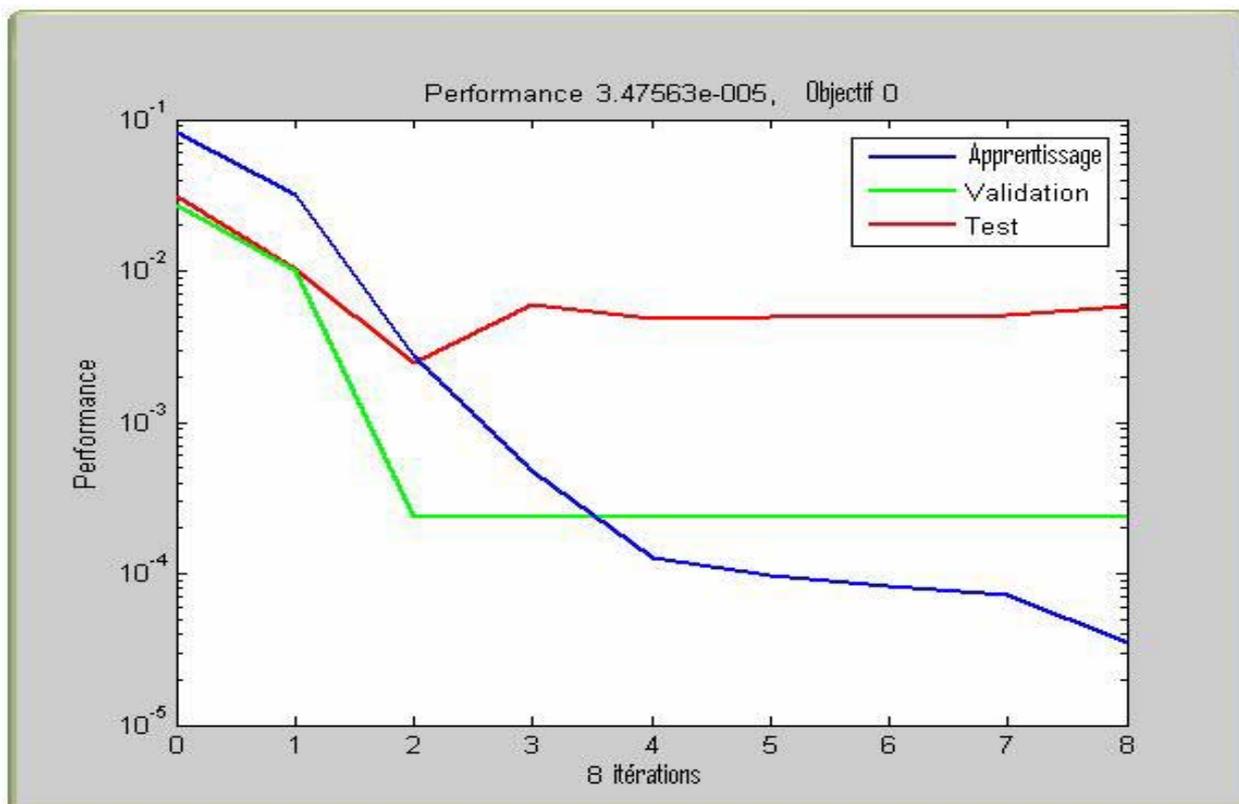


Figure V.33 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à neurones cachés, essai N°3.

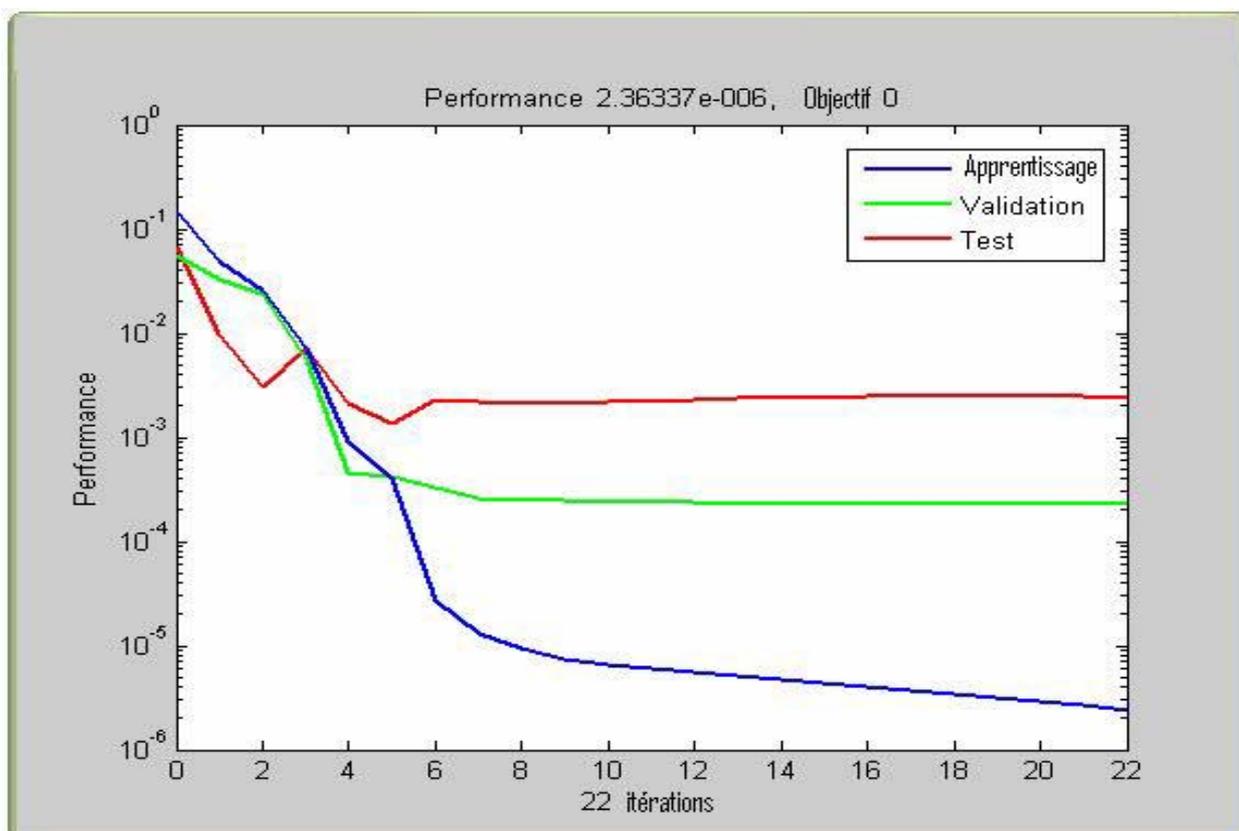


Figure V.34 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à neurones cachés, essai N°4.

