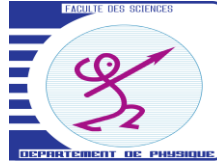




# UNIVERSITÉ D'ANTANANARIVO FACULTÉ DES SCIENCES



## FORMATION DOCTORALE EN PHYSIQUE



### DÉPARTEMENT DE PHYSIQUE

Laboratoire de Thermodynamique, Thermique et Combustion

## MEMOIRE

Pour l'obtention du :

### DIPLOME D'ETUDES APPROFONDIES EN PHYSIQUE

*Option* : Energétique

*Sur*

### OPTIMISATION D'ESTIMATION DU DEBIT D'UNE RIVIERE PAR LA THEORIE DES RESEAUX DE NEURONES ARTIFICIELS

*Présenté par :*

**ASSAANDI Saindou**

*Devant la commission d'examen composée de :*

<u>Président</u> :	Monsieur RAKOTOMALALA Minoson	Professeur titulaire
<u>Rapporteurs</u> :	Madame Randriamanantany Zely Arivelo	Professeur titulaire
<u>Examineurs</u> :	RANAIVO-NOMENJANAHARY Flavien Noël	Professeur titulaire
	RASAMIMANANA François de Salle	Maître de Conférences

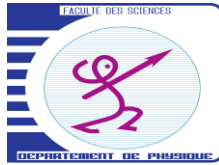
le 27 novembre 2010





# UNIVERSITÉ D'ANTANANARIVO FACULTÉ DES SCIENCES

FORMATION DOCTORALE EN PHYSIQUE



DÉPARTEMENT DE PHYSIQUE

Laboratoire de Thermodynamique, Thermique et Combustion

## MEMOIRE

Pour l'obtention du :

### DIPLOME D'ETUDES APPROFONDIES EN PHYSIQUE

*Option* : Energétique

*Sur*

**OPTIMISATION D'ESTIMATION DU DEBIT D'UNE RIVIERE PAR LA  
THEORIE DES RESEAUX DE NEURONES ARTIFICIELS**

*Présenté par :*

***ASSAANDI Saindou***



*Devant la commission d'examen composée de :*

<u>Président</u> :	Monsieur RAKOTOMALALA Minoson	Professeur titulaire
<u>Rapporteurs</u> :	Madame Randriamanantany Zely Arivelo	Professeur titulaire
<u>Examineurs</u> :	RANAIVO-NOMENJANAHARY Flavien Noël	Professeur titulaire
	RASAMIMANANA François de Salle	Maître de Conférences

le 27 novembre 2010



## REMERCIEMENTS

Je tiens à adresser mes sincères remerciements à Monsieur ANDRIANANTENAINA Bruno Jacques, Doyen de la Faculté des Sciences, de m'avoir permis de suivre la Formation Doctorale en Physique au sein de la Faculté des Sciences de l'Université d'Antananarivo et d'avoir autorisé la soutenance de ce mémoire.

Je tiens tout particulièrement à exprimer ma plus profonde gratitude à Monsieur RAKOTOMALALA Minoson, Professeur Titulaire, Directeur de l'Institut pour la Maîtrise de l'Energie, Responsable de l'option doctorale "ENERGETIQUE" de la Faculté des Sciences de m'avoir autorisé à poursuivre des études pour l'obtention du Diplôme d'Etudes Approfondies.

En toute sincérité et honnêteté, je remercie Madame RANDRIAMANANTANY Zely Arivelo, Professeur Titulaire, Chef du Département de Physique à la Faculté des Sciences de l'Université d'Antananarivo, qui m'a proposé ce sujet et m'a beaucoup aidé durant son accomplissement, malgré ses lourdes charges.

Je voudrais également exprimer mes vifs remerciements à Monsieur RANAIVONOMENJANAHARY Flavien Noël Professeur titulaire, d'avoir accepté d'examiner ce travail en tant que membre du jury.

J'exprime mes remerciements chaleureux à Monsieur RASAMIMANANA François de Salle, Maître de conférences à la Faculté des Sciences de l'Université d'Antananarivo, d'avoir bien voulu siéger en tant que membre du jury pour juger les résultats de mes travaux.

J'adresse aussi mes remerciements à Messieurs ANDRIAMAHITASOA Bernard Andriamparany et RAVELOMPANAHY Donat Hervé pour leur participation techniquement à la réalisation de cette mémoire.

Je voudrais également exprimer mes grands remerciements à tous les professeurs au sein de l'Institut pour la Maîtrise de l'Energie qui m'ont aidé par leurs enseignements pendant toutes ces années d'étude. Grâce à eux, j'ai été capable de finaliser ce présent mémoire de DEA.

J'adresse à ma famille, et spécialement à mes parents, toute mon affection en retour de leur aide, de leur prière et de leur confiance en moi. Puisse ce travail leur être la preuve de mon dévouement total à leur endroit.

Enfin, nous tenons à exprimer nos vifs remerciements à tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce mémoire j'adresse mes salutations les plus amicales

## TABLEAU DE MATIERE

REMERCIEMENTS.....	i
TABLEAU DE MATIERE .....	ii
LISTE DES ACRONYMES .....	viii
LISTE DES FIGURES .....	x
LISTE DES TABLEAUX.....	xi
Introduction générale.....	1
CHAPITRE 1 : .....	3
NOTIONS GENERALES SUR LA CLIMATOLOGIE, L’HYDROLOGIE ET LES MICROCENTRALES HYDROELECTRIQUES A MADAGASCAR .....	3
1.1. Situation géographique de Madagascar .....	4
1.1.1. Caractéristiques générales de Madagascar.....	5
1.1.2. Hautes Terres centrales .....	5
1.1.3. Versant oriental .....	5
1.2. Ressources naturelles.....	6
1.2.1. Climat.....	6
1.2.2 Saisons .....	6
1.2.3. Pluviométrie .....	6
1.3. Le potentiel hydroélectrique de Madagascar .....	7
1.3.1. Régimes hydrologiques.....	7
1.3.2. Sites hydroélectriques identifiés .....	8
1.4. Situation socio économique de l’électrification rurale .....	9
1.4.1. Contexte de la demande d’électricité à Madagascar .....	9
1.4.1.1. Consommation d’énergie selon les ressources.....	9
1.4.1.2. Consommation d’énergie selon les régions.....	10
1.4.2. Cadre de l’électrification rurale .....	11

1.4.2.1. Législation relative à l'électrification rurale .....	11
1.4.2.2. Politique nationale du développement de l'électrification rurale.....	11
1.4.2.2.1. Objectif .....	11
1.4.2.2.2. Stratégie : libéralisation du secteur électrique.....	12
1.4.3. Acteurs de l'électricité rurale.....	12
1.5. Microcentrales hydroélectriques.....	14
1.5.1. Définition et généralités .....	14
1.5.2 Fonctionnement des centrales hydrauliques .....	14
1.5.2.1. Principe de fonctionnement d'une centrale hydroélectrique.....	14
1.5.2.2. Puissance d'une chute d'eau .....	15
1.6. Les différentes composantes d'une microcentrale hydraulique.....	16
1.6.1. Ouvrages de génie civil.....	16
1.6.2. Eléments hydrauliques et électromécaniques .....	18
1.6.3. Eléments électriques .....	19
CHAPITRE 2 : .....	20
THEORIE DES RESEAUX DE NEURONES ARTIFICIELS .....	20
2.1. Introduction .....	21
2.2. Retour sur l'approximation de fonction par des polynômes.....	21
2.2.1. Cas d'une seule variable .....	21
2.2.2. Cas de plusieurs variables .....	22
2.3. Réseaux de neurones artificiels .....	22
2.3.1 Historique.....	22
2.3.1.1 Neurone formel .....	23
2.4. Structure des réseaux de neurones.....	25
2.3.1 Différents types de fonctions de transfert .....	27
2.4.2 Architectures des réseaux de neurones .....	29
2.4.2.1 Perceptron.....	29
2.4.2.2 Perceptrons multicouches (PMC).....	29

---

2.4.2.3 Réseau de Hopfield : .....	30
2.4.3 Classification des réseaux de neurones .....	31
2.4.3.1 Réseaux bouclé ou dynamique .....	31
2.4.3.2 Réseau non bouclé.....	32
2.4.4. Perceptrons Multicouches .....	33
2.4.4.1. Mise en œuvre .....	33
2.4.4.2. Approximation universelle .....	33
2.4.4.3. La propriété de parcimonie .....	34
2.5. Phase d'apprentissage.....	34
2.5.1. L'apprentissage non supervisé.....	35
2.5.2. Le mode supervisé .....	35
2.5.3 Algorithmes de minimisation.....	36
2.6. Phase de généralisation.....	40
2.7. Le problème de surapprentissage .....	41
2.7.1. Définition de surapprentissage.....	41
2.7.2. Biais et variance .....	41
2.7.2.1. Early stopping .....	42
2.7.2.2. <i>Weight decay</i> .....	43
2.8. Approche neuronale bayésienne .....	43
2.8.1. Théorème de Bayes [19] .....	43
2.8.2. Approche probabiliste appliquée à l'apprentissage neuronal [25].....	44
2.8.3. Distribution de probabilité a priori [21] [21] .....	44
2.8.4. Fonction de vraisemblance [26] [27] .....	45
2.8.5. Distribution de probabilité a posteriori [26] [27].....	46
2.8.6. Approximation gaussienne du posterior [26] [27] .....	46
2.8.7. Détermination des hyper paramètres $\alpha$ et $\beta$ [26] [27] .....	47
2.8.8 Critères de performances pour les modèles .....	49

---

CHAPITRE 3 : .....	50
MODELISATION DU COMPORTEMENT DU DEBIT A PARTIR DES RESEAUX DE NEURONES ARTIFICIELS.....	50
3.1. Bases de données utilisées .....	51
3.2. Réseaux de neurones statiques .....	52
3.2. 1. Présentation du modèle neuronal classique .....	52
3.2.2. Organigramme du réseau classique.....	52
3.2.3. Résultats avec le modèle classique .....	53
3.2.3.1. La station d'Anosizato .....	54
3.2.3.2. La station d'Anjeva .....	56
3.2.3.3. La station d'Ambohimanambola.....	58
3.3. Modélisation avec l'approche bayésienne .....	59
3.3.1. Organigramme de l'apprentissage bayésien .....	59
3.3.2. Résultats avec l'approche bayésienne.....	61
3.3.2. La station d'Anosizato .....	61
3.3.4. La station d'Anjeva.....	63
3.3.4. La station d'Ambohimanambola.....	65
3.2.4.3 Recherche des modèles parcimonieux .....	66
3.2.4.3.1. Station d'Anjeva .....	66
3.2.4.3.1. Station d'Ambohimanambola.....	67
3.2.5.1. Comparaison des performances des modèle bayésien et optimisé.....	69
CONCLUSION GENERALE .....	71
REFERENCES	
LISTE DES ANNEXES	





## NOMENCLATURE

$P_H$	la puissance hydraulique	kw
$Q$ ,	le débit d'eau	$m^3/s$
$H$	la hauteur de chute	kw
$W$	l'énergie électrique produite	kwh
$P_{inst}$	la puissance installée	w
$t$ ,	la durée de fonctionnement	[h]
$f_1$	le coefficient lié aux variations saisonnières de débit pour des installations au fil de l'eau sans réservoir régulateur.	
S	somme pondérée	
d	sortie désirée	
Nc	nombre de neurones nuerons caches	
$\alpha_k$	le pas de descente	
xi	entrées	
$y(x)$	la sortie du modèle	
$J(w)$	fonction de cout	
$w_0$	Biais	
$w_i$	Poids d'une connexion	
$p(\theta/D)$	fonction de ressemblance	
$p(D)$	l'évidence	
$p(\theta)$	priori	
$y(x)$	la sortie du modèle	
$E_w$	terme régulateur	
$p$	le nombre de données	
$\alpha$	hyper paramètre	
$E_D$ ,	la fonction d'erreur	
$\beta$ ,	un autre hyper paramètre lié à la sortie.	
$S(w)$	fonction régularisée	
$\alpha_{MP}$ , $\beta_{MP}$	correspondent au maximum de ce terme <i>Evidence</i> .	
A :	sortie prédite	
T :	sortie souhaité	

**LETTRE GRECQUE**

$\eta$	le rendement de l'ensemble turbine-générateur variant	[-]
$\eta_{tu}$	le rendement de la chute	[-]
$\eta_{rd}$	le rendement du multiplicateur	[-]
$\eta_{al}$	le rendement du transformateur	[-]
$\eta_{tr}$	le rendement du transformateur	[-]
$\eta_{li}$	le rendement de la ligne.	[-]
$\alpha$	hyper paramètre	
$\beta$ ,	un autre hyper paramètre lié à la sortie.	

## LISTE DES ACRONYMES

<b>ADER</b>	L'Agence de <b>D</b> éveloppement de l' <b>E</b> lectrification <b>R</b> urale
<b>APIPA</b>	L' <b>A</b> utorité pour la <b>P</b> rotection contre les <b>I</b> nondations de la <b>P</b> lain d'Antananarivo
<b>fdp</b>	fonction de <b>d</b> ensité de <b>p</b> robabilité
<b>FNE</b>	<b>F</b> ond <b>N</b> ational de l' <b>E</b> lectricité
<b>JIRAMA</b>	<b>J</b> iro sy <b>R</b> ano <b>M</b> alagasy
<b>NRA</b>	<b>R</b> éseaux de <b>N</b> eurons <b>A</b> rtificiels
<b>MCH</b>	<b>M</b> icro <b>C</b> entrale <b>H</b> ydroélectrique
<b>MECIE</b>	<b>M</b> ise en <b>C</b> ompatibilité des <b>I</b> nvestissements avec l' <b>E</b> nvironnement
<b>MEM</b>	<b>M</b> inistère de l' <b>E</b> nergie et des <b>M</b> ines
<b>MLP</b>	<b>M</b> ulti <b>L</b> ayer <b>P</b> erceptron
<b>ORE</b>	L' <b>O</b> ffice de <b>R</b> égulation de l' <b>E</b> lectricité
<b>PDP</b>	<b>P</b> arallel <b>D</b> istributed <b>P</b> rocessing
<b>PMC</b>	<b>P</b> erceptrons <b>M</b> ulticouches
<b>UNPEDE</b>	L' <b>U</b> nion <b>I</b> nternationale des <b>P</b> roducteurs et <b>D</b> istributeurs d' <b>E</b> nergie <b>E</b> lectrique

## LISTE DES FIGURES

Figure 1. 1: Carte physique de Madagascar .....	4
Figure 1. 2: Sites ayant des potentialités hydroélectriques déjà identifiées. ....	8
Figure 1. 3: Principe de fonctionnement d'une centrale hydroélectrique .....	15
Figure 1. 4: illustration des composantes d'une centrale hydroélectrique .....	16
Figure 2. 1: Illustration du surapprentissage. ....	22
Figure 2. 2: Neurone biologique et cerveau humain. ....	23
Figure 2. 3: La ressemblance entre un neurone biologique et un neurone artificiel .....	24
Figure 2. 4: Différents types de neurones .....	25
Figure 2. 5: Structure d'un neurone artificiel.....	26
Figure 2. 6: La fonction « hardlim » .....	27
Figure 2. 7: La fonction « linéaire ».....	27
Figure 2. 8: La fonction « log-sigmoïd » .....	28
Figure 2. 9: La fonction « tansig » .....	28
Figure 2. 10: Réseau multicouche .....	30
Figure 2. 11: Réseau de Hopfield totalement connecté.....	30
Figure 2. 12: Forme canonique d'un réseau de neurones bouclé. ....	31
Figure 2. 13: Schéma synoptique d'un réseau de neurone non bouclé .....	32
Figure 2. 14: Réseau PMC à n entrées, Nc neurones cachés et un neurone de sortie .....	32
Figure 2. 15: Mode non supervisé .....	35
Figure 2. 16: Mode supervisé .....	36
Figure 2. 17: Principe de la descente du gradient .....	37
Figure 2. 18: Illustration du compromis biais/variance.....	41
Figure 2. 19: Early stopping .....	42
Figure 3. 1: service d'annonce de crues dans le bassin de l'Ikopa en amont de Bevomanga (4300 km <sup>2</sup> ) .....	51
Figure 3. 2: organigramme du réseau classique .....	53

---

Figure 3. 3: MLP en phase d'apprentissage .....	54
Figure 3. 4: droite de régression MLP en apprentissage .....	55
Figure 3. 5: MLP en phase de test.....	55
Figure 3. 6: Droite de régression MLP en phase de test.....	56
Figure 3. 7: MLP en phase d'apprentissage .....	56
Figure 3. 8: droite de régression MLP en apprentissage .....	57
Figure 3. 9: MLP en phase de test.....	57
Figure 3. 10: Droite de régression MLP en phase de test.....	58
Figure 3. 11: MLP en phase d'apprentissage .....	58
Figure 3. 12: MLP en phase de test.....	59
Figure 3. 13: Organigramme suivi lors de l'apprentissage bayésien. ....	60
Figure 3. 14: Droite de régression approche bayésienne en apprentissage .....	62
Figure 3. 15: Approche bayésienne en phase test .....	62
Figure 3. 16: Droite de régression approche bayésienne en phase test .....	63
Figure 3. 17: Approche bayésienne en phase d'apprentissage.....	63
Figure 3. 18: Droite de régression approche bayésienne en apprentissage .....	64
Figure 3. 19: Approche bayésienne en phase test .....	64
Figure 3. 20: Droite de régression approche bayésienne en test .....	65
Figure 3. 21: Approche bayésienne en phase d'apprentissage.....	65
Figure 3. 22: Approche bayésienne en phase test .....	66
Figure 3. 23: la détermination des hyper-parametres .....	67
Figure 3. 24: Nombre de neurones cachés .....	67
Figure 3. 25: la détermination des hyper-parametres.....	68
Figure 3. 26: Evidence en fonction du nombre de neurones cachés. ....	68
Figure 3. 27: Approche bayésienne optimisé en phase test.....	69
Figure 3. 28: Droite de régression approche bayésienne optimisée en apprentissage .....	70

## LISTE DES TABLEAUX

Tableau 1. 1: Consommation d'énergie à Madagascar en l'an 2000 .....	9
Tableau 1. 2: Sources d'énergie à Madagascar .....	10
Tableau 3. 1: les entrées du modèle .....	53
Tableau 3. 2: Evaluation des critères de performance pour le modèle classique.....	54
Tableau 3. 3: Evaluation des critères de performance pour le modèle bayesienne.....	61
Tableau 3. 4: Comparaison des erreurs des modèles.....	69

## **INTRODUCTION GENERALE**

Dans le monde actuel, les énergies renouvelables prennent une place de plus en plus importante dans notre société. L'exploitation du potentiel hydraulique est une des solutions envisageables pour résoudre le problème énergétique auquel fait face actuellement Madagascar.