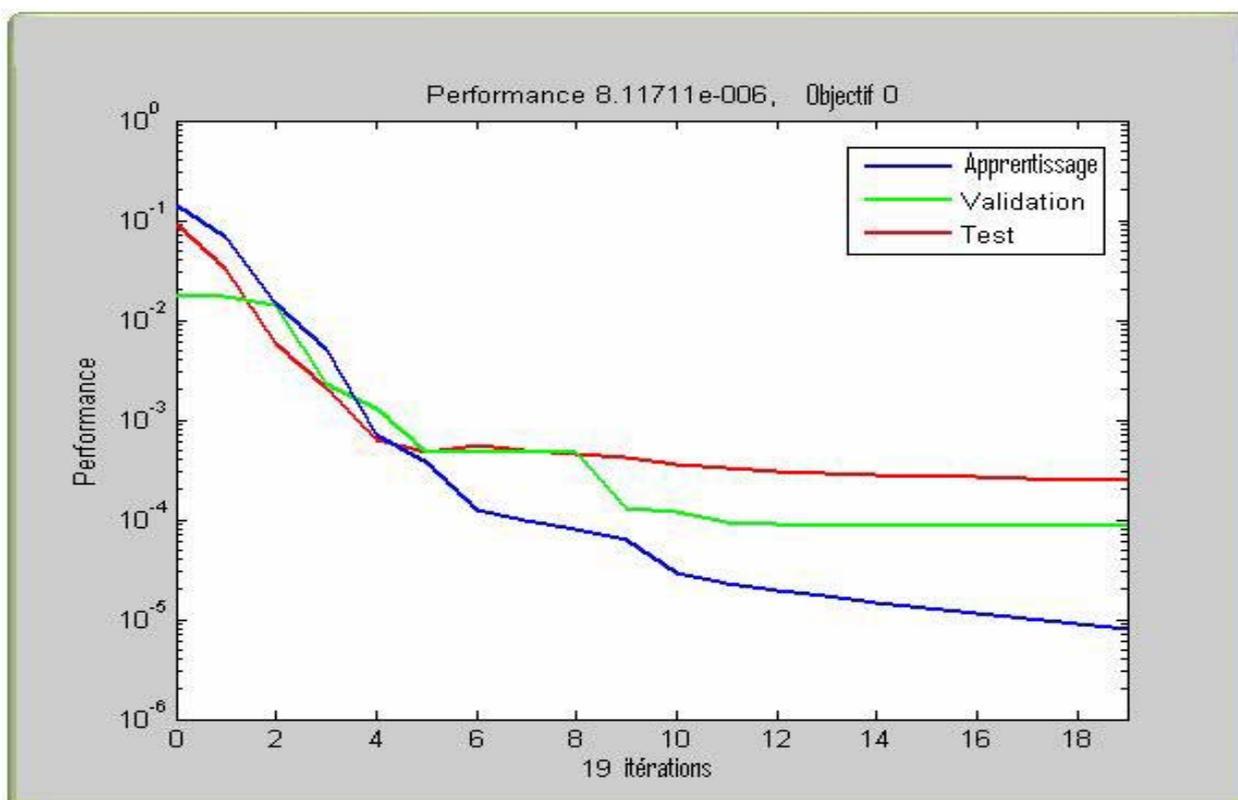


Figure V.35 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à sept neurones cachés, essai N°5.

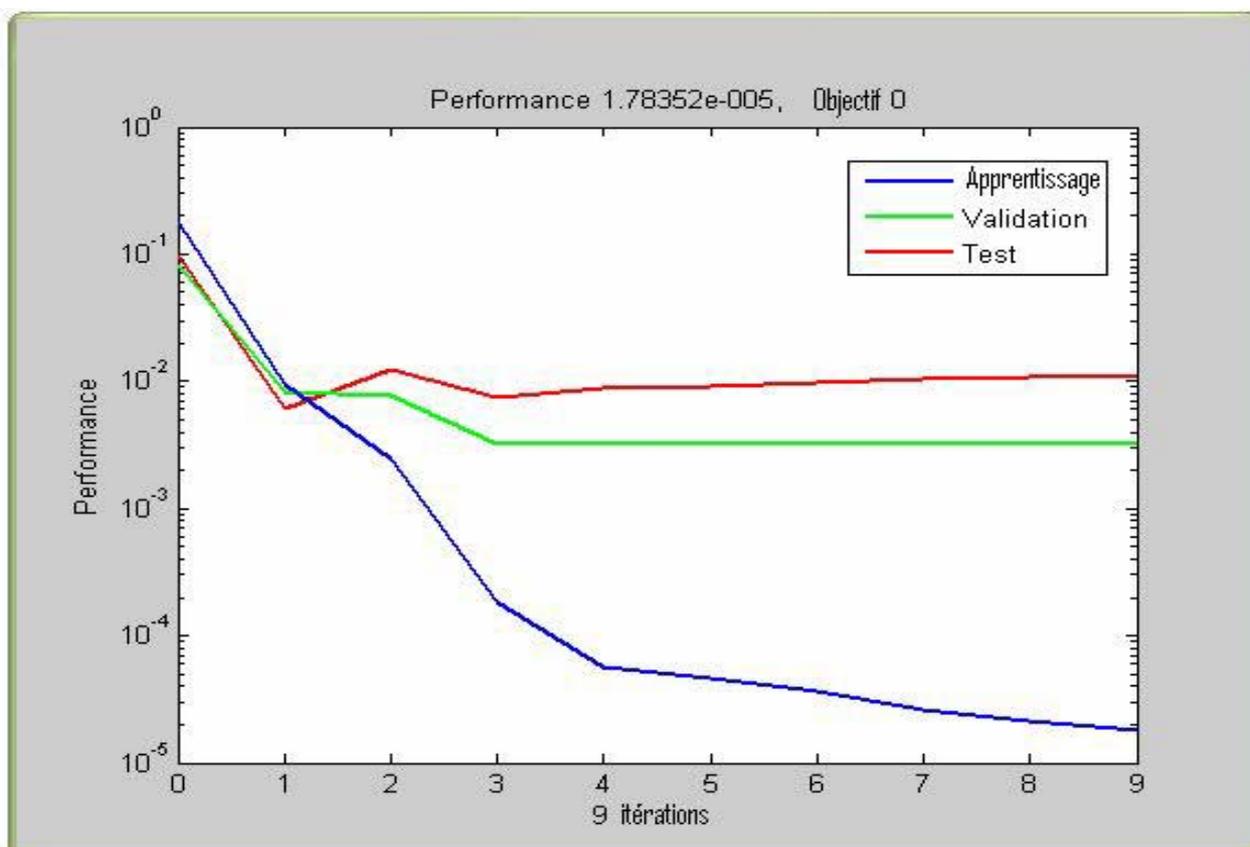


Figure V.36 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à huit neurones cachés, essai N°1.

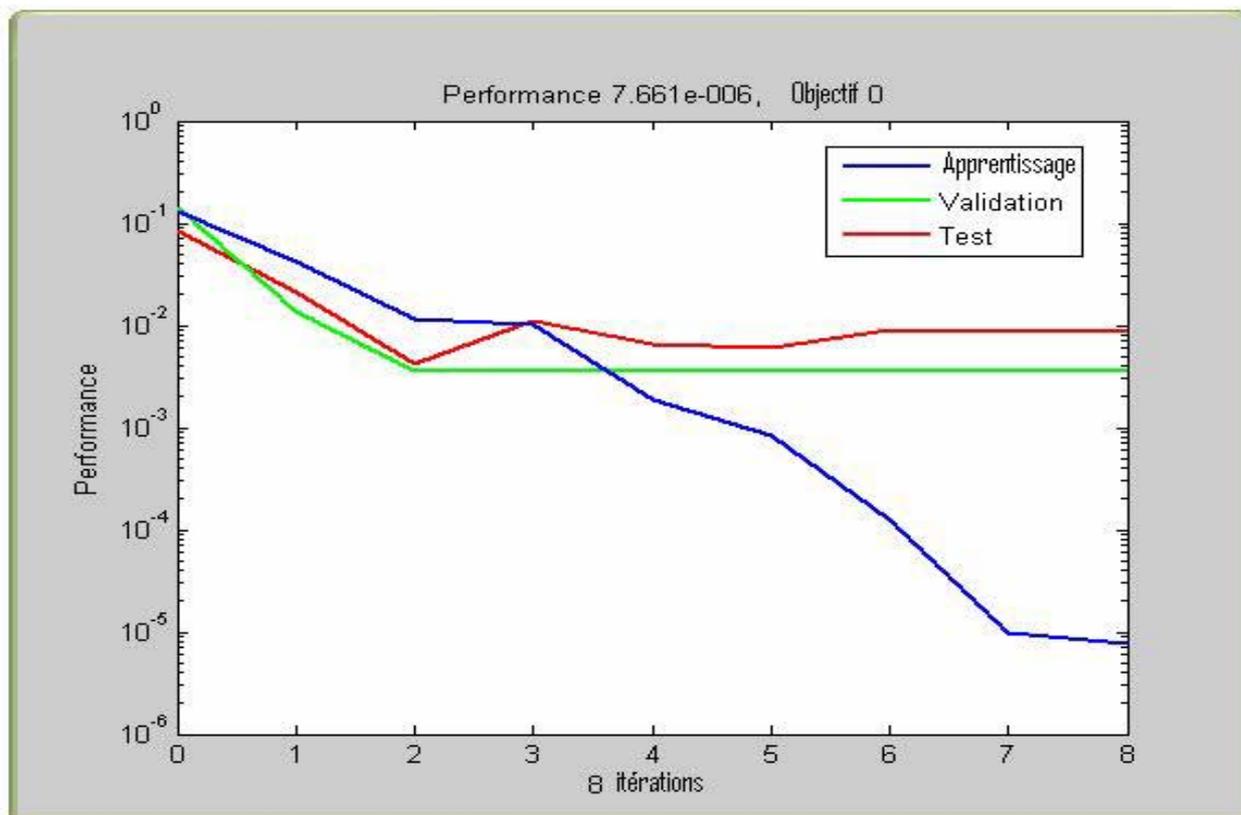


Figure V.37 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à huit neurones cachés, essai N°2.

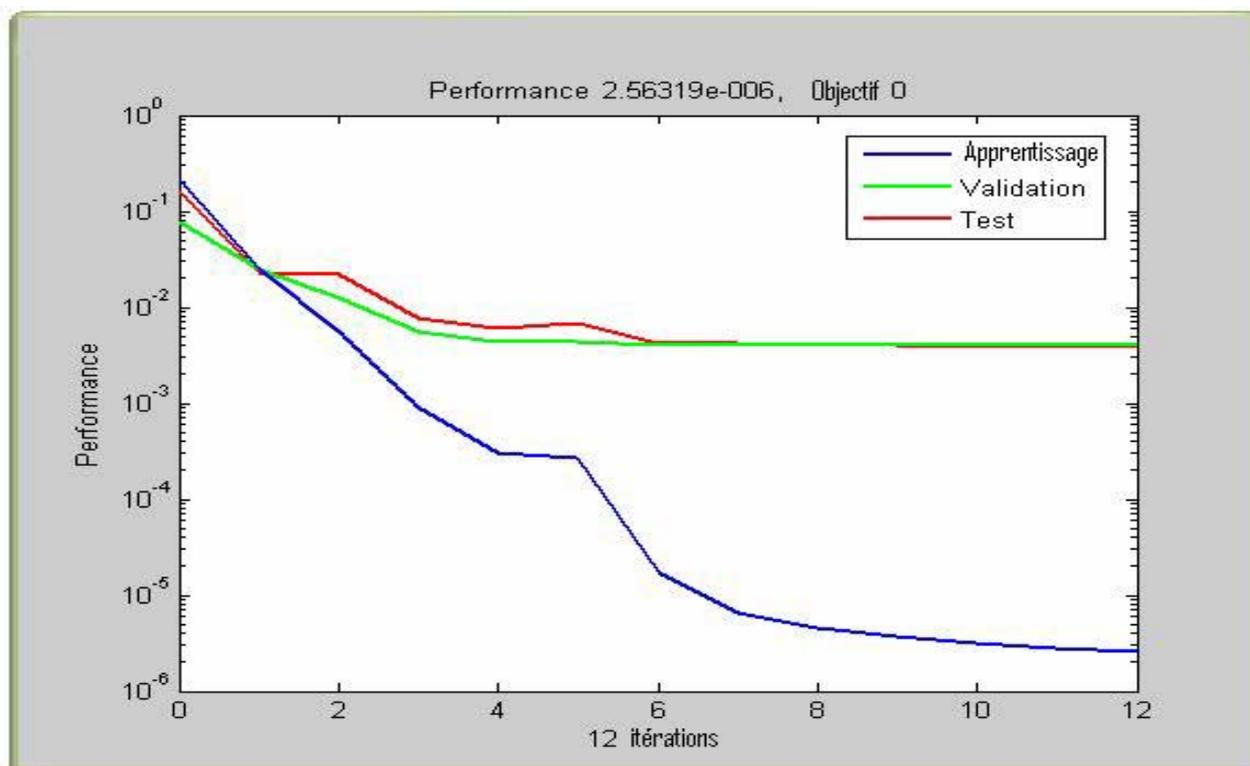


Figure V.38 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à huit neurones cachés, essai N°3.

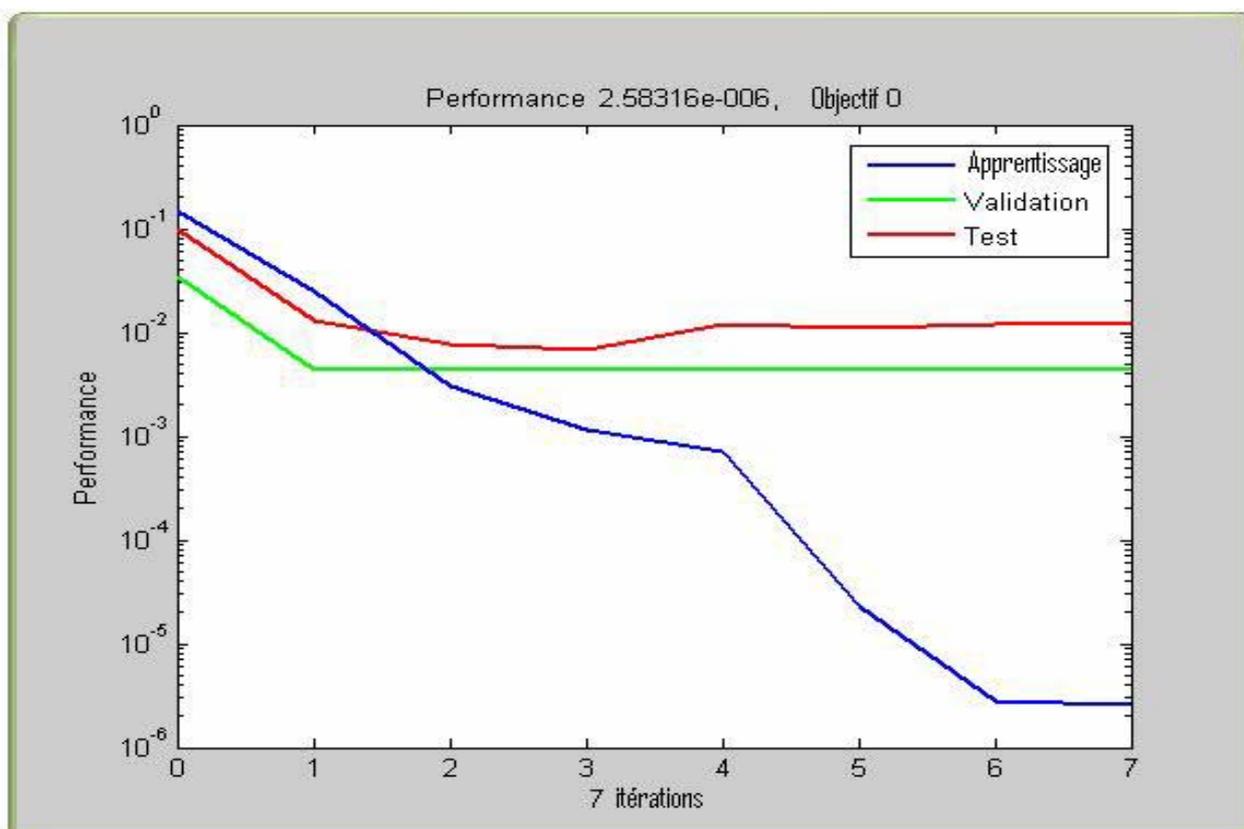


Figure V.39 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à huit neurones cachés, essai N°4.