En effet, classée dans la catégorie des énergies renouvelables, l'énergie hydroélectrique est la plus propre et la moins chère (en terme prix de revient) comparée à toute autre forme de production d'énergie électrique.

L'énergie constitue une des facteurs de la croissance économique et de l'augmentation des biens de consommation à travers le monde.

Madagascar possède d'importantes ressources hydroélectriques, estimées à 7500MW, alors que seulement 24% de la population ont accès à l'électricité et dont 4% se trouvent en milieu rural.

La plupart des villages malgaches, qui sont composés seulement de quelques foyers, sont susceptibles d'être électrifiés par des ressources renouvelables isolées. Plus particulièrement, l'implantation des MicroCentrales Hydroélectriques (MCH) est parmi ces solutions proposées. Mais leurs installations sont confrontées à un certain nombre de problèmes dont celui le plus crucial est le manque de données hydrologiques pour les petites rivières de Madagascar. Notre choix porte sur l'utilisation des Réseaux de Neurones Artificiels (RNA) pour pouvoir y accéder.

A la fin de l'année 1997, une trentaine d'articles portant sur l'application des RNA à la gestion des ressources en eau ont été publiés. Une des premières applications fut celle de la prévision de la demande en eau de la ville de Canberra en Australie [1]. Ensuite, d'autres applications à la prévision de la consommation d'eau ont mené à des résultats jugés satisfaisants [2] [3]. Par la suite, la moitié des applications hydrologiques concernent les prévisions des débits, crues et étiages des rivières. La prédiction de la qualité de l'eau représente environ 20% des applications des RNA en hydrologie.

Maier et Dandy (1996) ont utilisé un réseau Perceptron MultiCouche (PMC) pour la prévision de deux (2) semaines en avance du degré de salinité de l'eau de la rivière Murray en Australie [4] et seulement 10% des applications recensées concernent la prévision des apports naturels dans les réservoirs hydroélectrique [5].

Dans tous les travaux entrepris, les bases de données étaient complètes. A Madagascar, le nombre de stations d'acquisition de données est très insuffisant, d'où notre objectif qui

consiste, entre autres, à exploiter les capacités des RNA à modéliser les régimes annuels et interannuels des cours d'eau en comblant le vide créé par le manque de données et ceci, lors des phases d'étude et d'exploitation d'une MCH.

Cet ouvrage est divisé en trois chapitres, en plus des introduction et conclusion générales :

- Le premier chapitre illustre les notions générales sur la climatologie, l'hydrologie et les microcentrales hydroélectriques à Madagascar
- Le deuxième chapitre se consacre à l'étude de réseaux de neurones artificiels et les différentes méthodes d'utilisation dans le domaine de l'hydrologie
- Le dernier chapitre traite la modélisation du comportement du débit à partir de la technique bayésienne pour la prédiction du débit.



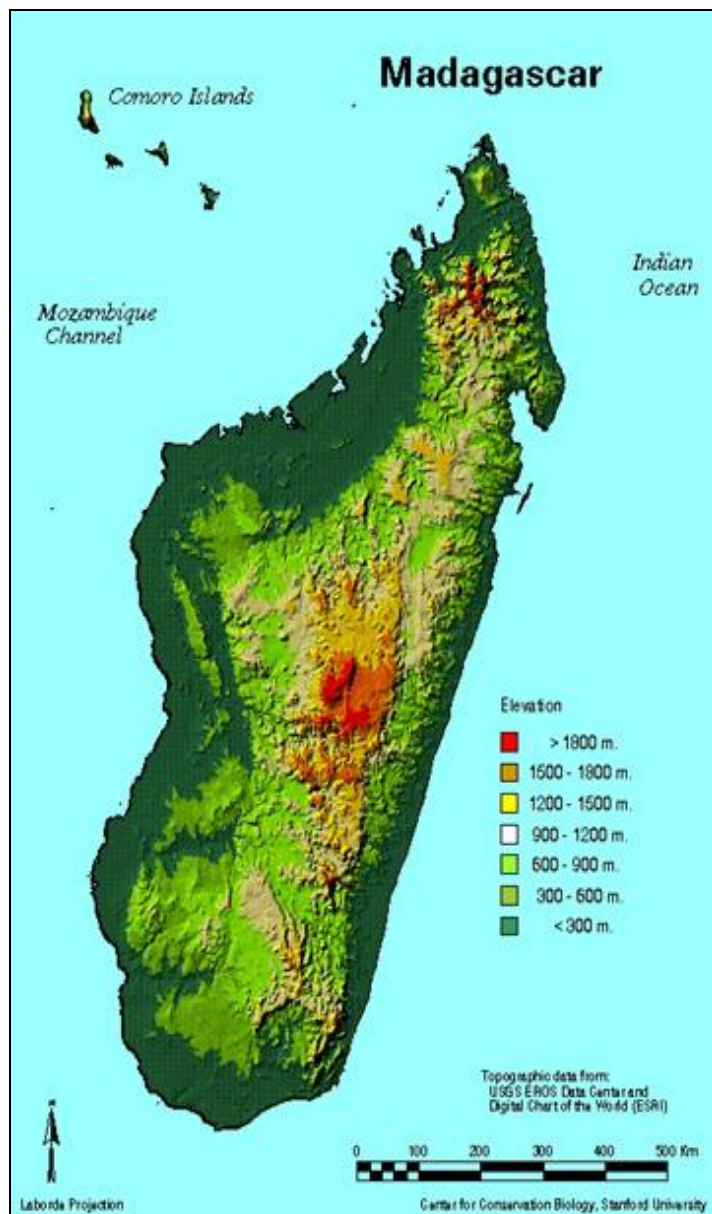
CHAPITRE 1 :  
NOTIONS GENERALES SUR LA CLIMATOLOGIE,  
L'HYDROLOGIE ET LES MICROCENTRALES  
HYDROELECTRIQUES A MADAGASCAR

## 1.1. Situation géographique de Madagascar

Madagascar, avec une superficie de 587041 Km<sup>2</sup>, est situé dans l'Océan Indien et séparé de l'Afrique par le canal de Mozambique. Du Nord au Sud, sa largeur est de 1600 km et sa plus grande largeur d'Est à l'Ouest mesure 560 km. Madagascar se trouve dans la zone intertropicale humide avec deux principales saisons :

- La saison pluviale de Novembre à Avril qui est aussi la saison chaude.
- La saison sèche de Mai à Octobre qui est la saison fraîche.

La figure 1.1 indique la carte physique de Madagascar au niveau de l'Océan Indien [6].



Source : [6]

**Figure 1. 1: Carte physique de Madagascar**

### **1.1.1. Caractéristiques générales de Madagascar**

Madagascar possède une faible densité moyenne de population, caractérisé par un faible degré de développement économique. Dans le domaine de l'énergie, ceci se traduit par de faibles niveaux de consommation et par une grande dispersion de la demande, ce qui implique pour les régions éloignées et peu peuplées des coûts élevés pour la fourniture d'énergies conventionnelles comme les produits pétroliers et l'électricité thermique. Le relief joue un rôle essentiel direct sur le tracé du réseau hydrographique ainsi que sur les variables climatiques.

Madagascar peut être divisé en trois grands ensembles topographiques : Les Hautes Terres centrales, le versant oriental et le versant occidental mais c'est le versant oriental qui renferme une potentialité hydrologique plus intéressante.

### **1.1.2. Hautes Terres centrales**

Les Hautes Terres Centrales de Madagascar sont montagneuses de 800 à 1 200 m d'altitude. Partiellement volcaniques, les montagnes s'élèvent jusqu'à 2 876 m au Maromokotro, dans le Tsaratanana au Nord. La chaîne de l'Ankaratra, près de la ville d'Antananarivo, atteint 2 643 m d'altitude. Les Hautes Terres malgaches présentent un relief très varié. Dans ces massifs prennent source la plus part des grands fleuves de Madagascar. Les terres les plus fertiles se trouvent dans les vallées fluviales du plateau central. L'aspect massif du relief et la complexité du système de drainage des eaux ne favorisent pas l'établissement de voies de communications [8].

### **1.1.3. Versant oriental**

De la baie d'Antongil à Taolagnaro, le versant oriental peut être divisé en trois régions d'importance inégale suivant le relief [9]. De Taolagnaro à la rivière Isandra s'étend une zone relativement étroite, puis entre la rivière Isandra et l'embouchure du Mangoro existe une vaste plaine et enfin au nord à partir de l'embouchure du Mangoro se placent les reliefs les plus complexes du Haut Mangoro et de l'Alaotra qui sont une vaste dépression à 750m d'altitude, le reste d'une partie du lac Alaotra étant en voie d'assèchement.

Du Nord au Sud, suivant le littoral, il existe une série de petits abaissements marécageux isolés les uns des autres par des basses collines. Les rivières se jettent dans un système de lagunes, dont les points les plus élevés de séparation sont creusés pendant l'aménagement du canal des Pangalanes.

## **1.2. Ressources naturelles**

### **1.2.1. Climat**

Madagascar est soumis au climat tropical caractérisé par une alternance de saison des pluies durant la période du mois de Novembre au mois de Mars et de saison sèche durant celle du mois d'Avril au mois d'Octobre. Les durées et caractéristiques de ces saisons varient légèrement d'une région à l'autre, par exemple, une saison tempérée à l'intérieur des terres et une saison aride dans le Sud. L'altitude accentue par ailleurs les variations de température. La saison sèche peut ainsi être particulièrement fraîche sur les Hautes-Terres où, sporadiquement, il peut geler dans les régions d'Antsirabe et Ambatolampy.

Madagascar subit annuellement au cœur de la saison des pluies (du mois de janvier au mois de mars), les dégâts engendrés par les cyclones qui proviennent de l'Océan Indien ou du canal de Mozambique.

### **1.2.2 Saisons**

On distingue principales deux saisons : la saison fraîche en hiver austral et la saison chaude. Les deux saisons se différencient essentiellement entre ces deux saisons résulte de l'intervention de la zone de basses pressions intertropicales en période chaude. A cause de son influence, les caractères des deux saisons sont bien tranchés. La saison chaude est la saison de l'instabilité orageuse et des cyclones tropicaux et la saison fraîche, celle de la sécheresse dans les régions occidentales et des pluies des nuages de basse altitude dans les régions orientales. La saison fraîche a lieu normalement d'Avril à Octobre, la saison chaude couvrant le reste de l'année. Toutefois, la deuxième quinzaine du Mars, les mois d'Avril et d'Octobre constituent des périodes de transition très variables d'une année à une autre.

### **1.2.3. Pluviométrie**

La pluviométrie présente une grande diversité avec des pluies moyennes annuelles de 377 mm à Faux Cap au Sud et 3 792 mm à Maroantsetra au Nord, dans la baie d'Antongil. Les observations montrent que la zone la plus arrosée se situe dans la baie d'Antongil, avec des maxima annuels supérieurs à 5 000 mm à Maroantsetra, alors que la partie ouest est encore plus sèche causé par la perte d'une grande partie de l'humidité des alizés. Le Sud-Est et le Sud sont semi-désertiques. En effet, Toliara ne reçoit que 300 mm de précipitations par année.

### 1.3. Le potentiel hydroélectrique de Madagascar

Madagascar possède d'importantes ressources d'énergies renouvelables. Mais dans la suite, nous nous intéressons uniquement à l'énergie hydroélectrique appelée aussi «houille blanche» [10]. Cette dernière est estimée à 7 800 MW mais seulement une puissance de 276 MW est mise à la disposition des consommateurs jusqu'à présent.

#### 1.3.1. Régimes hydrologiques

Madagascar est presque entièrement situé dans la zone tropicale appartenant à la zone des alizés. Mais les influences du relief et de la latitude créent une très grande diversité du climat entraînant une complexité extrême des régimes hydrologiques.

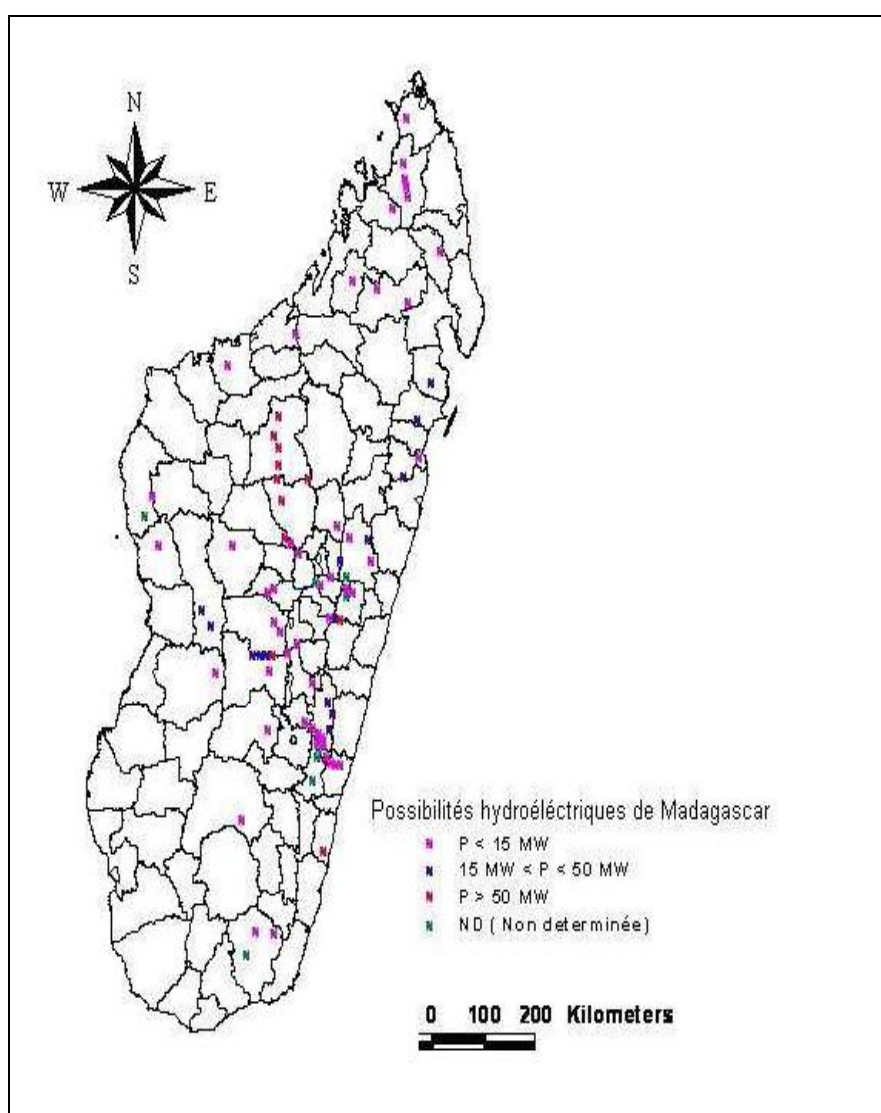
Au Nord-Ouest de l'île, une saison pluvieuse et une saison sèche coexistent tandis qu'au Sud-Ouest, le caractère subdésertique du climat est observé. La présence de couverture végétale à l'Est de Madagascar provoque une pluie abondante, ce qui n'est pas le cas à l'Ouest car la précipitation se réduit en allant progressivement vers l'Ouest. Les plateaux font transition entre les deux régimes.

- Sur les plateaux : la pluviométrie moyenne est de 1 300 mm/an (à Antananarivo). Dans cette région, deux saisons sont nettement observées : une saison de pluies de 6 à 7 mois et une saison sèche de 5 à 6 mois présentant un déficit en eau pendant 4 mois. Cette pluviométrie passe à 1 800 mm à Moramanga pour atteindre 2 800 à 3 000 mm au voisinage de la côte.
- La côte Est : il y existe une forte pluviométrie au Nord (3 400 à 3 500 mm de pluies) et la saison sèche n'y est pas observée pratiquement. Le besoin en eau de cette zone étant satisfait, on n'y rencontre pas de déficit en eau mais au contraire, on y observe de la pluviométrie excédentaire. Tandis qu'au Sud de cette côte Est, la pluie n'est pas assez forte.
- La côte Ouest : on y rencontre une faible pluviométrie surtout au Sud car il n'y a que 500 à 600 mm de pluies chaque année. Elle est encore plus faible dans les régions côtières du Sud-Ouest, mais elle croît vers le Nord.

Dans le Sud, les saisons sont marquées de trois mois de pluies irrégulières contre neuf mois de sécheresse presque totale. On trouve un fort déficit en eau et les nappes souterraines ne sont pas alimentées à cause de l'inexistence des apports occultes des cours d'eau.

### 1.3.2. Sites hydroélectriques identifiés

Beaucoup des sites à potentiel hydroélectrique ont déjà fait l'objet d'une étude préliminaire en estimant leurs puissances brutes [6]. En effet, le Ministère de l'Énergie et des Mines (MEM) a identifié près de 300 sites dont la puissance annuelle combinée est d'environ 7 000 MW pouvant produire 60 000 GWh d'électricité. Actuellement, la puissance installée ne dépasse pas 250 MW sur tout le territoire toutes puissances confondues, soit seulement 3% de la puissance potentielle. Toutes les régions de l'île sont dotées de ressources hydrauliques. Trente (30) sites peuvent développer une puissance au-delà de 50 MW, 70 sites, entre 1 à 5 MW et 200 sites, inférieure à 100 KW.



Source : MEM

**Figure 1. 2: Sites ayant des potentialités hydroélectriques déjà identifiées.**

Malgré leur identification, ces sites nécessitent encore beaucoup d'études des données pour pouvoir confirmer leurs potentiels réels. Ces sites sont identifiés par le service hydrologie du

Ministère de l'Énergie et des Mines. L'aperçu général montre le potentiel des régions des Hautes Terres centrales. La figure 1.2 illustre bien les sites plus grands et susceptibles d'être exploités par l'Etat, par les opérateurs compétents dans le domaine de l'hydrologie ou des grands groupes multinationaux. Ils ont pour inconvénient majeur de se trouver très loin des centres de consommation.

Des sites à faible puissance sont beaucoup plus nombreux et répondent aux besoins de l'hydroélectricité rurale à Madagascar avec des cours d'eau permanents et des possibilités d'aménagement selon leur situation géographique.

#### 1.4. Situation socio économique de l'électrification rurale

##### 1.4.1. Contexte de la demande d'électricité à Madagascar

###### 1.4.1.1. Consommation d'énergie selon les ressources

L'énergie constitue le moteur de toutes les initiatives de la lutte contre la pauvreté. Elle satisfait les besoins quotidiens des individus. Sa rareté et son coût relativement élevé contribuent à l'accroissement du taux de pauvreté à Madagascar.

Selon le bilan national d'énergie indiqué sur le tableau 1.1, le bois reste encore, la source la plus utilisée à Madagascar [10]. Ceci est le reflet de la répartition géographique de la population et aussi celui de la consommation de l'énergie domestique. Seul le milieu urbain est le mieux desservi en matière d'électricité. L'électricité est la ressource la plus utilisée par les industries. Les ménages s'en servent pour les besoins en éclairage et en appareils électroménagers.

**Tableau 1. 1: Consommation d'énergie à Madagascar en l'an 2000**

Type de combustibles	Consommation (kTEP)	Proportion (%)
Bois	1 594	72
Charbon de bois	203	9
Produit pétroliers	238	11
Electricité	107	5
Divers	69	3

Source : MEM

Aujourd'hui, l'épuisement des ressources d'énergies conventionnelles et leur caractère polluant sont à l'origine des graves problèmes de développement et de dégradation environnementale. Ainsi, il faut réussir impérativement à mettre en place une politique énergétique soutenue et à réduire leurs impacts négatifs sur l'environnement et la santé. Les actions correspondantes consistent, entre autres, à promouvoir des solutions profitables à tous actuellement sans compromettre la vie des futures générations.

#### 1.4.1.2. Consommation d'énergie selon les régions

Un déséquilibre est aussi observé sur la consommation urbaine et celle du monde rural. La région centrale de l'île bénéficie du réseau interconnecté de la Jiro sy Rano Malagasy (JIRAMA) tandis que les autres villes dépendent simplement des centrales thermiques à gasoil ou à fuel. Le réseau interconnecté d'Antananarivo développe une puissance totale de 120 MW. La répartition de la consommation nationale en biomasse est la même mais l'utilisation de l'électricité est une image de la forme future de la demande en énergie. Le tableau 1.2 montre la puissance électrique installée sur le réseau interconnecté d'Antananarivo selon la JIRAMA [11].

**Tableau 1. 2: Sources d'énergie à Madagascar**

Site	Puissance (MVA)	Total (MW)	Source
<b>Andekaleka</b>	2*32	51.2	Hydraulique
<b>Antelomita</b>			Hydraulique
<b>Antelomita</b>	3*1.17	8.16	Hydraulique
<b>Antelomita</b>	3*1.17		Hydraulique
<b>Mandraka</b>	4*7.5	24	Hydraulique
<b>Manandona</b>		1.6	Hydraulique
<b>Ambohimanabola</b>	3*7.5	18	Thermique
<b>Antsirabe</b>	9.1+7.3	13.12	Thermique
<b>Mandroseza</b>	4.12	3.3	Thermique
<b>Total Hydraulique</b>	84.96		
<b>Total Thermique</b>	34.32		
<b>Total R-I Tana</b>	119.39		
<b>Total Madagascar</b>	250		

Source : JIRAMA



## **1.4.2. Cadre de l'électrification rurale**

### **1.4.2.1. Législation relative à l'électrification rurale**

Le type d'exploitant et la puissance maximale produite selon les installations sont conformes à la loi n° 98-032 portant réforme du secteur de l'Electricité comme suit :

- Etablissement et exploitation des installations de production de puissance maximale inférieure ou égale à 50 KW pour les installations thermiques et à 150 kW pour les installations hydrauliques.
- Etablissement et exploitation des installations de distribution d'une puissance de pointe inférieure ou égale à 500 kW.

Au-delà de ces puissances, le régime de la concession doit être tenu compte.

Le décret relatif à la Mise en Compatibilité des Investissements avec l'Environnement (MECIE) est conforme à la loi n° 90-033, portant sur la Charte de l'Environnement et ses modifications. Ainsi tout projet d'investissement doit être sujet d'étude d'impact environnemental, surtout celui relatif au secteur de production d'énergie. Spécialement pour les centrales hydroélectriques, leur réalisation doit suivre des guides spécifiques visant entre autres les :

- Partage de l'eau avec d'autres utilisateurs environnants autres que ceux destinés à la conversion énergétique
- Préservation de l'écosystème aquatique.

### **1.4.2.2. Politique nationale du développement de l'électrification rurale**

#### **1.4.2.2.1. Objectif**

L'Etat malgache a récemment réaffirmé sa volonté de lutter contre la pauvreté à travers les développements du monde rural et de son économie, tout en protégeant efficacement son environnement. Pour ce faire, la promotion de l'électrification rurale est l'une des stratégies à adopter.

Effectivement, en zone rurale, il faut la présence d'électricité pour améliorer de façon rapide les conditions de vie de la population (éclairage, réfrigération, pompes hydrauliques, etc...). Par ailleurs, elle permet aussi un développement économique de la région électrifiée, grâce à la création des petites et moyennes industries et entreprises destinées spécialement aux transformations locales des produits agricoles.

#### **1.4.2.2. Stratégie : libéralisation du secteur électrique**

Avant 1975, la société d'intérêt national JIRAMA, détenait le monopôle dans les domaines d'exploitation et de production d'électricité. Ainsi, elle exploitait presque la totalité des installations de production, de transport et de distribution d'électricité. Depuis, la situation économique du Pays se pouvait que se dégrader progressivement pour devenir actuellement catastrophique. Par ailleurs, cette unique société ne pouvait pas assurer les besoins du Pays avec efficacité dans un contexte favorable au développement économique et social du Pays, d'où la nécessité de procéder à une réforme du secteur de l'énergie électrique, en 1999. Cette réforme avait pour but de faire appel à de nouveaux opérateurs afin de les inciter à s'investir dans ce secteur. Cette réforme a permis le renflouement du financement de l'infrastructure électrique par les sociétés privées.

Cette entreprise entraîne nécessairement la promotion du secteur électrique dans un environnement favorable à la concurrence et avantageux aux usagers.

Par ailleurs, cette nouvelle loi doit aussi sécuriser l'investissement de nouveaux opérateurs :

- elle instaure un régime d'octroi d'Autorisations et de Concessions par voie d'appels d'offre ou de candidature spontanée,
- elle reconnaît le statut de propriétaire des installations aux permissionnaires ou concessionnaires,
- elle institue l'Office de Régulation de l'Electricité (ORE) un organe qui assure la régulation, le contrôle et le suivi des activités relatives audit secteur (normes, qualité, tarification, etc.).

#### **1.4.3. Acteurs de l'électricité rurale**

Pour développer l'électrification rurale, le Ministère de l'Energie et des Mines, en concertation avec d'organismes et institutions publics et privés, a créé différents comités et agences assurant la mise en œuvre de la politique de l'électrification rurale à Madagascar. Pratiquement, les rôles de ces différents comités sont complémentaires pour atteindre l'objectif unique qui consiste à électrifier n'importe quelle zone de Madagascar, qu'elle soit rurale ou publique. Il s'agit des :

- Agence de Développement de l'Electrification Rurale : l'Agence de Développement de l'Electrification Rurale (ADER) dont l'objectif majeur consiste à promouvoir et à encourager la soumission de projets en matière d'électrification rurale. De plus, elle statue périodiquement sur les demandes d'octroi de financement et de subvention à la réalisation desdits projets.

- **Fond National de l'Électricité** : le Fond National de l'Électricité (FNE) visant le financement des programmes de développement d'électrification rurale. Ce financement doit permettre, en partie, de subventionner les investissements accordés aux exploitants titulaires d'autorisation ou de concession selon des conditions et modalités fixées par voie réglementaire.
- **Office de Régulation de l'Électricité** : l'Office de régulation de l'Électricité (ORE) est l'organisme régulateur chargé de :
  - Déterminer et publier les textes et les prix réglementaires de l'électricité avec le montant des redevances de transit conformément aux dispositions tarifaires de la loi de réforme,
  - Surveiller leur application correcte et le respect des normes de qualité du service.
  - Contrôler et faire respecter les principes de la concurrence.
  - Calculer les tarifs conformément aux lois et les textes en vigueur.
- **Office National pour l'Environnement** : l'Office National pour l'Environnement (ONE) coordonne les opérations de mise en œuvre des programmes environnementaux nationaux, ceci sous la tutelle du Ministère chargé de l'Environnement. Il assure les suivis, évaluation et application des normes selon la procédure mise en vigueur du secteur électrique et compatible aux investissements avec l'environnement.
- **Les entreprises privées** : la réforme créant la loi de libéralisation du secteur électrique malgache entraîne la croissance du nombre d'entreprises d'électrification rurale dont la majorité n'a aucune expérience en la matière. Ainsi, plus d'une vingtaine de réseaux ruraux sont créés par une quinzaine de compagnies privées depuis 1999. Malheureusement, les centrales électriques installées sont, dans la plupart des cas, des centrales thermiques qui sont tributaires des incessantes augmentations du prix du pétrole. Devant ce contexte, elles ne peuvent pas garantir leur avenir à cause du prix de vente du kWh élevé qui est parfois près de 2 fois supérieur à celui de la ville : soit plus de 1000Ar/kWh au lieu des 600Ar/kWh en ville). En outre, le taux de connexion de ces réseaux est resté très bas, ce qui entraîne un temps d'amortissement très long qui n'est pas favorable à l'investissement.
- **Les associations villageoises** : Aujourd'hui, seule une association villageoise est en charge de la gestion d'un réseau électrique rurale (village d'Antetazambato, près Ambositra 42 kW installé, seulement une dizaine distribués). Il s'agit d'ailleurs du seul réseau rural alimenté par une microcentrale hydroélectrique. Le projet a été

accompagné par différents organismes (IEPF, FONDEM et le cabinet malgache Mihitra) qui ont formé l'association.

## **1.5. Microcentrales hydroélectriques**

### **1.5.1. Définition et généralités**

Il s'agit d'une centrale de production d'électricité dans laquelle le flux créé par l'eau amenée par écoulement libre (canaux) ou par des conduites en charge (conduites forcées) actionne la rotation des turboalternateurs avant de retourner à la rivière.

On classe les Petites Centrales Hydroélectriques (PCH) en fonction de leur puissance installée selon l'Union Internationale des Producteurs et Distributeurs d'Energie Electrique (UNIPED). Il s'agit d'une :

- Petite centrale pour une puissance comprise entre 2 000 kW et 10 000 kW,
- Mini centrale pour une puissance comprise entre 500 kW et 2 000 kW
- Micro centrale Hydroélectrique (MCH) pour une puissance comprise entre 20 kW et 500 kW
- Pico centrale pour une puissance inférieure à 20 kW

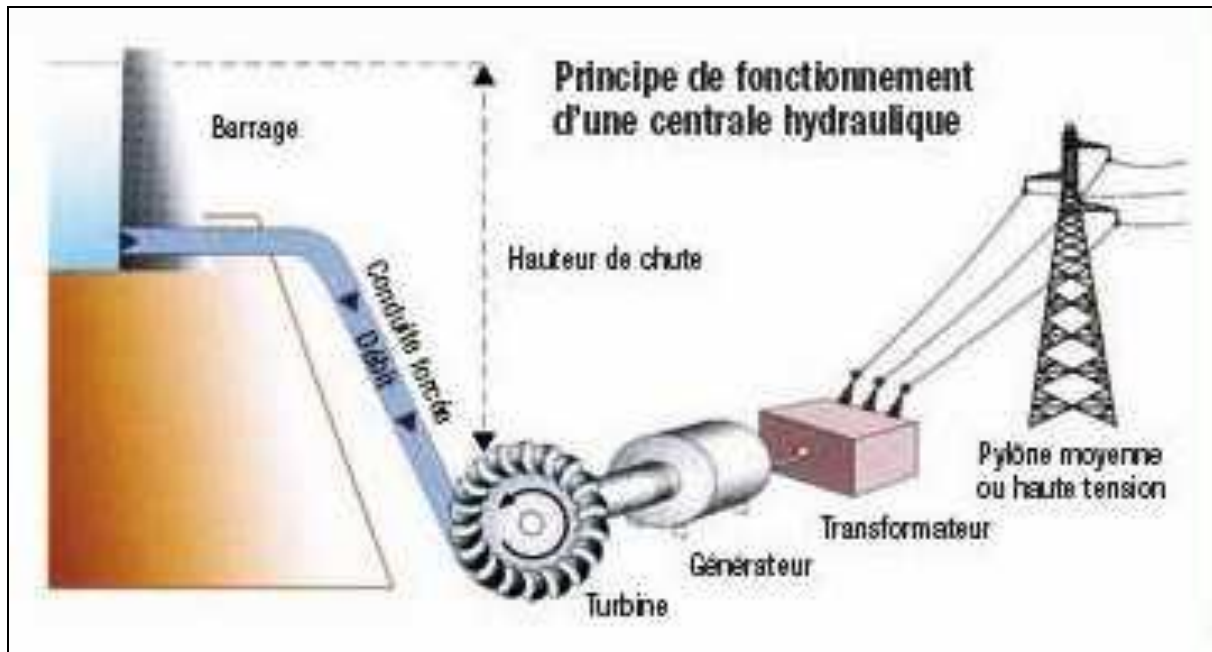
Mais, au sens large, la Microcentrale hydraulique MCH est l'ensemble des installations d'une puissance inférieure à 5000 kW [12].

### **1.5.2 Fonctionnement des centrales hydrauliques**

#### **1.5.2.1. Principe de fonctionnement d'une centrale hydroélectrique**

L'eau accumulée dans les barrages ou dérivées par les prises d'eau, dispose d'une énergie potentielle capable d'entraîner en rotation la turbine d'une génératrice. L'énergie hydraulique se transforme alors en énergie mécanique. Cette turbine accouplée mécaniquement à un alternateur l'entraîne en rotation convertissant ainsi l'énergie mécanique en énergie électrique. A part des ouvrages de génie civil, la MCH est essentiellement constituée par trois machines telles que : la turbine, le générateur et le transformateur.

Le principe de fonctionnement d'une centrale hydroélectrique est montré sur la figure 1.3.



**Figure 1. 3: Principe de fonctionnement d'une centrale hydroélectrique**

### 1.5.2.2. Puissance d'une chute d'eau

La puissance disponible résulte de l'existence simultanée de deux facteurs qui sont :

- la hauteur de la chute
- le débit de la chute.

Dans un site, connaissant le débit et la hauteur de chute nette, on en peut estimer la puissance et la production d'énergie électrique en utilisant les formules suivantes :

$$P_H = QHg \quad (1.1)$$

$$P_{inst} = \eta P_H \quad (1.2)$$

$$W = P_H t \eta f_1 \quad (1.3)$$

En désignant par :

$P_H$  , la puissance hydraulique

$Q$  , le débit d'eau

$H$  , la hauteur de chute

$W$  , l'énergie électrique produite

$P_{inst}$  , la puissance installée

$t$  , la durée de fonctionnement [h]

$f_1$  , le coefficient lié aux variations saisonnières de débit pour des installations au fil

de l'eau sans réservoir régulateur.

$\eta$  , le rendement de l'ensemble turbine-générateur variant de 0,6 à 0,9, [12].