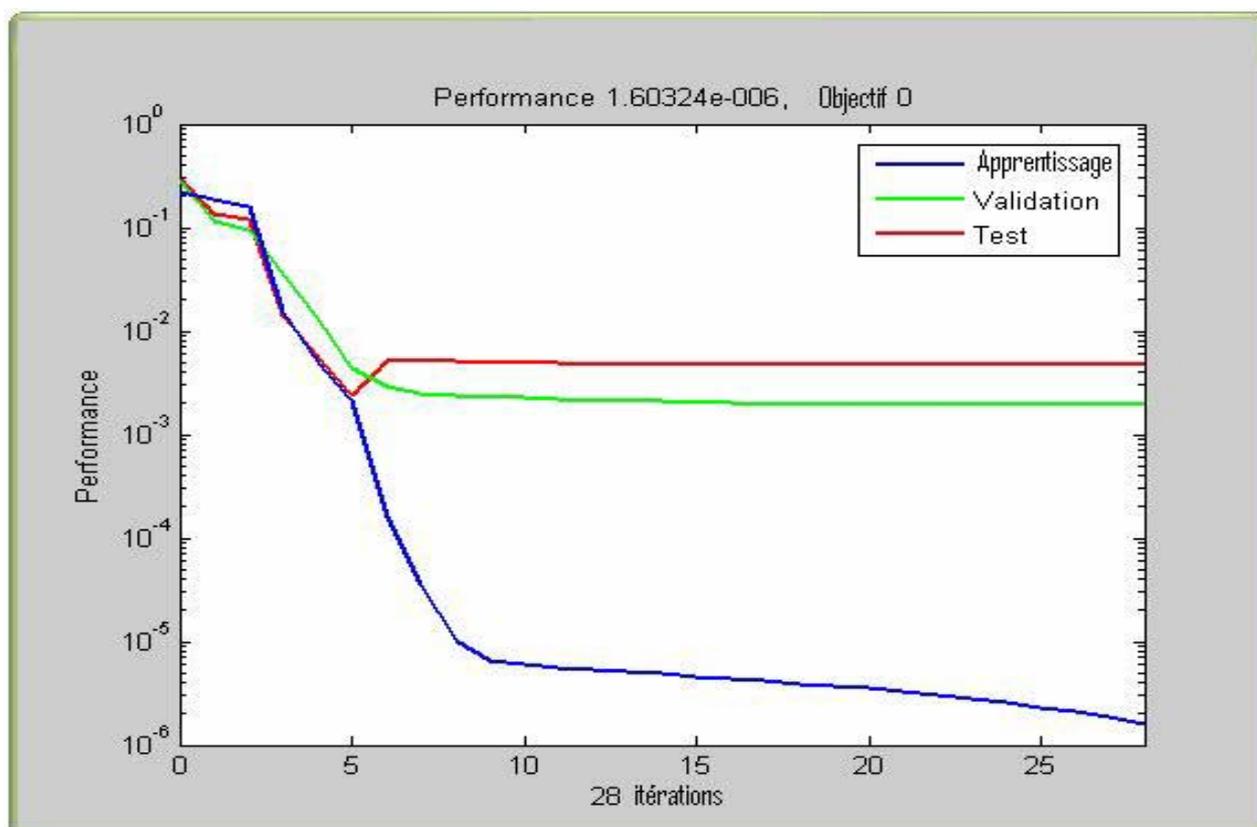


Figure V.40 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à huit neurones cachés, essai N°5.

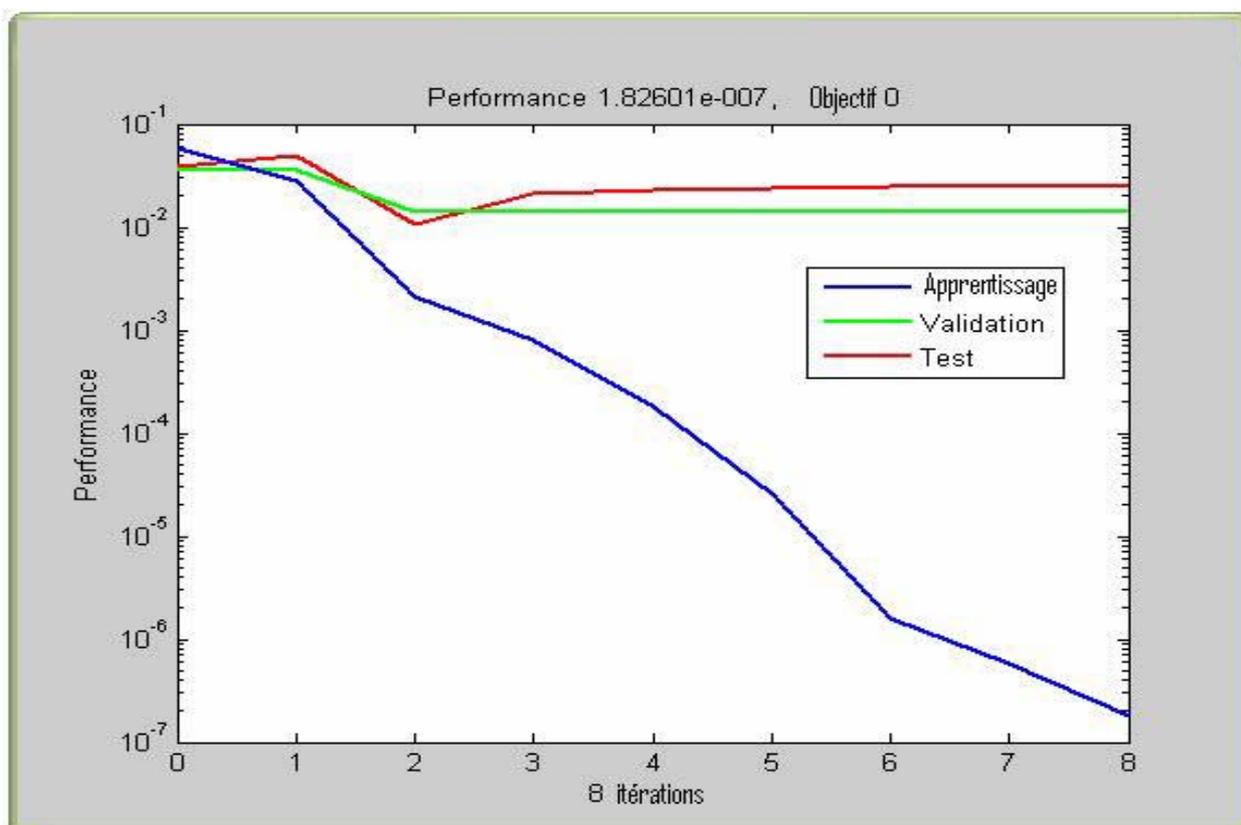


Figure V.41 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à neuf neurones cachés, essai N°1.

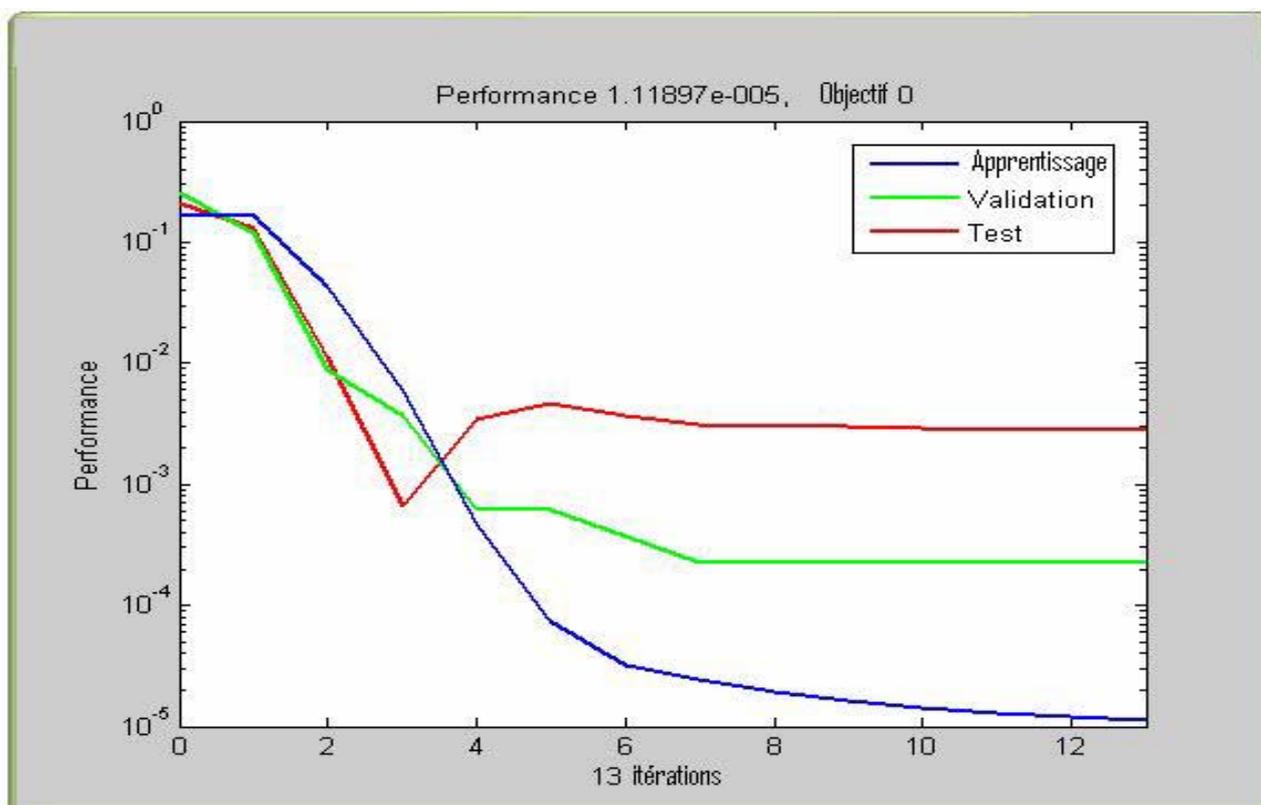


Figure V.42 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à neuf neurones cachés, essai N°2.

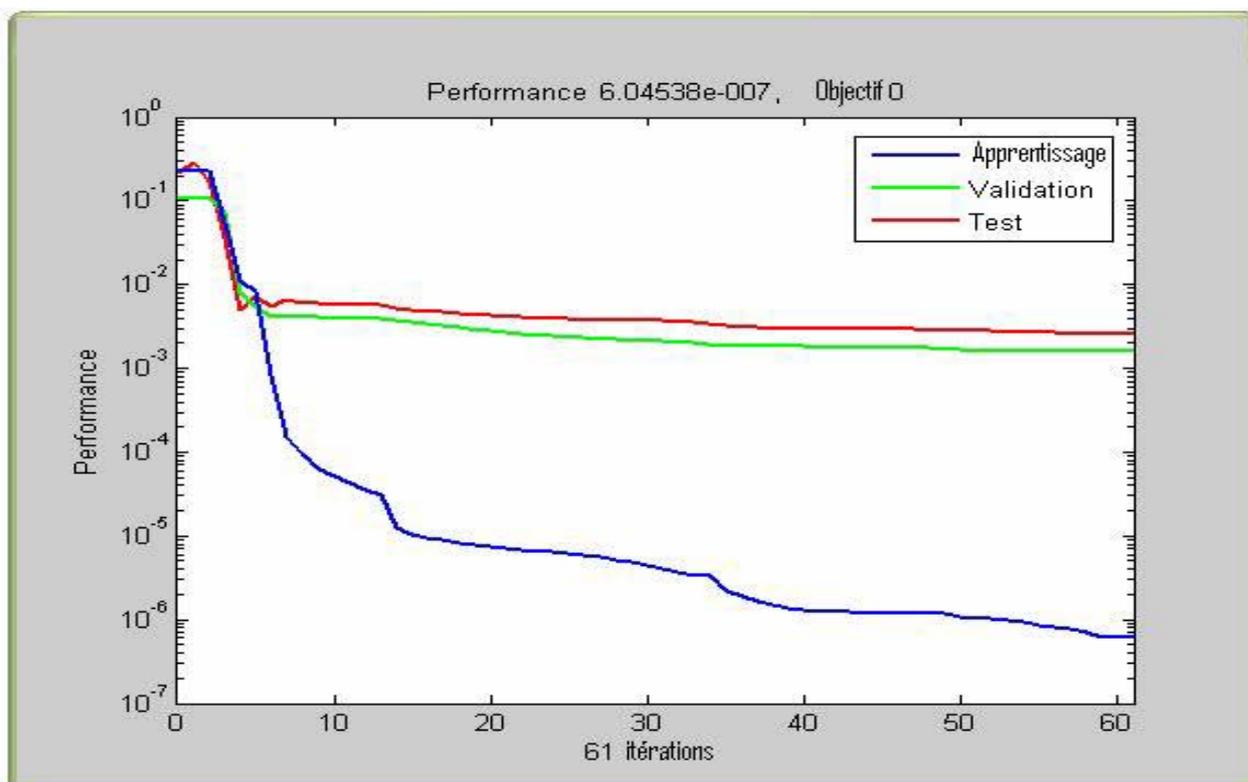


Figure V.43 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à neuf neurones cachés, essai N°3.