Le rendement de l'ensemble est donné par :

$$\eta = \eta_{tu} \cdot \eta_{rd} \cdot \eta_{al} \cdot \eta_{tr} \cdot \eta_{li} \quad (1.4)$$

Avec :

$\eta_{tu}$  , le rendement de la chute

$\eta_{rd}$  , le rendement du multiplicateur

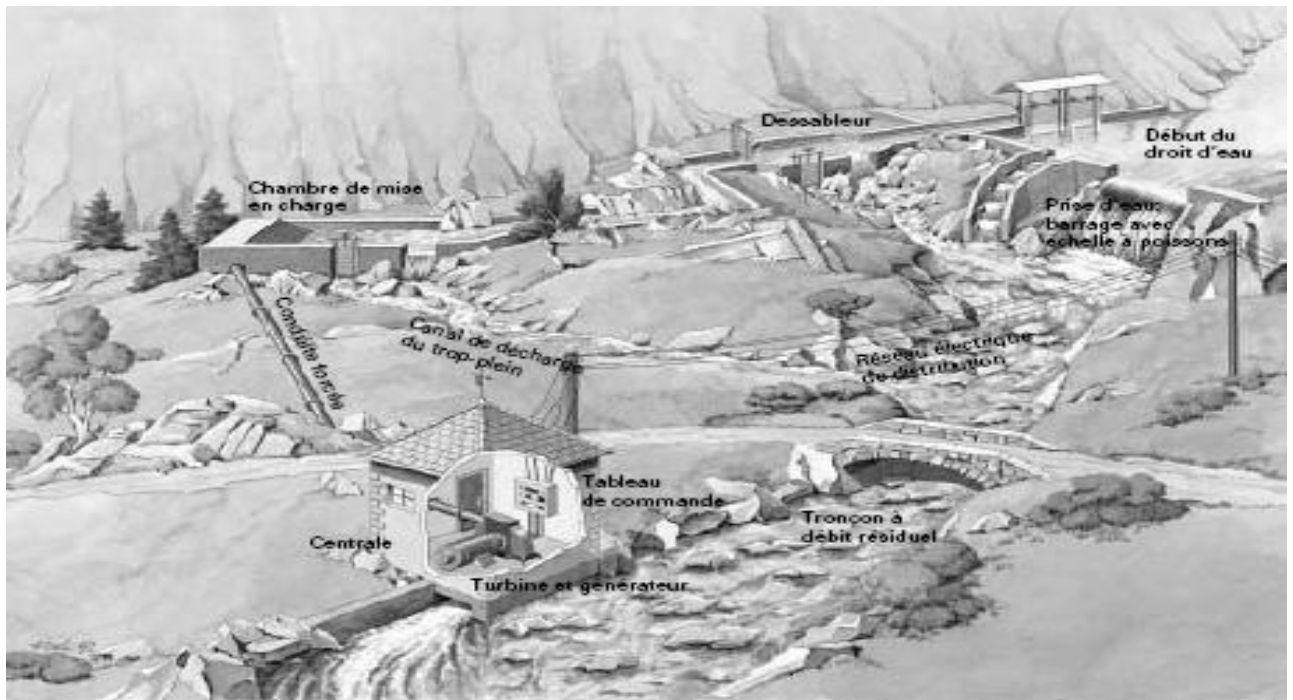
$\eta_{al}$  , le rendement du transformateur

$\eta_{tr}$  , le rendement du transformateur

$\eta_{li}$  , le rendement de la ligne.

## 1.6. Les différentes composantes d'une microcentrale hydraulique

Une installation hydroélectrique, quelque soit la puissance délivrée, comporte toujours en suivant l'écoulement de l'eau, les éléments qui sont indiqués sur la figure 1.4.



**Figure 1. 4 : Illustration des composantes d'une centrale hydroélectrique**

### 1.6.1. Ouvrages de génie civil

Les ouvrages de génie civil sont composés des différents éléments qui sont construits les uns après les autres et dont les rôles sont bien assignés afin d'assurer l'écoulement d'eau de l'endroit de prélèvement à celui où l'électricité est produite.

➤ **Barrage**

Il s'agit d'un ouvrage essentiellement en béton et en moellon permettant la rétention d'eau créant ainsi une hauteur seuil du cours d'eau pour pouvoir garantir l'alimentation de prise d'eau d'une installation hydroélectrique

➤ **Prise d'eau**

Un ouvrage de prise d'eau a pour rôle de diriger l'eau tant qu'elle est en quantité suffisante dans la rivière. Elle est munie d'une vanne qui permet de couper l'arrivée d'eau en cas d'arrêt de la turbine ou en cas de travaux à faire sur l'infrastructure.

➤ **Canal d'amenée**

C'est un canal plus ou moins long qui va conduire l'eau de l'ouvrage de prise à l'ouvrage de tête ou chambre de mise en charge. Ce canal permet aussi de créer une hauteur de chute plus importante en choisissant un tracé approprié.

➤ **Mise en charge**

Elle assure le remplissage de la canalisation qui relie cet ouvrage à la turbine. La conception correcte de l'ouvrage de mise en charge permet de maintenir noyée l'admission de la conduite forcée. L'ouvrage est pourvu d'un trop plein de manière à permettre l'évacuation de l'excès d'eau lors de la réduction des besoins en eau de la turbine. S'il y a présence de régulateur de débit, c'est en générale sur cette chambre que se situera le niveau utile à la régulation.

➤ **Grille et dégrilleur le dessableur**

Ce sont des dispositifs qui permettent de protéger toute l'installation des débris de matière emportés par l'eau. La grille est prévue pour les grosses matières qui seront enlevées manuellement ou automatiquement par le dégrilleur. Les particules fines et le sable qui ne peuvent pas arrêter par la grille sont déposés dans le dessableur. Ces trois éléments sont donc installés en amont de la conduite forcée.

➤ **Conduite forcée**

Généralement en matière synthétique pour les centrales à faible chute et en métal pour les hautes chutes. Elle crée la colonne d'eau qui permet la mise en charge de la turbine. La conduite forcée amène l'eau en écoulement forcé de la chambre de mise en charge à la centrale, et crée en quelque sorte artificiellement la chute d'eau. Lorsque la hauteur de la colonne d'eau est faible, il n'y a pas de conduite forcée mais une construction en béton dans laquelle s'écoule l'eau pour la mise en charge la colonne d'eau.

➤ **Canal de fuite et de restitution**

A la sortie de l'aspirateur-diffuseur, le débit d'eau turbiné sera remis au cours d'eau par l'intermédiaire de ce canal

### 1.6.2. Éléments hydrauliques et électromécaniques

➤ **Vanne de garde**

Placer à l'entrée de la centrale, elle permet de réguler le débit d'eau de la turbine et de la mettre à sec en cas de répartition de cette dernière.

Il y a trois types de vannes selon ses différentes utilisations :

- Protection contre les crues
- Isolation du canal
- Isolation de la turbine

➤ **Distributeur**

Elle reçoit l'eau sous pression prise soit au bas d'une conduite forcée pour une usine de dérivation, soit au fond d'une retenue pour une usine au pied de barrage. Et distribue l'eau autour de la roue pour une turbine à injection totale ou sur une partie seulement pour une turbine à injection partielle. Grâce à des organes mobiles réglables, il modifie la vitesse d'entrée de l'eau dans la roue pour l'alimenter dans les meilleures conditions possibles et en quantité suffisante selon les besoins de l'exploitation.

➤ **Turbine**

Elle est mise en rotation par le débit d'eau qui la traverse. C'est une roue qui transforme l'énergie de l'eau sortant du distributeur en énergie mécanique disponible recueillie sur l'arbre et transmise au générateur de courant à travers un organe de transmission de puissance appelé multiplicateur si la vitesse de rotation de la roue est différente de celle du générateur.

Il y a différentes sortes de turbines :

- Les turbines radiales : **BANKI** (Cross Flow)
- les turbines axiales : **BULBES** (roue, hélice)
- les turbines mixtes : **FRANCIS** ; entrée radiale et sortie axiale,  
**KAPLAN** ; entrée radiale et sortie axiale (roue en hélice)
- les turbines coniques : **DERIAZ**
- les turbines tangentielles : **PELTON**



➤ **Aspirateur-diffuseur**

C'est un gros tube d'échappement (cylindrique ou évasée) qui prend l'eau détendue sortant de la roue et la conduit jusqu'au canal de fuite. Pour les turbines qui ne possèdent pas d'aspirateur-diffuseur, on dit qu'ils sont à échappement libre. C'est le cas par exemple pour les turbines PELTON.

### 1.6.3. Éléments électriques

➤ **Générateur de courant**

Communément appelé aussi alternateur, il converti l'énergie mécanique de la turbine en énergie électrique. Un alternateur fournit le courant alternatif, type de courant plus répandu que le courant continu. Entre cet alternateur et la turbine, qui tourne beaucoup moins vite, est installée une boîte de vitesse afin de rendre le fonctionnement des deux machines compatibles. L'alternateur est dit synchrone ou asynchrone selon qu'il détermine ou non la fréquence des cycles du courant sur le réseau qui est de 50 cycles par seconde sauf sur le continent américain où il est de 60 cycles. Si l'installation est la seule source d'alimentation du réseau ou la source principale, il faut installer un alternateur synchrone.

➤ **Transformateur**

En vue de transporter l'énergie fournie par l'alternateur, il est parfois nécessaire d'élever la tension électrique en raison des chutes de tension en ligne et ensuite de l'abaisser arrivant à proximité des centres de consommation.

➤ **Lignes de transport et de distribution**

Dans la majorité des cas, la centrale de production se trouve éloigné du lieu d'agglomération des consommateurs. Donc les lignes de transport servent à relier la centrale et le poste source de distribution. Par contre, les lignes de distributions ont pour rôle de connecter la station de distribution aux consommateurs.

➤ **Armoire électrique de commande et de distribution**

C'est une armoire où l'on trouve toutes les bornes de connexion des parties sous-tension pour assurer la sécurité et pour centraliser les matériels susceptibles d'opérer en cas de pannes éventuelles, et les matériels électriques de protection et le tableau de commande de l'installation. On peut aussi trouver les blocs de commande des systèmes de régulation de la centrale.

**CHAPITRE 2 :**  
**THEORIE DES RESEAUX DE NEURONES**  
**ARTIFICIELS**

## 2.1. Introduction

Les réseaux de neurones formels, issus de la modélisation simplifiée des mécanismes d'interaction entre les cellules nerveuses donnaient des résultats intéressants de modélisation [13]. Les Réseaux de Neurones Artificiels (RNA) se classent ainsi parmi les bons candidats à la modélisation physique du fait de leur très fort non linéarité et de leur aspect de généralisation. A travers ce chapitre, nous allons découvrir la capacité des réseaux de Neurones Artificiels dans la modélisation physique [14].

## 2.2. Retour sur l'approximation de fonction par des polynômes

Les concepts liés aux réseaux de neurones peuvent être introduits de manière simple dans le cadre d'approximation de fonctions par des polynômes [15].

### 2.2.1. Cas d'une seule variable

Nous disposons d'une base de données de  $N$  points ou couples  $(x^n, t^n)$  et nous cherchons à déterminer le polynôme d'ordre  $m$  qui approche au mieux ces points. Ces polynômes s'écrivent sous la forme :

$$y(x) = y(x, \omega) = \omega_0 + \omega_1 x + \omega_2 x^2 + \dots + \omega_m x^m = \sum_{j=0}^m \omega_j x^j \quad (2.1)$$

Où  $y(x)$  est la sortie du modèle.

Ce polynôme permet d'établir une relation non linéaire entre  $x$  et la sortie  $y$ . La forme précise de la relation est déterminée par la valeur des coefficients (poids dans la terminologie des RNA) du polynôme. Ces coefficients sont regroupés ici dans le vecteur  $w$ .

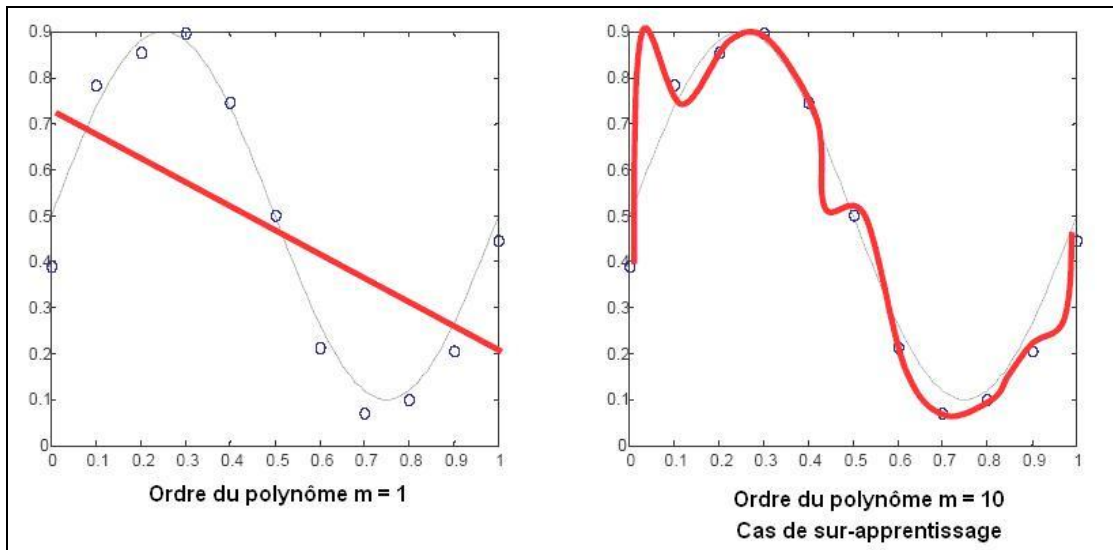
Afin de déterminer les valeurs de ces coefficients, les procédures standards minimisent le critère quadratique suivant :

$$E = \frac{1}{2} \sum_{n=1}^N \{y(x^n; \omega) - t^n\}^2 \quad (2.2)$$

La sortie  $y$  étant linéaire par rapport aux coefficients,  $E$  est une fonction quadratique en  $w$  et il est alors possible de trouver le vecteur  $w^*$  qui minimise  $E$  en résolvant un système d'équations linéaires.

Une fois déterminé, le jeu de paramètre optimal  $w^*$ , il est intéressant d'évaluer la capacité du modèle sur des nouvelles données.

En fait, les performances du modèle dépendent de sa complexité ou plus exactement ici de l'ordre du polynôme.



**Figure 2. 1: Illustration du surapprentissage.**

### 2.2.2. Cas de plusieurs variables

Nous pouvons étendre le modèle du polynôme au cas multivariable. Si nous disposons de  $d$  variables d'entrée et en nous limitant à un polynôme d'ordre 3, nous pouvons écrire :

$$y(x) = y(x, \omega) = \omega_0 + \sum_{i_1=1}^d \omega_{i_1} x_{i_1} + \sum_{i_1=1}^d \sum_{i_2=1}^d \omega_{i_1 i_2} x_{i_1} x_{i_2} + \sum_{i_1=1}^d \sum_{i_2=1}^d \sum_{i_3=1}^d \omega_{i_1 i_2 i_3} x_{i_1} x_{i_2} x_{i_3} \quad (2.3)$$

L'inconvénient de ce type de modèle est que le nombre de coefficients du modèle croît très rapidement avec la dimension de l'espace des entrées. Le nombre de coefficients croît en une loi de puissance du type  $d^M$ . Alors il est nécessaire de disposer de beaucoup de données pour que les coefficients soient correctement déterminés.

Ce problème n'existe pas avec les réseaux de neurones. Pour un nombre fixé de neurones cachés (qui correspond à l'ordre du polynôme), le nombre de coefficients à ajuster croît linéairement avec la dimension de l'espace des entrées. C'est pour cette raison que les réseaux de neurones sont appelés approximateurs universels parcimonieux.

## 2.3. Réseaux de neurones artificiels

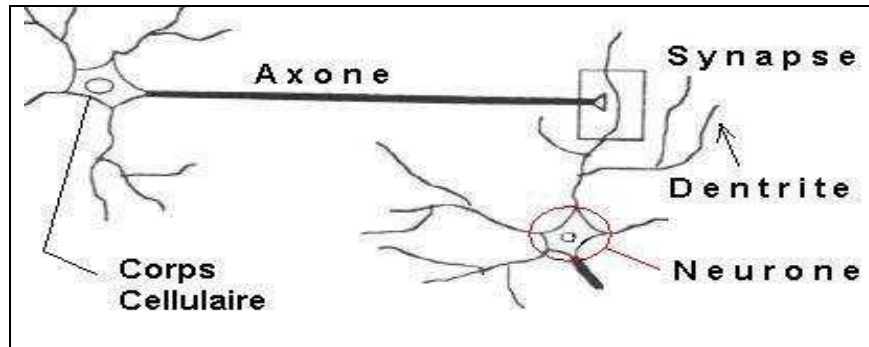
### 2.3.1 Historique

Le début des travaux sur les réseaux de neurones remonte aux travaux de Warren Sturgis Mac Culloch et Walter Pitts vers 1943, bien qu'on puisse trouver d'autres ancêtres. Ils ont donné dans un article une modélisation du neurone biologique. Ils étaient les premiers à montrer théoriquement que les neurones simples peuvent réaliser des fonctions logiques et arithmétiques. A l'époque déjà le cognitivisme cherchait à cerner les principes rendant possibles tout phénomène d'intelligence. Partant de la constatation que la pensée est produite par le cerveau ou en est une propriété, la question qui se présente naturellement

est : « pourquoi ne pas imiter la nature et créer l'équivalent artificiel de l'organe concerné ? » Ainsi, les réseaux de neurones formels sont à l'origine d'une tentative de modélisation mathématique du cerveau humain [16].