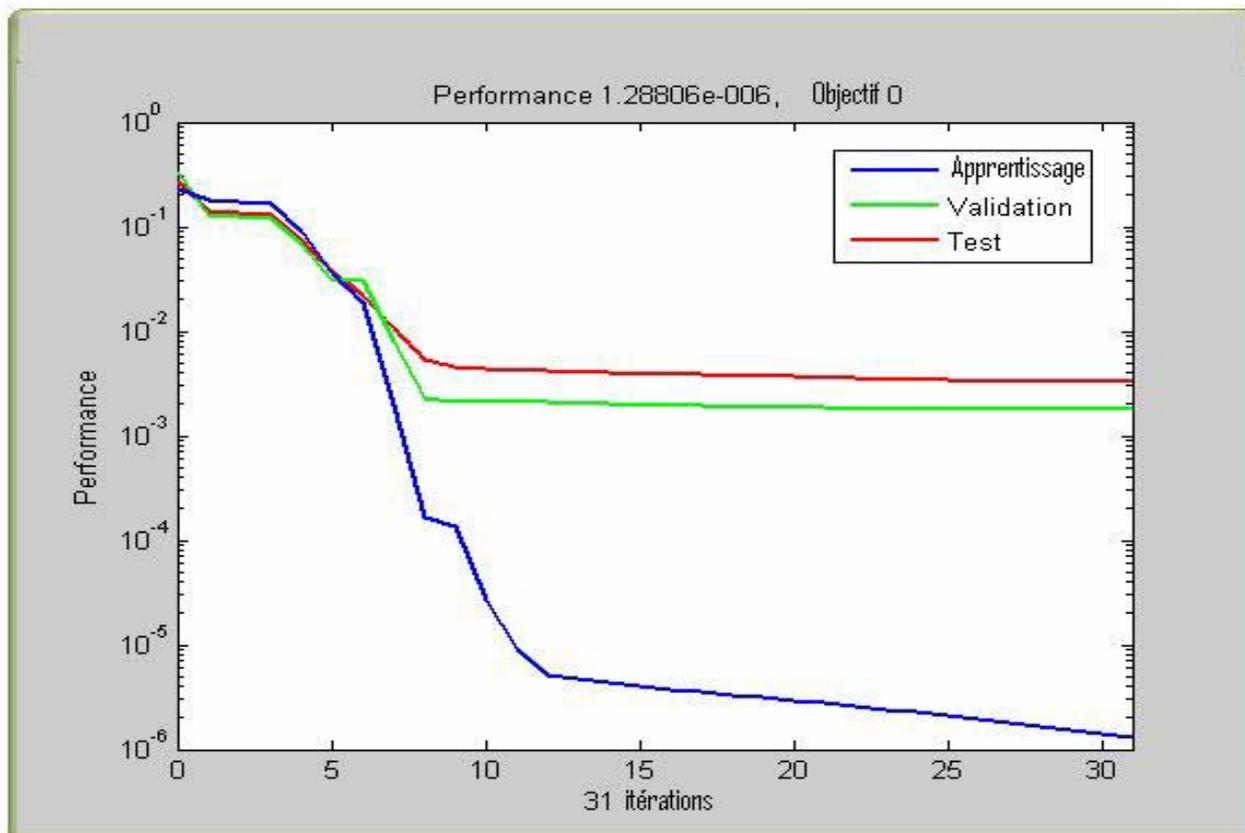


Figure V.44 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à neuf neurones cachés, essai N°4.

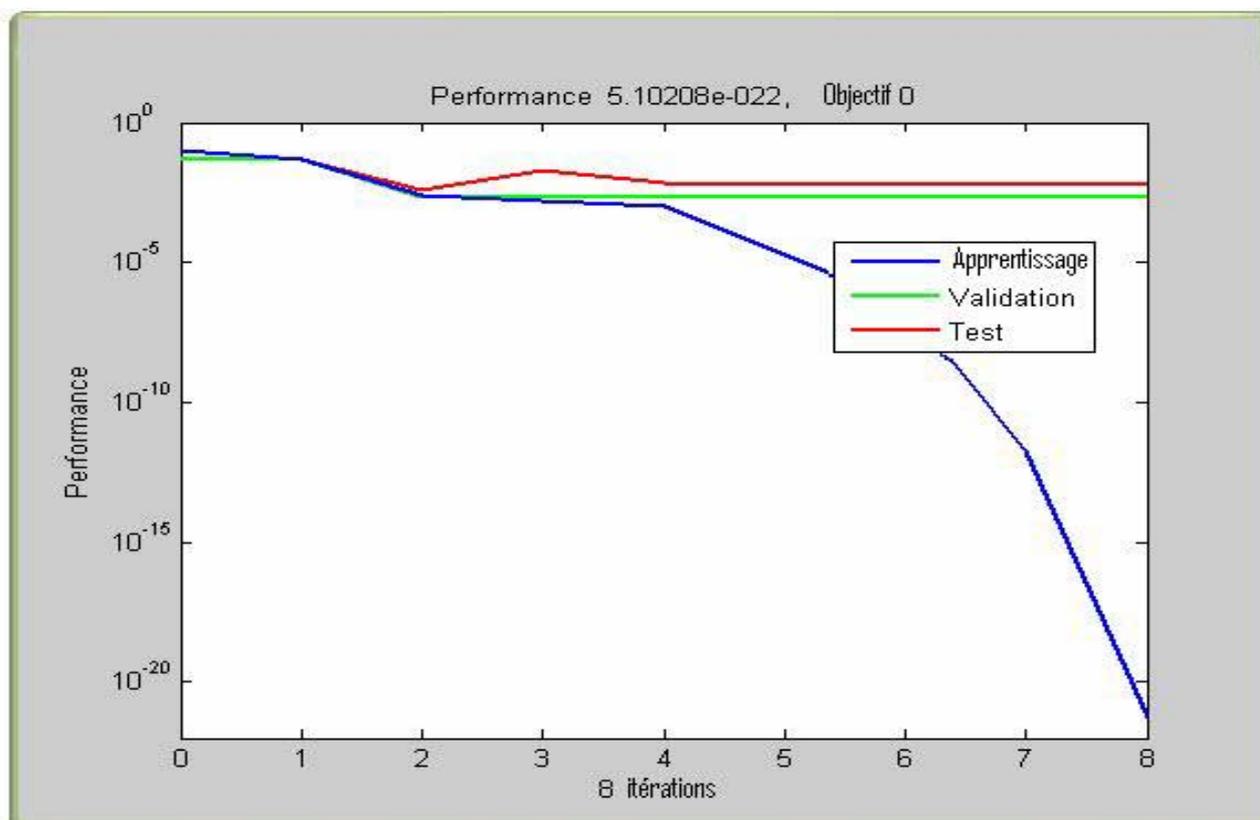


Figure V.45 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à neuf neurones cachés, essai N°5.