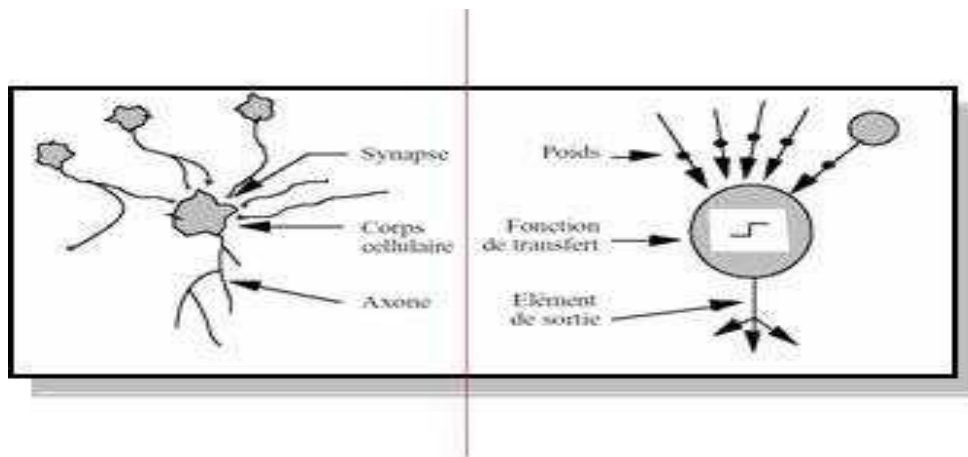
### 2.3.1.1 Neurone formel

La figure 2.1 montre un modèle simplifié des neurones biologiques qui a conduit aux modèles des premiers neurones formels.



**Figure 2. 2: Neurone biologique et cerveau humain.**

Le cerveau humain est composé d'un grand nombre de cellules nerveuses telles que chacune d'elles contient des neurones. Ces derniers ont une partie « cellulaire » et une partie « noyau ». La partie cellulaire se ramifie pour former ce que l'on nomme les dendrites. C'est par les dendrites que l'information s'achemine vers le corps de neurone. Une fois l'information traitée, elle est envoyée vers l'axone. Et la jonction entre les deux neurones s'appellent synapse [4]. En fait un neurone émet un signal en fonction des signaux qui lui proviennent des autres neurones. Au niveau d'un neurone, il y a intégration des signaux reçus au cours du temps, c'est-à-dire une sorte de sommation des signaux. En général, quand une somme dépasse un certain seuil, le neurone émet à son tour un signal électrique. 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 électrique ne peut pas se propager. La transmission se fait alors par l'intermédiaire des substances chimiques : les neuromédiateurs. Quand 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 inter-synaptique. Quand suffisamment des molécules se sont fixées, un signal électrique est émis de l'autre côté et il y a alors une transmission. A partir d'une analyse assez simple de ce modèle biologique, les neurologues Warren Sturgis Mac Culloch et Walter Pitts présentent le modèle de neurone formel de la figure (2.1) [1]. Les réseaux de neurones artificiels sont alors construits sur ce paradigme biologique. [5]



**Figure 2. 3: La ressemblance entre un neurone biologique et un neurone artificiel**

Ils ont modélisé le neurone biologique par un comportement booléen, constitué d'un modèle simplifié de neurones biologiques appelé neurone formel. Ils ont par la suite montré théoriquement que les neurones formels simples pouvaient réaliser des fonctions logiques, arithmétiques et symboliques complexes. Ainsi, le neurone artificiel effectue un automate binaire qui réalise une somme pondérée de stimuli 'S' provenant d'autres neurones. Si la somme des excitations est supérieure à un seuil ' $B_0$ ' donné, alors le neurone est activé, sinon il ne transmet aucune information et ceci selon la fonction suivante :

- Si  $S > B_0$ , la sortie vaut 1, le neurone est actif
- Si  $S < B_0$ , la sortie vaut -1, le neurone est inactif.

En 1949, le psychologue canadien Hebb propose dans son ouvrage que les cellules apprennent en modifiant l'intensité des connexions qui les relient en fonction de leur activité simultanée, comme sous le nom de « loi de Hebb ».

En 1951, Minsky construit le premier modèle réel d'un réseau de neurones appelé « *Snarl* ». Ce dernier était capable d'apprendre en corrigeant l'intensité de ses connexions qui étaient en grande partie aléatoire.

En 1957, Rosenblatt construit le premier neuro-ordinateur nommé « *Mark I* » basé sur le modèle Perceptron et l'applique au domaine de la reconnaissance de forme.

En 1960, l'automaticien Widrow analyse une règle d'apprentissage proche de celle du perceptron connu par « loi de Widrow-Hoff » et développe le modèle Adaline ou Madaline, utilisé pour atténuer le bruit dans les lignes téléphoniques.

Entre 1967 et 1982, il y a eu abandon financier des recherches dans le domaine des réseaux de neurones. En fait, les ordinateurs neuronaux étaient extrêmement lents à cette époque, ils coûtaient très cher, alors que leur performance n'était pas si impressionnante. De plus, en

1969, Minsky et Papert publient un ouvrage qui met en exergue les limitations théoriques du Perceptron et réseaux des neurones artificiels aux problèmes non linéaires.

Toutes les recherches ne sont, bien sûr, pas interrompues. Elles se poursuivent, mais déguisées, sous le couvert de divers domaines comme : le traitement adaptatif du signal, la reconnaissance de forme, la modélisation en neurobiologie, etc.

En 1982, Hopfield est reconnu comme celui à qui l'on doit, de nouveau, le nouvel intérêt aux réseaux de neurones artificiels : il propose de fixer préalablement le comportement à atteindre pour un modèle et construire à partir du résultat escompté, la structure et la loi d'apprentissage.

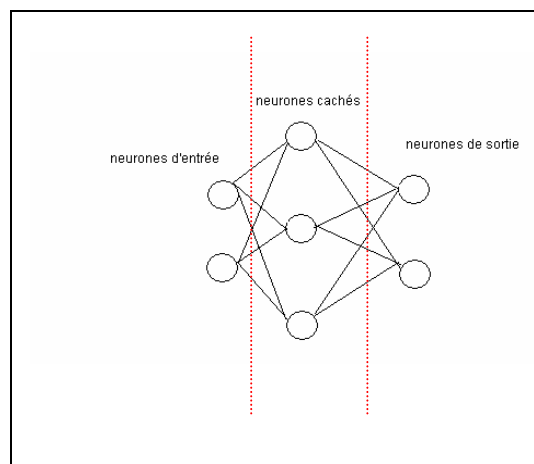
En 1985, la rétropropagation de gradient apparaît. C'est un algorithme d'apprentissage, adapté aux réseaux multicouches, qui permet de résoudre les problèmes non linéaires.

Dès la fin des années quatre-vingts à nos jours, l'intérêt pour les réseaux de neurones ne cessent de croître. Les réseaux multicouches et la rétro-propagation de gradient reste maintenant le modèle le plus étudié et le plus productif au niveau des applications.

#### 2.4. Structure des réseaux de neurones

En général on peut dire que les réseaux de neurones peuvent être caractérisés comme des modèles de calculs apprenant, généralisant et organisant des données. Un réseau de neurones artificiel contient un grand nombre d'unités, les neurones, qui communiquent entre eux en s'envoyant des signaux à travers de liens, appelées connexions synaptiques. En général le système de neurone possède trois types des neurones (voir la figure 3) :

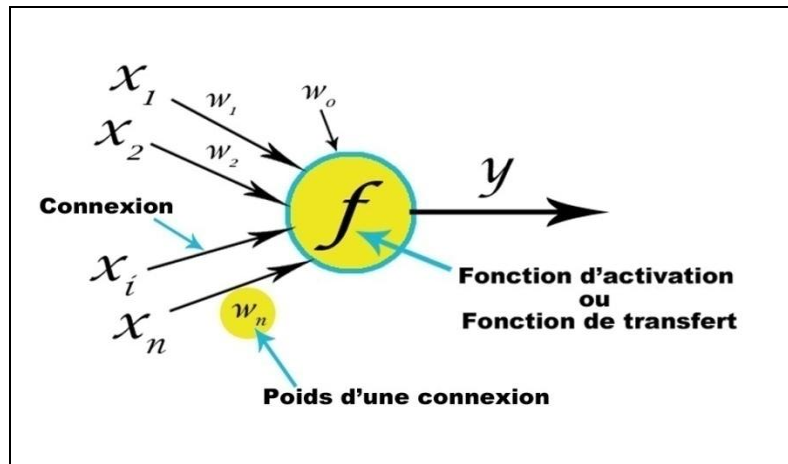
- les neurones d'entrée qui reçoivent les données ;
- les neurones de sortie qui envoient les données par la sortie du système ;
- les neurones cachés, dont les signaux d'entrée et de sortie demeurent dans le système.



**Figure 2. 4: Différents types de neurones**

La structure de ce neurone formel s'est inspirée des études descriptives du neurone biologique. Chaque neurone reçoit des valeurs d'entrées (ou des signaux d'entrée) par le

neurone en amont et les utilise pour calculer le signal de sortie qui se propage par les autres neurones. En concordance avec ce processus, on doit ajuster les poids de connexion synaptique noté ' $W_i$ ' (Abréviation de *weight* en anglais), représentatif de la force de la connexion. Sur la figure 2.5, on peut voir la structure d'un neurone artificiel. Le neurone calcule la somme de ses entrées puis cette valeur passe à travers la fonction d'activation pour produire sa sortie.



**Figure 2. 5: Structure d'un neurone artificiel**

Ainsi, on peut caractériser le neurone formel par :

- Les connexions : Un ensemble de connexions, appelées les connexions synaptiques qui permettent d'accepter les signaux d'entrée et qui sont définies par des poids synaptiques. La connectivité des réseaux de neurones peut être totale (tous les neurones sont connectés entre eux) ou par couche (les neurones d'une couche ne sont connectés qu'aux couches suivantes). Il existe des réseaux monocouches ou multicouches.
- La fonction d'activation (Transfert) : Une fonction d'activation  $f$  qui anime le neurone en déterminant son activation. Le déclenchement de l'activité intervient si la somme des excitations (activité, des neurones excitateurs, pondérée par les poids synaptiques) dépasse un certain seuil propre au neurone.
- Le biais ' $w_0$ ' : Le biais est une activation équivalente de la sortie de neurone. Il peut être envisagé comme une sorte de "poids local" qui est employé dans plusieurs types de fonctions d'activation.
- Les paramètres  $\{x_i\}$  sont attachés aux entrées du réseau, la sortie est une fonction non-linéaire due à la combinaison des entrées  $\{x_i\}$  pondérées par les poids  $\{w_i\}$  traversant à travers la fonction d'activation.



Conformément à l'usage, cette combinaison linéaire est appelée «potentiel». Le potentiel  $v$  le plus fréquemment utilisé est la somme pondérée, à laquelle s'ajoute un terme constant dénommé «biais» et noté par  $w_0$  :

$$v = w_0 + \sum_{i=1}^n w_i x_i \quad (2.4)$$

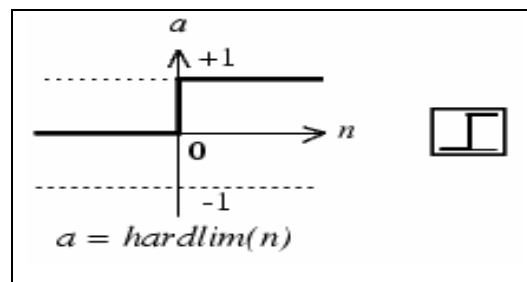
La sortie d'un neurone a ainsi pour équation :

$$y = f(v) = f\left(\sum_{i=1}^n w_i x_i + w_0\right) \quad (2.5)$$

### 2.3.1 Différents types de fonctions de transfert

Il y a plusieurs types de fonction de transfert et les plus utilisées sont généralement les suivantes :

- La fonction "hard limit transfert function" est représentée par la figure 2.6.

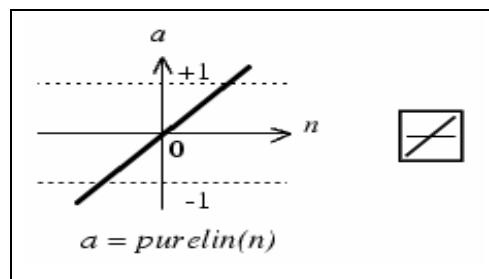


**Figure 2. 6: La fonction « hardlim »**

Cette fonction est de type binaire (fonction de Heaviside) d'équation :

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases} \quad (2.6)$$

- La fonction linéaire "purelin" est de type linéaire.

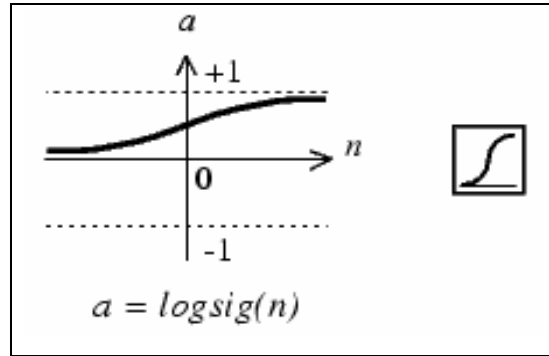


**Figure 2. 7: La fonction « linéaire »**

Elle est représentée mathématiquement par :

$$f(x) = A \cdot x \quad (2.7)$$

- La première fonction sigmoïde « logsig » se présente suivant la figure 2.8.

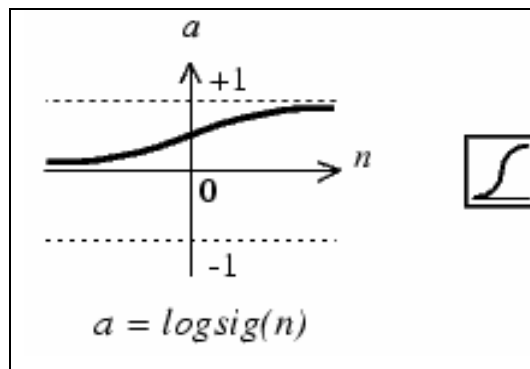


**Figure 2. 8: La fonction « log-sigmoïd »**

Elle est représentée par la fonction mathématique :

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.8)$$

Il existe une deuxième fonction sigmoïde « tansig » comme le montre la figure 2.9



**Figure 2. 9: La fonction « tansig »**

La fonction correspondante est donnée par :

$$f(x) = \frac{2}{1 + e^{-x}} - 1 \quad (2.9)$$

Nous remarquons bien qu'à la différence des neurones biologiques dont l'état est binaire, la plupart des fonctions d'activation sont continues, offrant une infinité de valeurs possibles comprises dans l'intervalle  $[0, +1]$  ou  $[-1, +1]$ .

## 2.4.2 Architectures des réseaux de neurones

### 2.4.2.1 Perceptron

C'est un des premiers réseaux de neurones, conçu en 1958 par Rosenblatt. Il est linéaire et monocouche. Il est inspiré du système visuel. La première couche (d'entrée) représente la rétine. Les neurones de la couche suivante sont les cellules d'association et la couche finale, les cellules de décision. Les sorties des neurones ne peuvent prendre que deux états -1 et 1 ou 0 et 1.

Seuls les poids des liaisons entre la couche d'association et la couche finale peuvent être modifiés.

La règle de modification des poids utilisée est la règle de Widrow-Hoff :

- Si la sortie (celle d'une cellule de décision donc) est égale à la sortie désirée, le poids de la connexion entre ce neurone et le neurone d'association qui lui est connecté n'est pas modifié.
- Dans le cas contraire le poids est modifié en fonction de l'entrée :

$$W_i \leftarrow w_i + k(d - s) \quad (2.10)$$

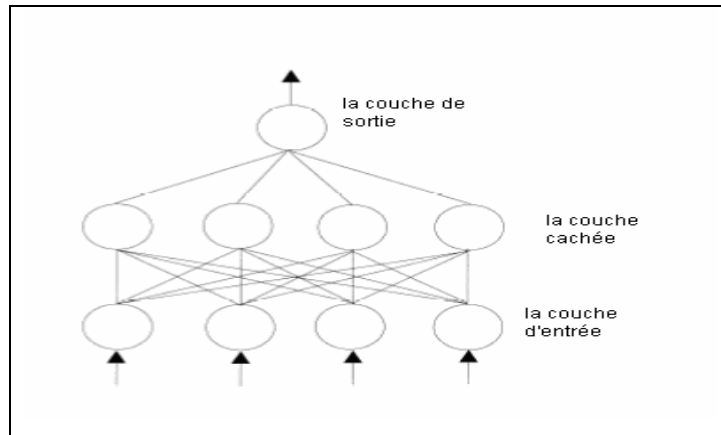
Avec  $k$ , une constante positive,  $s$ , la sortie et  $d$ , la sortie désirée.

### 2.4.2.2 Perceptrons multicouches (PMC)

Parmi les types de neurones les plus connus, le MLP est à la fois le plus célèbre et le plus utilisé. Le perceptron multicouche est un réseau orienté de neurones vers la couche de sortie. Ils sont une amélioration du perceptron comprenant une ou plusieurs couches intermédiaires dites couches cachées. Ils utilisent, pour modifier leurs poids, un algorithme de rétropropagation du gradient, qui est une généralisation de la règle de Widrow-Hoff. Il s'agit toujours de minimiser l'erreur quadratique, ce qui est assez simple quand on utilise une fonction  $F$  dérivable (la sigmoïde par exemple). On propage la modification des poids de la couche de sortie jusqu'à la couche d'entrée.

Les PMC (ou MLP pour *Multi Layer Perceptron*) agissent comme un séparateur non linéaire et peuvent être utilisés pour la classification, le traitement de l'image ou l'aide à la décision. L'idée principale est de grouper des neurones dans une couche. On place ensuite bout à bout plusieurs couches et on connecte complètement les neurones de deux couches adjacentes. Les entrées des neurones de la deuxième couche sont en fait les sorties des neurones de la première couche. Les neurones de la première couche sont reliés au monde extérieur et reçoivent tous le même vecteur d'entrée (c'est en fait l'entrée du réseau). Ils calculent alors

leurs sorties qui sont transmises aux neurones de la deuxième couche, etc. Les sorties des neurones de la dernière couche forment la sortie du réseau.