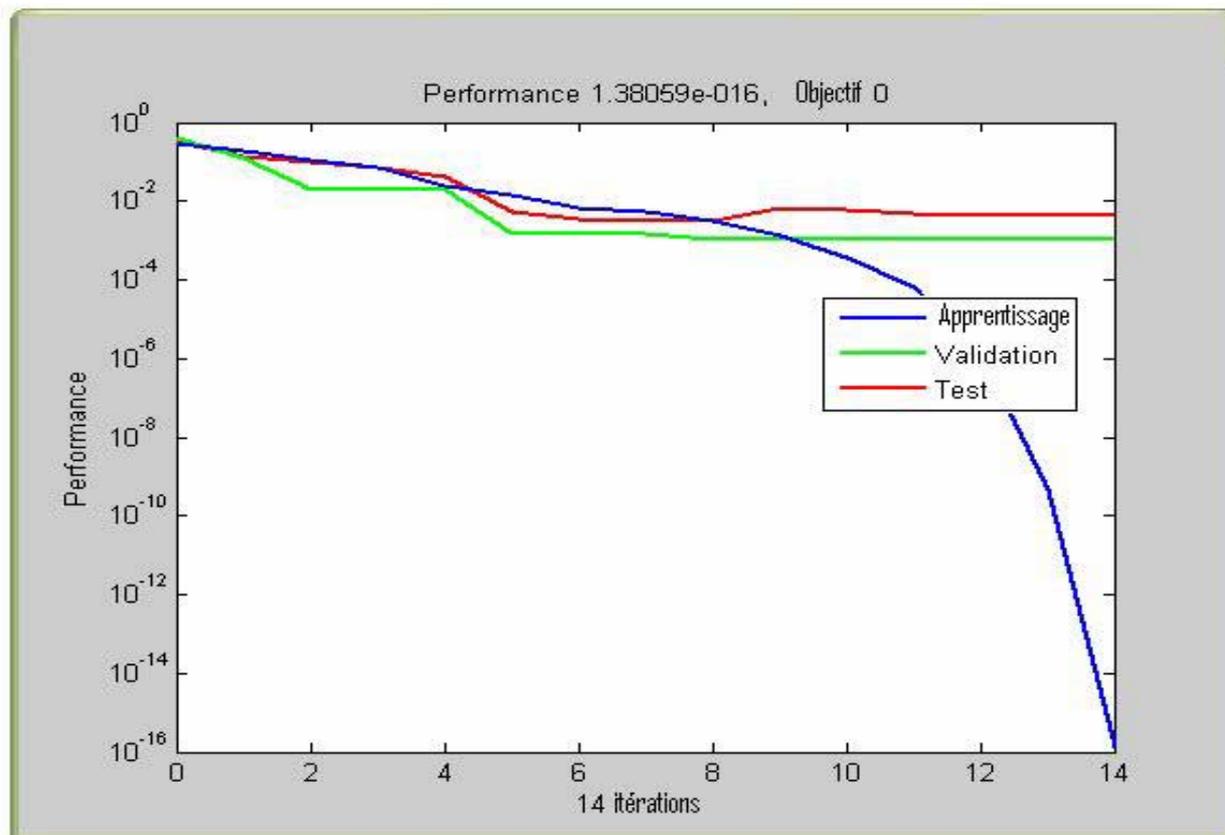


Figure V.46 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à dix neurones cachés, essai N°1.

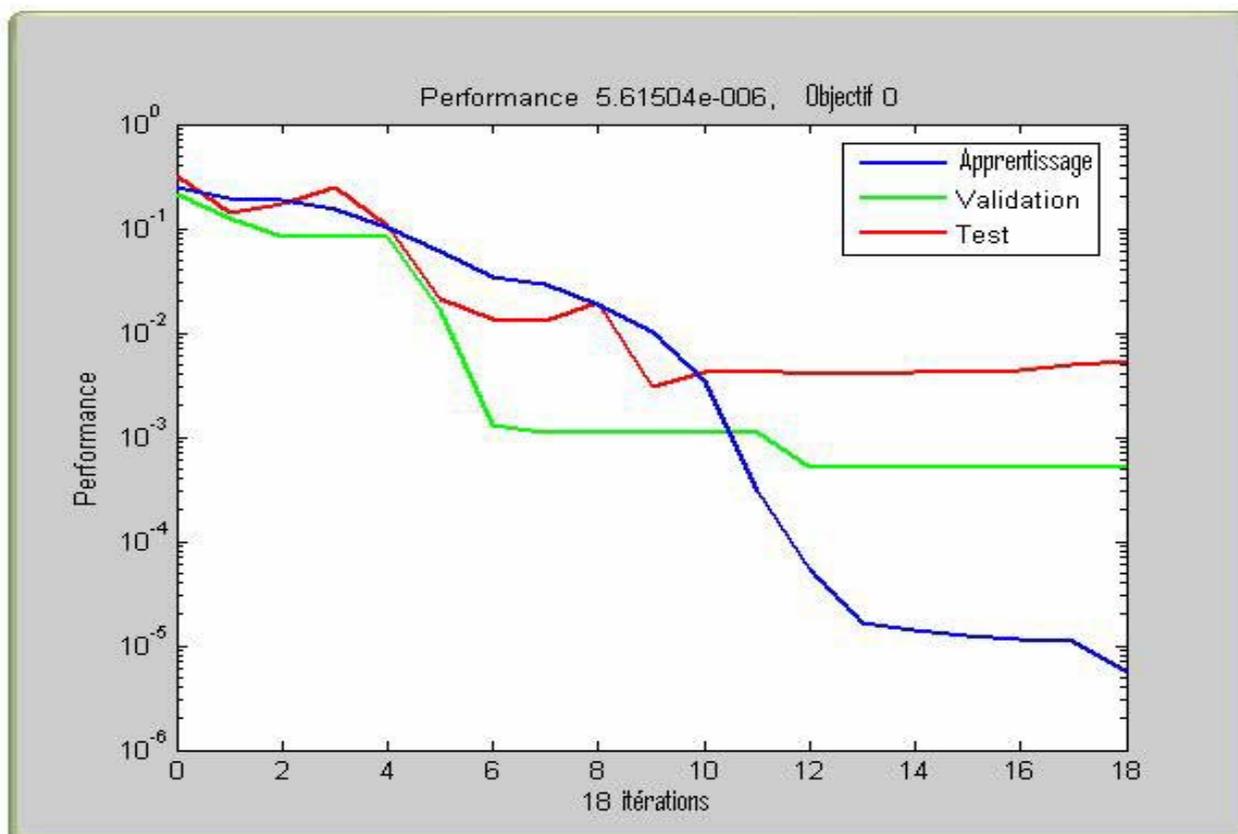


Figure V.47 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à dix neurones cachés, essai N°2.

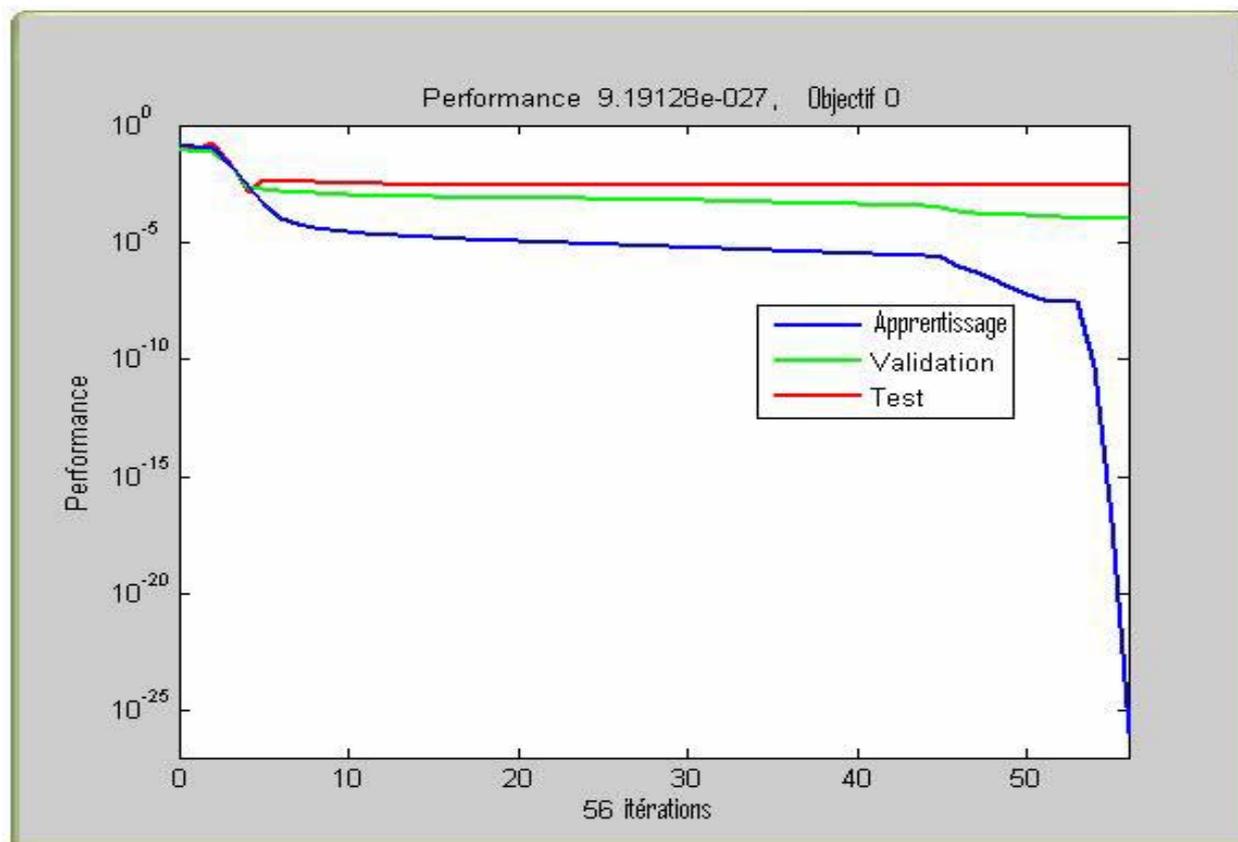


Figure V.48 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à dix neurones cachés, essai N°3.

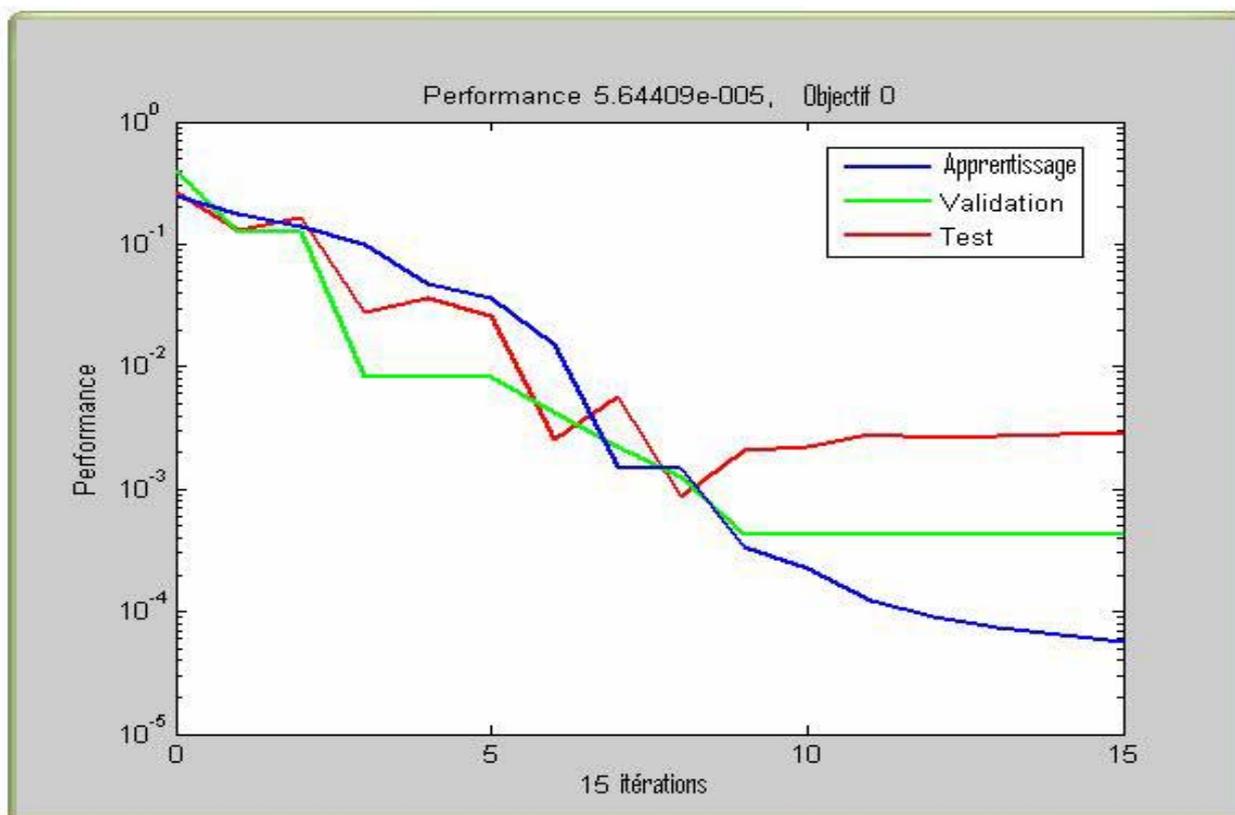


Figure V.49 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à dix neurones cachés, essai N°4.

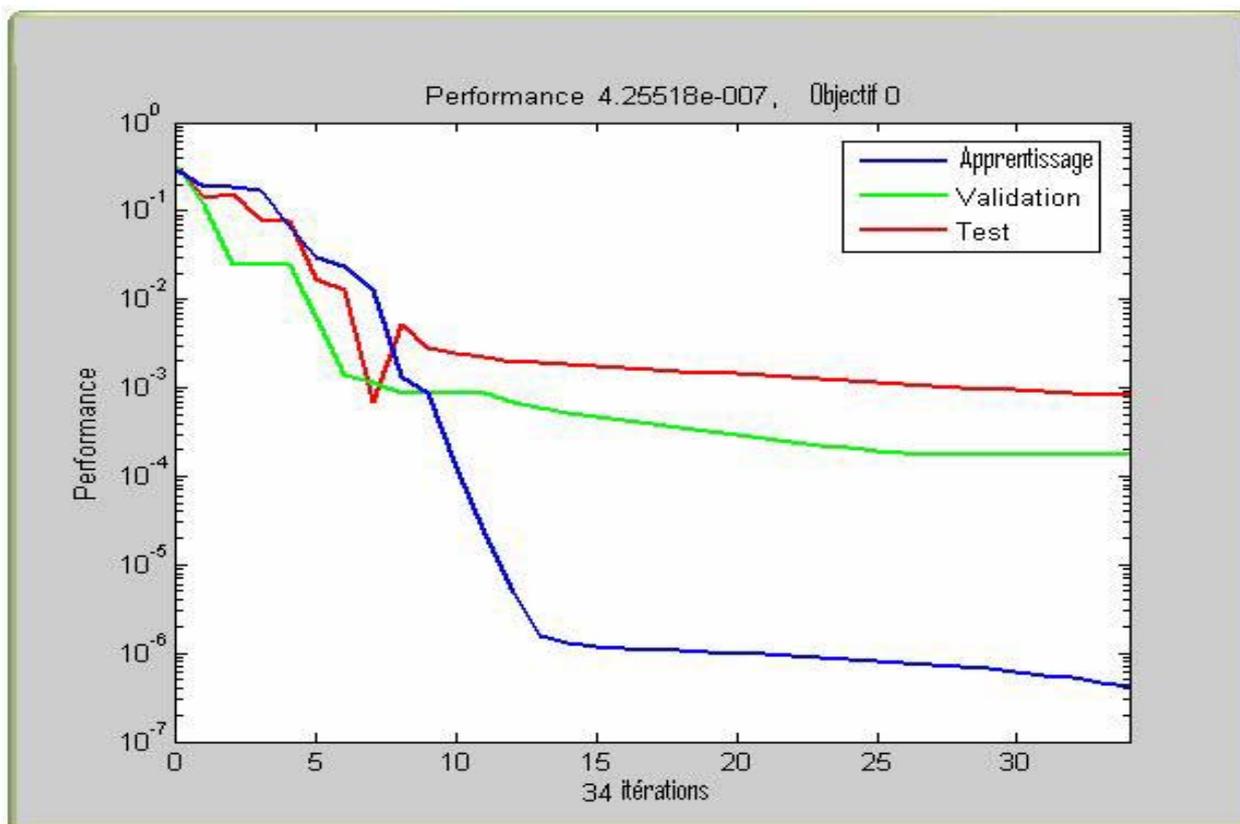


Figure V.50 : Evolutions de l'EQMA, de l'EQMV et de l'EQMT au cours du processus d'apprentissage du réseau à dix neurones cachés, essai N°5.