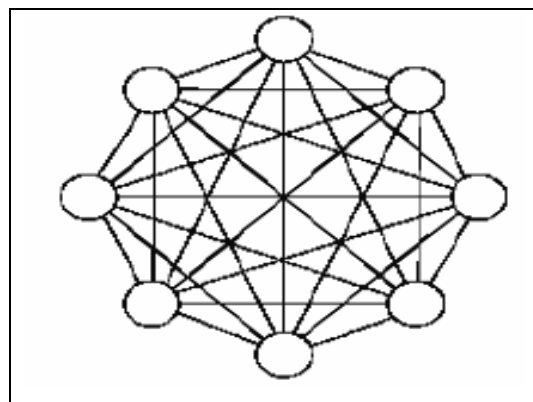


**Figure 2. 10: Réseau multicouche**

Un perceptron multicouche calcule une fonction vectorielle. On peut ajouter les valeurs des connexions synaptiques et des seuils afin de modifier la fonction calculée.

#### 2.4.2.3 Réseau de Hopfield :

Un réseau de Hopfield réalise une mémoire adressable par son contenu. Il s'agit d'un réseau constitué de neurones de Mac. Culloch et Pitts (à deux états, -1 et 1 ou 0 et 1), dont la loi d'apprentissage est régie par la règle de Hebb (1949), qui veut qu'une synapse améliore son activité si et seulement si l'activité de ses deux neurones est corrélée (ou le poids  $w_{ij}$  d'une connexion entre un neurone  $i$  et un neurone  $j$  augmente quand les deux neurones sont activés en même temps).



**Figure 2. 11: Réseau de Hopfield totalement connecté**

### 2.4.3 Classification des réseaux de neurones

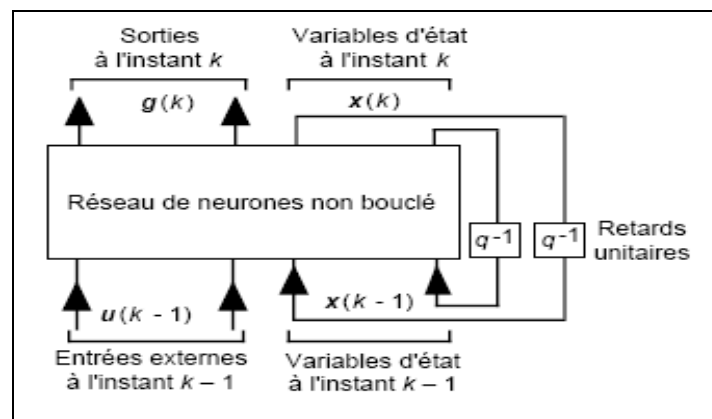
Suivant la logique d'interconnexion entre les couches, on distingue deux grandes familles d'architecture de réseaux :

- Réseaux bouclés ou récurrents (dynamiques)
- Réseaux non bouclés.

#### 2.4.3.1 Réseaux bouclé ou dynamique

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

Ainsi, à chaque connexion d'un réseau de neurones bouclé (ou à chaque arête de son graphe) est attaché, outre un poids comme pour les réseaux non bouclés, un retard qui est un multiple entier ou 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. « Tout réseau de neurones bouclé, aussi complexe soit-il, peut être mis sous une forme canonique (voir figure 2.13), comportant un réseau de neurones non bouclé dont certaines sorties (les variables d'état) sont ramenées aux entrées par des bouclages de retard unité » [17].

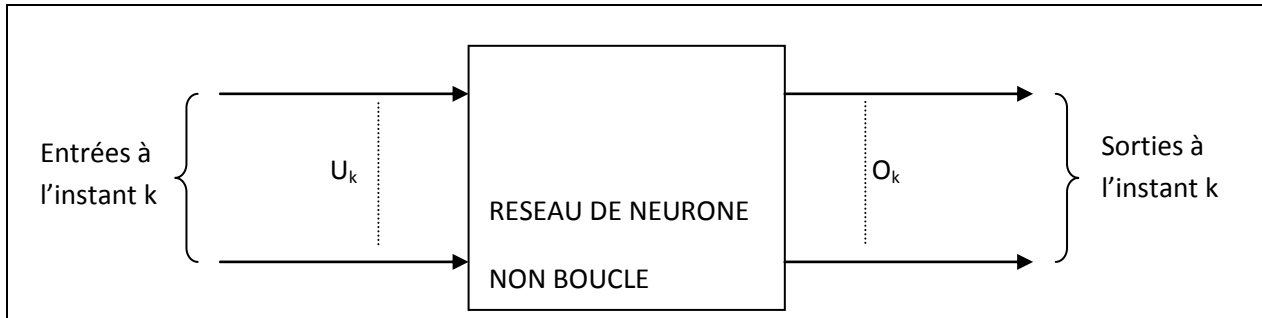


**Figure 2. 12: Forme canonique d'un réseau de neurones bouclé.**

Le symbole  $q^{-1}$  représente un retard d'une unité de temps.

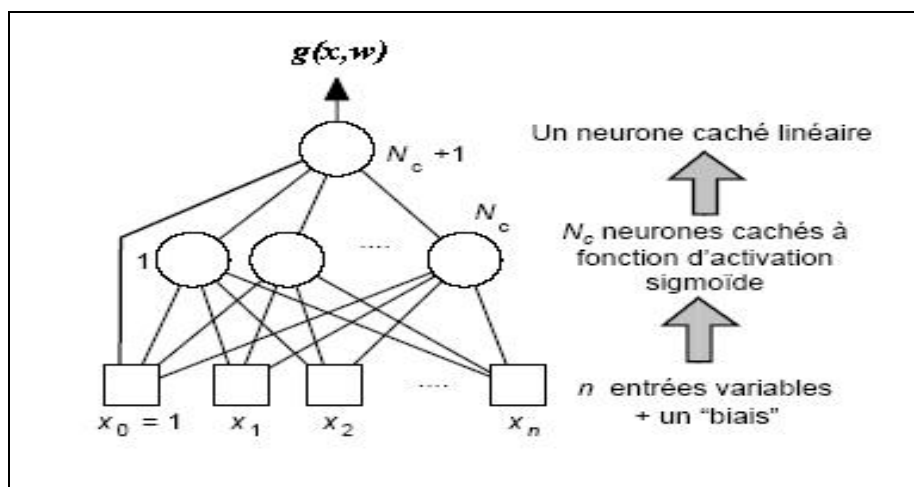
### 2.4.3.2 Réseau non bouclé

Les réseaux non bouclés produisent les signaux de sortie à partir d'information émanant directement d'autres cellules qui ne sont l'objet d'aucune contre réaction. On parle alors de réseaux à propagation directe (*feedforward* en anglais). Les neurones qui effectuent le dernier calcul de la composition de fonction s'appellent neurones de sorties et ceux qui effectuent le calcul intermédiaire sont les neurones cachés [12].



**Figure 2. 13: Schéma synoptique d'un réseau de neurone non bouclé**

Une autre façon d'assembler les neurones entre eux consiste à constituer des couches de neurones en interdisant toute connexion entre neurones de la même couche et aussi les connexions entre les neurones de deux couches consécutives. Cette structure particulière est donnée par les RNA multicouches "Perceptrons Multicouches", PMC (*Multi Layer Perceptron, MLP*). Elle est la plus utilisée dans le cadre d'approximation des fonctions. Il a été en effet démontré qu'un réseau de type *feedforward* avec seulement une couche cachée et un neurone de sortie linéaire est capable d'approcher toute fonction non linéaire suffisamment régulière avec une précision fixée, sous réserve que le nombre de neurones sur la couche cachée soit suffisant [18].



**Figure 2. 14: Réseau PMC à n entrées,  $N_c$  neurones cachés et un neurone de sortie**

En effectuant toutes opérations, nous arrivons à la sortie du réseau la fonction :

$$\begin{aligned}
 g(x, w) &= \sum_{i=1}^{N_c} \left[ w_{N_c+1,i} f \left( \sum_{j=1}^n w_{ij} x_j + w_{j0} \right) \right] + w_{N_c+1,0} \\
 &= \sum_{i=1}^{N_c} \left[ w_{N_c+1,i} f \left( \sum_{j=0}^n w_{ij} x_j \right) \right] + w_{N_c+1,0}
 \end{aligned}
 \tag{2.11}$$

Contrairement aux réseaux récurrents, le temps ne joue aucun rôle fonctionnel dans un réseau de neurones non bouclé. Pour cette raison, les réseaux non bouclés sont souvent appelés "réseaux statiques", par opposition aux réseaux bouclés ou "dynamiques".

## 2.4.4. Perceptrons Multicouches

### 2.4.4.1. Mise en œuvre

Cette mise en œuvre du perceptron multicouche comporte une partie conception dont l'objectif principal est de permettre de choisir la meilleure architecture possible et une partie numérique destinée à réaliser l'apprentissage d'un réseau de neurones. Dans le cas général, un PMC peut posséder un nombre quelconque de couches et aussi un nombre quelconque de neurones par couche, mais en vue de perfectionner le fonctionnement d'un PMC d'un côté et minimiser autant que possible le temps de calcul d'autre côté, il faut chercher la meilleure architecture optimale. A partir de cette architecture et d'une base d'exemples disponibles (Base d'apprentissage), on détermine le réseau optimal [19].

### 2.4.4.2. Approximation universelle

La famille des réseaux de neurones à une couche de neurones cachée possède la propriété d'approximation universelle. Cette approximation a été démontrée [19], par [Cybenko, 1989] et [Funahashi, 1989] et [HORNIK *et al.* 1989], [HORNIK *et al.* 1990], [HORNIK 1991] et peut s'énoncer comme suit « Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation et un neurone de sortie linéaire. »

C'est ainsi que le nombre de neurones cachés doit être choisi convenablement pour obtenir la précision voulue.

Cette propriété, qui n'est qu'un théorème d'existence et ne donne pas de méthode pour trouver les paramètres du réseau, n'est pas spécifique aux réseaux de neurones. C'est la propriété suivante qui leur est particulière et fait tout leur intérêt.

#### **2.4.4.3. La propriété de parcimonie**

Dans la pratique, pour la recherche d'une modélisation d'un processus à partir des données, on s'efforce toujours d'obtenir les résultats les plus satisfaisants possibles avec un nombre minimum de paramètres ajustables. C'est de ce sens que [Hornik et al. 1994] ont montré que [20] : « Si le résultat de l'approximation (c'est-à-dire la sortie du réseau de neurones) est une fonction non linéaire des paramètres ajustables, elle est plus parcimonieuse que si elle est une fonction linéaire de ces paramètres. De plus, pour des réseaux de neurones à fonction d'activation sigmoïdale, l'erreur commise dans l'approximation varie comme l'inverse du nombre de neurones cachés et elle est indépendante du nombre de variables de la fonction à approcher. Par conséquent, pour une précision donnée, donc pour un nombre de neurones cachés donné, le nombre de paramètres du réseau est proportionnel au nombre de variables de la fonction à approcher ».

Les réseaux de neurones, en raison de leurs propriétés mentionnées précédemment sont des bons « approximateurs » de fonction de régression. C'est ce qui justifie leur utilisation pratique à partir d'un nombre fini des points expérimentaux comme base d'apprentissage.

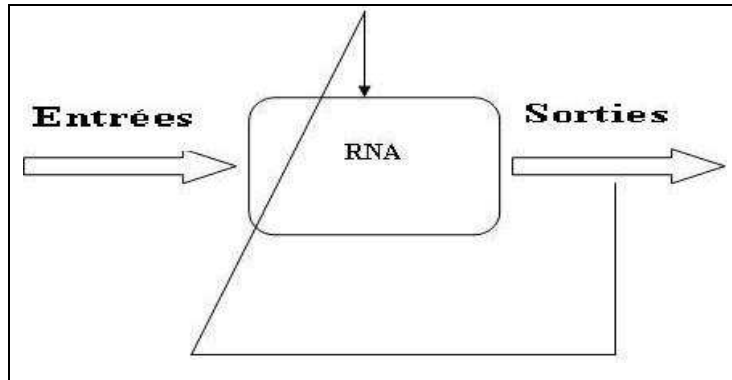
#### **2.5. Phase d'apprentissage**

Une fois l'architecture du réseau de neurones choisie, il est nécessaire d'effectuer un apprentissage. Ce dernier se définit comme étant une phase de développement du réseau de neurones pendant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré. Au cours de cette phase, les valeurs de poids et biais de toutes les couches du réseau permettant à la sortie du réseau de neurones d'être aussi proche que possible de l'objectif fixé sont déterminées moyennant d'algorithmes d'apprentissage. Deux types d'apprentissages sont à distinguer :

- L'apprentissage non supervisé.
- L'apprentissage supervisé.



### 2.5.1. L'apprentissage non supervisé

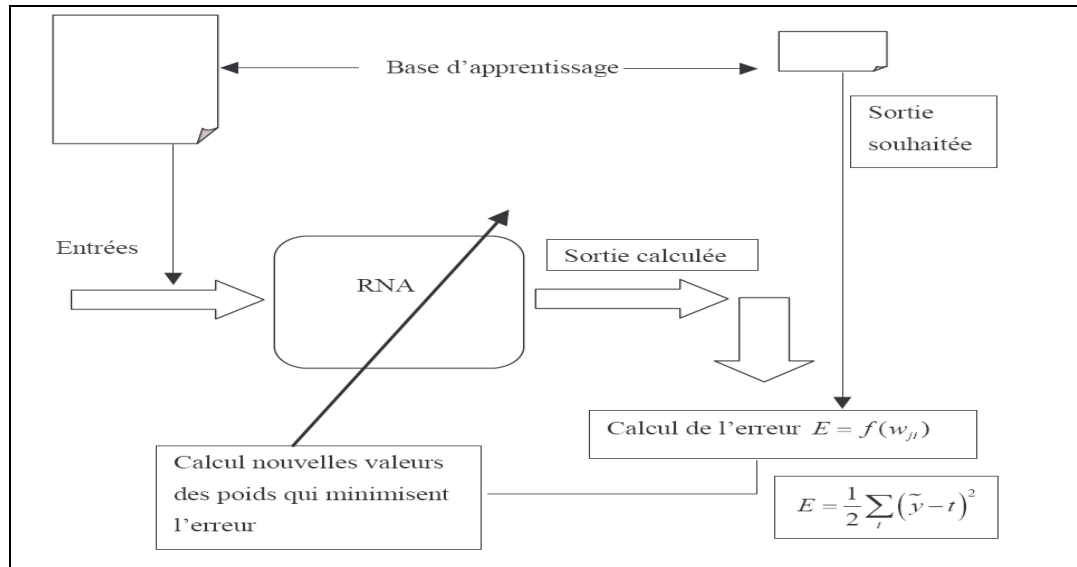


**Figure 2. 15: Mode non supervisé**

L'apprentissage est basé sur des probabilités. Le réseau adaptatif va se modifier en fonction des régularités statistiques de l'entrée et établir des catégories, en attribuant et en optimisant une valeur de qualité, aux catégories reconnues. On ne sait pas à priori si la sortie est valable ou non. Les entrées sont projetées sur l'espace du réseau.

### 2.5.2. Le mode supervisé

Le réseau adaptatif RNA compare les résultats qu'il a calculés en fonction des entrées fournies et la réponse attendue en sortie. Ainsi le réseau va se modifier jusqu'à ce qu'il trouve la bonne sortie souhaitée, c'est-à-dire celle attendue, correspondant à une entrée donnée. Les différentes réponses sont connues à priori. On dispose d'une base d'apprentissage qui contient un ensemble d'observation sous forme des couples entrées/sorties associées. Les poids sont modifiés en fonction des sorties désirées [2]. Pour ce mode d'apprentissage, il existe plusieurs algorithmes pour la minimisation de l'erreur d'apprentissage entre sorties et cibles. Ces techniques et exemples d'apprentissage sont traités dans l'ouvrage du groupe de recherche PDP (*Parallel Distributed Processing*) [21] [22], avec deux ouvrages de référence sur les modèles connexionnistes.



**Figure 2. 16: Mode supervisé**

### 2.5.3 Algorithmes de minimisation

- Principe des algorithmes utilisés : Si  $J(w)$  la fonction de coût ou fonction d'erreur, les algorithmes utilisés nécessitent que  $J(w)$  soit dérivable par rapport aux poids. Le principe c'est de situer en un point initial et de trouver une direction de descente du coût dans l'espace des paramètres  $w$  en se dirigeant d'un pas vers cette direction. Un nouveau point est ainsi atteint et l'on itère la procédure jusqu'à satisfaction du critère d'arrêt. Ainsi, à l'itération  $k$ , nous calculons le poids par :

$$w_k = w_{k-1} + \alpha_{k-1} \cdot d_{k-1} \quad (2.12)$$

Avec  $\alpha_k$ , le pas de descente

$d_k$ , la direction de la descente

A partir de cette équation nous avons les différents algorithmes se distinguant par le choix de ces deux quantités qui sont :

- Descente du gradient : L'algorithme le plus simple consiste à choisir comme direction de descente l'opposé du gradient de la fonction de coût ( $d_k = -\nabla J(w) = -\text{Grad}(J(w_k))$ ). Cette méthode est efficace loin du minimum et permet uniquement de s'en approcher. Pour cette raison, la détermination du pas n'est pas cruciale : loin du minimum, il faut seulement vérifier que le pas n'est ni trop petit ni trop grand. Et pratiquement, nous avons utilisé, selon les cas, deux méthodes :  
Soit un asservissement par la norme du gradient :

$$\alpha_k = \frac{\alpha_0}{1 + |\nabla J(w_k)|} \quad (2.13)$$

Où  $\alpha_0$  est une constante qui vaut typiquement 0,01.

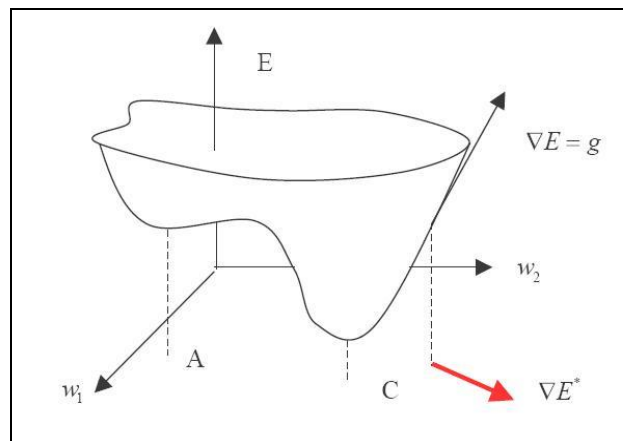
Soit la méthode de Goldstein pour laquelle le pas est adapté afin de satisfaire deux conditions [18] :

$$J(w_k + \alpha_k d_k) \prec J(w_k) + m_1 \alpha_k \nabla J^T(w_k) d_k \quad (2.14)$$

$$J(w_k + \alpha_k d_k) \succ J(w_k) + m_2 \alpha_k \nabla J^T(w_k) d_k$$

La première condition nous rassure que le pas choisi n'est pas trop grand, alors que la deuxième, qu'il n'est pas trop petit. Et habituellement, les valeurs des deux paramètres  $m_1$  et  $m_2$  sont respectivement égaux à 0,1 et 0,7.

Ces deux méthodes de recherche du pas sont "économiques", car elles ne demandent pas de calculs inutiles de gradient (seul celui dans la direction de descente est nécessaire). Le principe de la descente du gradient est illustré sur la figure suivante.



**Figure 2. 17: Principe de la descente du gradient**

- Méthode de Newton : La méthode de Newton utilise la courbure de la fonction de coût du développement du 2ème ordre pour atteindre le minimum. La modification des paramètres s'écrit ainsi :

$$w_k = w_{k-1} - H_{k-1}^{-1} \nabla J(w_{k-1}) \quad (2.15)$$

La direction de la descente est  $-H_{k-1}^{-1} \nabla J(w_{k-1})$  où  $H_{k-1}^{-1}$  est l'inverse du Hessien de la fonction de coût et le pas constant est fixé à un. Cet algorithme converge en une seule itération pour une fonction quadratique. C'est donc un algorithme qui est inefficace loin du minimum de la fonction et très efficace près du minimum.

En pratique, les calculs du Hessien et surtout de son inverse sont à la fois complexe et source d'instabilités numériques et nous avons préféré utiliser la méthode de "quasi-Newton" qui est basée sur celle de Newton.

- La méthode de quasi-Newton : Les méthodes de quasi-Newton consistent à approcher l'inverse du Hessien plutôt que de calculer sa valeur exacte. La modification des paramètres s'écrit :

$$w_k = w_{k-1} - \alpha_{k-1} \cdot M_{k-1} \cdot \nabla J(w_{k-1}) \quad (2.16)$$

Avec la suite  $M_k$  qui est construite de façon à converger vers l'inverse du Hessien avec  $M_0$  égale à la matrice identité. Cette suite est construite grâce à la méthode dite BFGS dont la vitesse de convergence est beaucoup plus grande que celle de la méthode du gradient. De plus, elle est relativement insensible au choix du pas, qui peut être déterminé économiquement par la méthode de Goldstein [23].

- Problème des minima locaux : Les minima trouvés par les algorithmes précédents sont des minima locaux. Le minimum trouvé dépend du point de départ de la recherche c'est-à-dire de l'initialisation des poids. En pratique, il faut effectuer plusieurs minimisations avec des initialisations différentes, pour trouver plusieurs minima et retenir le "meilleur". Il est néanmoins impossible et généralement inutile, de s'assurer que le minimum choisi est le minimum global. Les réseaux de neurones à couches présentent des symétries, si bien que l'on peut montrer que pour une architecture avec  $N_c$  neurones cachés, il existe  $2^{N_c} N_c!$  minima équivalents [Bishop, 1995].
- L'erreur quadratique : Pour les problèmes de régression, l'ensemble d'apprentissage est constitué d'exemples pour lesquels la sortie désirée  $t$  est une variable continue. La fonction de coût la plus utilisée est l'erreur quadratique sur la base d'apprentissage : elle consiste à minimiser la somme des carrés des erreurs entre la sortie du réseau et la valeur réelle de la sortie.

$$J(w) = \frac{1}{2} \sum_{i=1}^N (y_i(w) - t_i)^2 \quad (2.17)$$

Cette fonction de coût est issue du principe de maximum de vraisemblance avec une hypothèse gaussienne sur la distribution des sorties.

- La technique de rétropropagation : La rétropropagation est actuellement l'outil le plus utilisé dans le domaine de réseaux de neurones. C'est une technique de calcul des dérivées qui peut être appliquée à n'importe quelle structure de fonctions dérivables.

Mathématiquement, cette méthode est basée sur l'algorithme de descente du gradient et utilise les règles de dérivation des fonctions dérivables.

Pour déterminer le gradient de la fonction de coût par rapport aux poids des couches cachées, une technique particulière est utilisée : elle est connue sous le nom de rétropropagation de l'erreur. Cette technique s'appuie sur la structure parallèle du RNA pour calculer le gradient relatif aux poids des couches qui précèdent la couche de sortie. D'où l'erreur observée sur la couche de sortie externe est rétro propagée sur les couches internes cachées et les poids sont modifiés selon leur degré de participation à l'erreur. [16][17]. Le but est de calculer la dérivée de l'erreur par rapport à tous les

poids du réseau. Pour un exemple, la dérivée donne le terme  $\frac{\partial E^n}{\partial w_{ji}}$ .

Durant la phase avant, avec les entrées  $x_i$  de poids  $w_{ji}$  faisant apparaître dans les RNA les fonctions intermédiaires  $a_j$  et  $z_j$  données et corrélées entre elles, la sortie  $y_k$  est telle que :

$$\begin{aligned} a_j &= \sum_{i=1}^n w_{ji} x_i \\ z_j &= f(a_j) \\ a_k &= \sum_{j=1}^{N_c} w_{kj} z_j = \sum_{j=1}^{N_c} w_{kj} f(a_j) \\ y_k &= g(a_k) \end{aligned} \quad (2.18)$$

Pour la phase arrière, nous disposons de la valeur cible  $t_k$ :

$$E^n = E^n(g(a_k)) = \frac{1}{2} (y_k - t_k)^2 \quad (2.19)$$

Pour évaluer l'erreur, nous avons utilisé la relation :

$$\frac{\partial E^n}{\partial w_{kj}} = \frac{\partial E^n}{\partial a_k} \cdot \frac{\partial a_k}{\partial w_{kj}} = \frac{\partial E^n}{\partial a_k} \cdot z_j = \delta_k \cdot z_j \quad (2.20)$$

Où  $\delta_k$  est appelé signal d'erreur pour l'unité k et nous l'avons évalué par :

$$\delta_k = \frac{\partial E^n}{\partial a_k} = g'(a_k) \frac{\partial E^n}{\partial y_k} \quad (2.21)$$

Quant à l'évaluation du signal d'erreur pour les unités cachées, elle est déterminée par :

$$\frac{\partial E^n}{\partial w_{ji}} = \delta_j \cdot x_i \quad (2.22)$$

Avec  $\delta_j$  se déduisant des signaux d'erreur précédents, à partir :

$$\delta_j = \frac{\partial E^n}{\partial a_j} = \sum_k \frac{\partial E^n}{\partial a_k} \cdot \frac{\partial a_k}{\partial a_j} = \sum_k \delta_k \cdot \frac{\partial a_k}{\partial a_j} = \sum_k \delta_k w_{kj} f'(a_j) \quad (2.23)$$

La formule de rétro propagation du signal d'erreur est donc donnée par :

$$\delta_j = f'(a_j) \sum_k \delta_k w_{kj} \quad (2.24)$$

En résumé, l'algorithme de rétro propagation que nous avons adoptées dans notre démarche est le suivant :

- Appliquer les entrées et phase de propagation avant pour calculer les activations des neurones cachés et des sorties.
- Calcul des signaux d'erreur  $\delta_k$  pour tous les neurones de sortie.
- Rétro propager les signaux d'erreur  $\delta_k$  pour obtenir les  $\delta_k$  au niveau de neurones cachés
- Utiliser la relation générale  $\frac{\partial E^n}{\partial w_{ji}} = \delta_j \cdot x_i$

En définitive, la dérivée de l'erreur totale E est obtenue en répétant ces étapes début-fin pour chaque exemple et en sommant les dérivées pour chaque exemple, soit :

$$\frac{\partial E}{\partial w_{ji}} = \sum_1^{N^{\text{exp}}} \frac{\partial E^n}{\partial w_{ji}} \quad (2.25)$$

Les fonctions sigmoïdes non linéaires et différentiables, ont la particularité de voir leurs dérivées s'exprimer selon l'expression des fonctions elles-mêmes. Cette dernière propriété facilite les calculs lors de la mise œuvre de la phase d'apprentissage.

## 2.6. Phase de généralisation

A la fin de l'apprentissage, le réseau devrait être capable de généraliser, autrement dit avoir été capable de capturer les tendances incluses dans les données de la base d'apprentissage, ceci afin de produire des sorties viables à la présentation des nouvelles entrées. Les performances du réseau neuronal seront alors déterminées sur des exemples faisant partie de la base dite base de généralisation (test). Ces performances sont généralement données par le calcul de l'erreur en généralisation.

Le RNA est en phase d'utilisation, ainsi les poids obtenus lors de la phase d'apprentissage sont fixés.

## 2.7. Le problème de surapprentissage

### 2.7.1. Définition de surapprentissage

Si l'on considère un ensemble d'apprentissage et une fonction de coût quadratique, en vertu de la propriété d'approximation universelle exposée au paragraphe 2.3.4.2, il est toujours possible d'obtenir une fonction de coût aussi petite que l'on veut sur l'ensemble d'apprentissage, à condition de mettre suffisamment de neurones cachés, mais choisir un nombre très grand de neurones cachés est inapproprié : c'est le contexte de « surapprentissage ».

Il faut apprendre au réseau suffisamment pour atteindre le degré de flexibilité voulu, et aussi éviter ce surapprentissage.

### 2.7.2. Biais et variance

Le surapprentissage est souvent expliqué grâce aux concepts de biais et variance introduits dans la communauté des réseaux de neurones par [Geman *et al.*, 1992].[24]. Si l'on considère plusieurs ensembles d'apprentissage, le biais rend compte de la différence moyenne entre les modèles et l'espérance mathématique de la grandeur à modéliser. Le biais est donc lié à la valeur du bruit du processus que l'on cherche à modéliser. La variance rend compte des différences entre les modèles selon la base d'apprentissage utilisée. On parle souvent de compromis entre le biais et la variance. Si un modèle est trop simple par rapport au processus à modéliser, alors son biais est élevé, mais sa variance est faible puisqu'il est peu influencé par les données. Si un modèle est trop complexe, son biais est faible puisqu'il est capable de s'ajuster exactement à la base d'apprentissage, mais sa variance est élevée puisqu'une nouvelle base avec une réalisation différente du bruit peut entraîner un modèle très différent: c'est le cas du surapprentissage.

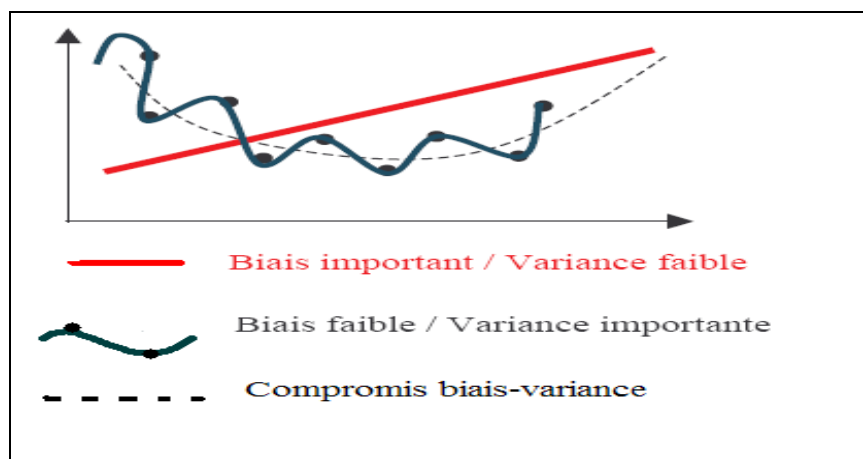


Figure 2. 18: Illustration du compromis biais/variance

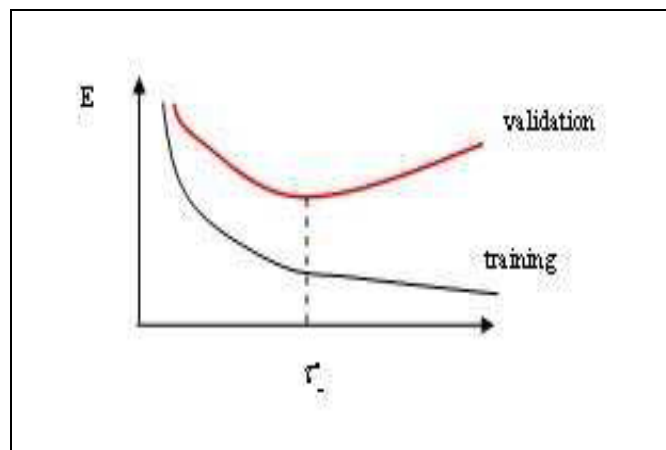
Ainsi, la complexité du modèle doit être ajustée pour trouver un compromis entre le biais et la variance. De ce fait, plusieurs techniques sont utilisées habituellement comme le Weight Decay ou l'Early stopping.

### 2.7.2.1. Early stopping

Comme nous l'avons vu précédemment, l'apprentissage consiste à minimiser, grâce à un algorithme itératif, une fonction de coût calculée sur la base d'apprentissage. La méthode de « *Early stopping* » consiste à arrêter les itérations avant la convergence de l'algorithme. Si la convergence n'est pas menée à son terme, le modèle ne s'ajuste pas trop finement aux données d'apprentissage : le surapprentissage est limité.

Pour mettre en œuvre cette méthode, il faut déterminer le nombre d'itérations à utiliser pendant l'apprentissage. La méthode la plus classique consiste à suivre l'évolution de la fonction de coût sur une base de validation, et à arrêter les itérations lorsque le coût calculé sur cette base commence à croître.

Cette base de validation dont généralement représentée par 10 à 25% tirée à partir de l'ensemble d'apprentissage est utilisée à la fin de chaque apprentissage. On mesure à la fois l'erreur d'apprentissage et l'erreur de validation (cf. figure).



**Figure 2. 19: Early stopping**

Cependant, cette méthode peut être inapplicable, car il est difficile de déterminer avec précision le moment exact où il faut arrêter l'apprentissage puisque les performances sur la base de validation ne se dégradent pas nettement.

On préfère donc utiliser les méthodes de régularisation, d'autant que [Sjöberg, 1994] a montré que l'arrêt prématuré était identique à un terme de pénalisation dans la fonction de coût.



### 2.7.2.2. Weight decay

Lorsque les poids du réseau sont grands en valeur absolue, les sigmoïdes des neurones cachés sont saturées, si bien que les fonctions modélisées peuvent avoir des variations brusques. Pour obtenir des fonctions régulières, il faut travailler avec la partie linéaire des sigmoïdes, ce qui implique d'avoir des poids dont la valeur absolue est faible.

La méthode de régularisation du « *weight decay* » limite la valeur absolue des poids en utilisant :

$$\Omega = \frac{1}{2} \sum w_i^2 \quad (2.26)$$

L'apprentissage s'effectue en minimisant :

$$E = E + \frac{\alpha}{2} \sum w_i^2 \quad (2.27)$$

$\alpha$  est un *hyper paramètre* qui détermine l'importance relative des deux termes dans la nouvelle fonction de coût (compromis entre biais variance). Si  $\alpha$  est trop grand, les poids

$\theta = (\theta_1, \theta_2, \dots, \theta_p)$ , l'équation, s'écrit : [20]

$$p(\theta | D) = \frac{p(D | \theta) \cdot p(\theta)}{p(D)} \quad (2.28)$$

Où :

$D = (d_1, d_2, \dots, d_n)$  Tendait rapidement vers zéro, le modèle ne tient plus compte des données. Si  $\alpha$  est trop petit, le terme de régularisation perd de son importance et le réseau de neurones peut donc être en surapprentissage. Dans le cas intermédiaire, les poids après l'apprentissage ont des valeurs modérées.

Cette méthode est appelée « *Ridge regression* » dans le cas de modèles linéaire par rapport aux paramètres [Saporta, 1990].

## 2.8. Approche neuronale bayésienne

### 2.8.1. Théorème de Bayes [19]

Le théorème de Bayes s'écrit:

$$prob(x | y) = \frac{prob(y | x) \times prob(x)}{prob(y)} \quad (2.29)$$

Les techniques bayésiennes peuvent être utilisées soit pour des problèmes de sélection de modèles, soit pour des problèmes d'estimation de paramètres appliqué à notre modèle défini par  $D$  Représente les données observées.

$p(\theta|D)$ , la fonction densité de probabilité (fdp) a posteriori appelée également *posterior*. Le maximum du posterior correspond aux meilleures estimations des paramètres  $\theta$  après avoir observé les nouvelles données  $D$ . La largeur de la fdp apporte également une indication sur l'incertitude associée à ces estimations.

$p(\theta|D)$ , la fonction de vraisemblance. Elle exprime l'influence des données,  $D$ , sur les paramètres  $\theta$ .  $p(D)$ , l'évidence. Elle permet de normaliser le posterior.

$p(\theta)$ , la fdp a priori, appelée prior. Elle représente la connaissance que l'on peut avoir sur les paramètres  $\theta$ , avant d'avoir observé les nouvelles données.

Dans le contexte bayésien, une probabilité n'est pas définie comme une fréquence d'occurrence, elle représente un degré de croyance, une connaissance a priori. L'inférence Bayésienne consiste à mettre à jour cette connaissance à priori, à partir de données observées. Ainsi, l'estimation des paramètres  $\theta$  s'appuie sur des données à la fois objectives et subjectives.

### 2.8.2. Approche probabiliste appliquée à l'apprentissage neuronal [25]

D'après Mac Kay, les techniques Bayésiennes et la modélisation neuronale sont deux outils complémentaires. L'apprentissage Bayésien consiste à déterminer les valeurs optimales des paramètres neuronaux dont les poids et biais, en leur attribuant des fdp. Ainsi, au lieu d'un jeu unique supposé optimal, de paramètres obtenu par les méthodes classiques à travers la minimisation d'une fonction d'erreur, la méthode Bayésienne fournit et associe une fonction de distribution de probabilité aux poids du RNA. L'estimation des paramètres neuronaux par inférence bayésienne consiste à déterminer la distribution de probabilité à posteriori des paramètres,  $p(w|D)$ , à partir de la distribution de probabilité a priori,  $p(w)$  et de la fonction de vraisemblance  $p(D|w)$ .

### 2.8.3. Distribution de probabilité a priori [21] [21]

Nous avons choisi d'exprimer la distribution de probabilité a priori et la fonction de vraisemblance comme des distributions gaussiennes. En effet, ce choix simplifie l'analyse et permet la dérivation analytique du posterior. De plus, le choix d'un prior gaussien apporte une interprétation probabiliste au terme de régularisation  $E_w$  (terme régulateur) qui peut être interprété comme étant l'opposé du logarithme du prior :

$$p(w|\alpha) = \frac{1}{Z_w(\alpha)} e^{(-\alpha \cdot E_w)} \quad (2.30)$$

Comme nous avons choisi un prior gaussien, le facteur de normalisation  $Z_w(\alpha)$  s'écrit :

$$Z_w(\alpha) = \int e^{(-\alpha E_w)} dw = \left( \frac{2\pi}{\alpha} \right)^{p/2} \quad (2.31)$$

Avec  $\int p(w)dw = 1$  et  $p$  le nombre de données

$\alpha$  est appelé hyper paramètre car il contrôle la distribution des paramètres neuronaux. Il représente l'inverse de la variance sur l'ensemble des poids et biais.

#### 2.8.4. Fonction de vraisemblance [26] [27]

L'expression de la fonction de vraisemblance est liée à la définition du modèle du bruit sur les données. Considérons des données d'apprentissage de la forme  $D = \{x^i, t^i\}$ . Le but de l'apprentissage neuronal est de déterminer la relation entre  $x^i$  et  $t^i$ . Puisqu'il y a des incertitudes sur les données, la relation  $R$  entre  $x^i$  et  $t^i$  est de la forme :

$$t^i = R(x^i) + \varepsilon^i \quad (2.32)$$

Où  $\varepsilon^i$  représente le bruit.

Dans notre cas, ce bruit est représenté par une fdp gaussienne, de moyenne nulle et d'écart type  $\sigma$ . En posant par  $\sigma^2 = \frac{1}{\beta}$  et  $y(x, w)$ , la sortie du réseau, chacune des données expérimentales  $t^i$  peut être exprimée à partir de la fdp suivante :

$$p(t^i | x^i, w, \beta) = \sqrt{\frac{\beta}{2\pi}} \exp\left(-\frac{\beta}{2} \{y(x^i, w) - t^i\}^2\right) \quad (2.33)$$

De plus, nous supposons que les réalisations du bruit  $\varepsilon^i$  sont indépendantes, alors la fonction de vraisemblance  $p(D|w, \beta)$  s'écrit :

$$p(D | w, \beta) = \prod_{k=1}^N p(t^k | x^k, w, \beta) = \frac{1}{Z_D(\beta)} \exp(-\beta E_D) \quad (2.34)$$

En indiquant par :

$E_D$ , la fonction d'erreur

$\beta$ , un autre hyper paramètre lié à la sortie.

Le choix d'un modèle gaussien pour représenter le bruit permet de calculer analytiquement le facteur de normalisation  $Z_D(\beta)$  :

$$Z_D(\beta) = \int e^{(-\beta \cdot E_D)} dD = \left( \frac{2\pi}{\beta} \right)^{N/2} \quad (2.35)$$

Le choix d'une fonction de vraisemblance gaussienne permet également une interprétation probabiliste de la fonction d'erreur. En effet,  $E_D$  peut être interprété comme étant l'opposé du logarithme de la fonction de vraisemblance :  $E_D \propto -\ln(p(D/w, \beta))$

### 2.8.5. Distribution de probabilité a posteriori [26] [27]

Une fois le prior et la fonction de vraisemblance définie, le posterior s'écrit :

$$p(w/\alpha, \beta, D) = \frac{p(D/w, \beta) \cdot p(w/\alpha)}{p(D/\alpha, \beta)} \quad (2.36)$$

Son facteur de normalisation est donné par :

$$Z_J(\alpha, \beta) = \int e^{-J(w)} dw \quad (2.37)$$

Lors de l'apprentissage neuronal, l'évaluation de  $Z_J(\alpha, \beta)$  ne peut se faire analytiquement. Mais pour que cette intégrale puisse être déterminée analytiquement, Mac Kay a proposé une approximation gaussienne du posterior.

Les valeurs optimales des poids et biais correspondent au maximum de la fdp du posterior. Ainsi, la fonction régularisée  $S(w)$  à minimiser est la suivante :

$$J(w) = \frac{\beta}{2} \sum_{i=1}^N \{y(x^i, w) - t^i\}^2 + \frac{\alpha}{2} \sum_{i=1}^p w_i^2 \quad (2.38)$$

Excepté les facteurs multiplicatifs,  $J(w)$  représente la somme des erreurs quadratiques avec un terme de régularisation qui dépend du coefficient  $\lambda = \frac{\alpha}{\beta}$ .

Dans le cas où le nombre de données,  $N$ , est important,  $J(w)$  converge vers  $N/2$ . Pratiquement, la quantité à minimiser est  $J'(w)$  qui est égale à  $(2/\beta) J(w)$ .

Le premier terme de  $J(w)$  est une fonction croissante du nombre  $N$  de données alors que le second ne l'est pas.

Un des avantages d'utilisation des méthodes bayésiennes est la modélisation d'un système à faible nombre de données.

### 2.8.6. Approximation gaussienne du posterior [26] [27]

L'inférence bayésienne appliquée à l'estimation des paramètres neuronaux nécessite d'évaluer des intégrales dans l'espace des poids. Mais le facteur de normalisation  $Z_J(\alpha, \beta)$  ne peut

être évalué analytiquement, d'où une approximation gaussienne du posterior est nécessaire. Cette approximation est obtenue à partir du développement du second ordre de Taylor de  $J(w)$ , autour de sa valeur minimale  $w_{MP}$  (*Most Probable*) :

$$J(w) = J(w_{MP}) + \frac{1}{2} (w - w_{MP})^T A (w - w_{MP}) \quad (2.39)$$

Où  $A$  est le Hessian de la fonction erreur régularisée.

Le posterior gaussien s'écrit alors :

$$P(w | \alpha, \beta, D) = \frac{1}{Z_J^*(\alpha, \beta)} \exp(-J(w_{MP}) - \frac{1}{2} \Delta w^T \cdot A \cdot \Delta w) \quad (2.40)$$

Sous cette approximation, le facteur de normalisation  $Z_J^*(\alpha, \beta)$  peut être évalué analytiquement par l'expression :

$$Z_J^*(\alpha, \beta) = \exp(-J(w_{MP})) (2\pi)^{p/2} |A|^{-1/2} \quad (2.41)$$

Où  $|A|$  est le déterminant de la matrice  $A$ .

### 2.8.7. Détermination des hyper paramètres $\alpha$ et $\beta$ [26] [27]

Pour déterminer  $P(w/D)$ , nous devons évaluer les valeurs les plus probables des hyper paramètres  $\alpha$  et  $\beta$  qui contrôlent la complexité du réseau de neurones à travers des densités de probabilités. Partant de la densité à posteriori et en appliquant la probabilité jointe pour faire apparaître les deux hyper paramètres, nous avons :

$$P(w/D) = \iint p(w, \alpha, \beta/D) \cdot d\alpha \cdot d\beta = \iint p(w/\alpha, \beta, D) \cdot p(\alpha, \beta/D) \cdot d\alpha \cdot d\beta \quad (2.42)$$

Cette intégrale double, étant difficilement calculable, nous proposons une approche simplificatrice pour la détermination des hyper paramètres, appelée "*approche Evidence*". En effet, le théorème de Bayes permet de déterminer les valeurs optimales des hyper paramètres, à partir des données  $D$  :

$$p(\alpha, \beta/D) = \frac{p(D/\alpha, \beta) \cdot p(\alpha, \beta)}{p(D)} \quad (2.43)$$

Où :  $p(\alpha, \beta)$ , la fdp a priori sur les paramètres, appelée *hyperprior*

$p(D|\alpha, \beta)$ , la fonction de vraisemblance, appelée également terme *Evidence*. Comme nous n'avons pas, a priori, les valeurs des hyper paramètres, nous leur attribuons, à chacune, une distribution uniforme à  $p(\alpha, \beta)$ . Par ailleurs,  $p(D)$  ne dépend pas des hyperparamètres, maximiser le posterior  $p(\alpha, \beta/D)$  revient à maximiser le terme *Evidence*  $p(D|\alpha, \beta)$ , exprimé par :

$$p(D|\alpha, \beta) = \int p(D|w, \alpha, \beta)p(w|\alpha, \beta)dw = \int p(D|w, \beta)p(w|\alpha)dw \quad (2.44)$$

Il vient :

$$p(D|\alpha, \beta) = \frac{1}{Z_D(\beta)Z_W(\alpha)} \int \exp(-J(w))dw \quad (2.45)$$

A partir de l'approximation gaussienne du posterior des poids, le terme *Evidence* s'écrit :

$$p(D|\alpha, \beta) = \frac{Z_S^*(\alpha, \beta)}{Z_D(\beta)Z_W(\alpha)} = \frac{(2\pi)^{\frac{p}{2}} \times |A|^{-\frac{1}{2}} \times \exp\{-J(w_{MP})\}}{\left(\frac{2\pi}{\beta}\right)^{\frac{N}{2}} \cdot \left(\frac{2\pi}{\alpha}\right)^{\frac{p}{2}}} \quad (2.46)$$

Les valeurs optimales des hyperparamètres  $\alpha_{MP}$  et  $\beta_{MP}$  correspondent au maximum de ce terme *Evidence*. Cette fdp peut également être interprétée comme une fonction erreur à minimiser, en prenant le logarithme de l'équation (2.39) :

$$\ln p(D|\alpha, \beta) = -\alpha E_W^{MP} - \beta E_D^{MP} - \frac{1}{2} \ln |A| + \frac{p}{2} \ln \alpha + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi) \quad (2.47)$$

Les valeurs optimales de ces hyperparamètres sont obtenues en dérivant le terme *Evidence* par rapport à ces hyperparamètres :

$$\alpha_{MP} = \frac{\gamma}{2 \cdot E_W^{MP}} \quad ; \quad \beta_{MP} = \frac{N - \gamma}{2 \cdot E_D^{MP}} \quad (2.48)$$

Où :

$$\gamma = \sum_{k=1}^p \left( \frac{\lambda_k}{\lambda_k + \alpha} \right) \quad (2.49)$$

Si  $\lambda_k$  sont les valeurs propres de la matrice Hessienne A.

$\gamma$ , le nombre effectif de paramètres neuronaux  $(\alpha, \lambda_k)$ , indiquant réellement la relation entre les entrées et la sortie du modèle.

A partir du nombre effectif des poids et biais, il est possible d'avoir le nombre de neurones cachés correspondants :

$$n_c = \frac{\gamma - 1}{n_i + 2} \quad (2.50)$$

Où :

$n_i$  est le nombre d'entrées

$n_c$  est le nombre de neurones cachés

Nous voyons ainsi qu'aucune base de validation n'est requise pour la détermination des valeurs optimales de  $\alpha$  et  $\beta$  et toutes les données sont disponibles pour déterminer les valeurs

des poids et les valeurs des hyper paramètres. De plus, l'approche *Evidence* permet de calculer le nombre effectif de paramètres neuronaux, lors de l'apprentissage même du RNA.

### 2.8.8. Critères de performances pour les modèles

Comme nous l'avons mentionné précédemment, il existe de très nombreux critères. Dans notre travail, nous nous sommes ramenés à des critères numériques faisant appel à deux (2) types de prise en compte de l'erreur du model qui sont :

- La racine carrée de l'erreur quadratique (RMSE)

Les critères en moindres carrée se basent sur une fonction suggérée par la régression linéaire ou la sommes des erreurs quadratiques du modèle définie par [28] :

$$RMSE = \sqrt{\frac{1}{N} \left( \sum_i^N (Q_{cal,i} - Q_{mes,i})^2 \right)} \quad (2.51)$$

Où  $Q_{cal, i}$  et  $Q_{mes, i}$  sont respectivement les débits calculés et les mesures au pas de temps  $i$

$N$ , le nombre total de données de la période d'étude.

- La moyenne de l'erreur absolue(MAPE)

L'erreur moyenne absolue MAPE d'un modèle peut être définie par :

$$MAPE = \frac{1}{N} \sum_{i=1}^n |Q_{cal,i} - Q_{mes,i}| \quad (2.52)$$

Elle représente la déviation absolue du débit simulée par rapport à celle mesurée en moyenne à chaque pas de temps. Ce critère peut être intéressant dans le contexte de la prévision de débit, pour être aussi proche que possible de la valeur observée à chaque pas de temps.

Ces critiques mesurent le degré d'adéquation entre les valeurs de la variable simulée par le modèle et de la variable mesurée.

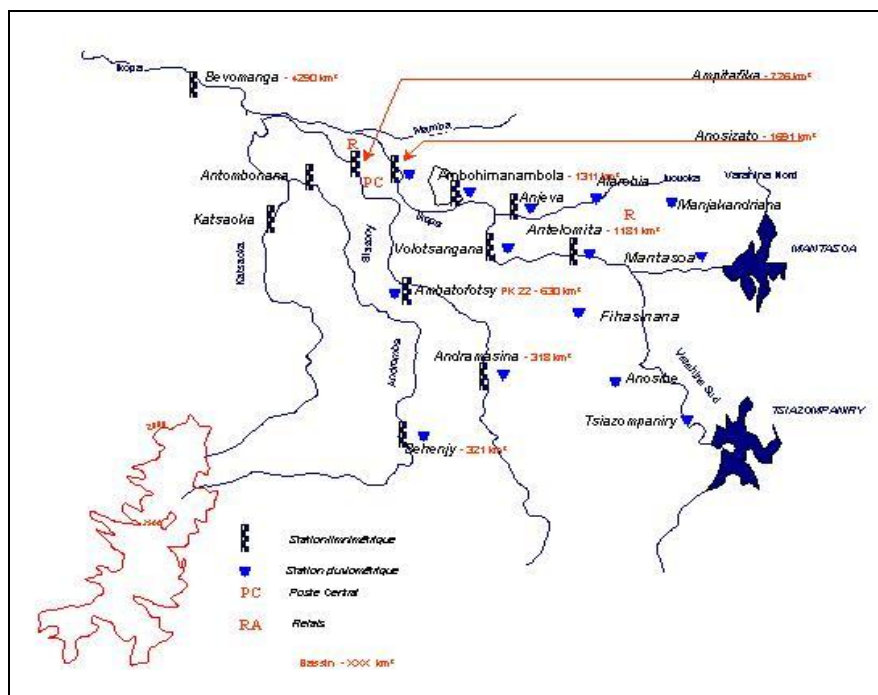
CHAPITRE 3 :  
MODELISATION DU COMPORTEMENT DU DEBIT A  
PARTIR DES RESEAUX DE NEURONES ARTIFICIELS



### 3.1. Bases de données utilisées

La disponibilité des données pour les simulations est assez rare ou inexistant. L'Autorité pour la Protection contre les Inondations de la Plaine d'Antananarivo (APIPA) a pour mission la gestion des ouvrages et équipements assurant la défense de la plaine d'Antananarivo contre les inondations. Elle dispose d'un réseau d'annonces de crues qui est composé de 18 stations de mesure automatique de pluie et/ou de débit dans le bassin versant de l'Ikopa en amont de Bevomanga (4 300 km<sup>2</sup>).

Ces stations sont reliées par radio à un poste central d'acquisition et de traitement des données au bureau de l'APIPA à Anosizato. Ce poste central collecte les données des 18 stations toutes les heures et les utilise pour la prévision des crues des rivières Ikopa et Sisaony pendant leurs traversées de l'agglomération. La figure 3.1 représente la carte de localisation des stations de mesure des hauteurs d'eau sur les affluents de l'Ikopa



Source : APIPA

**Figure 3. 1: service d'annonce de crues dans le bassin de l'Ikopa en amont de Bevomanga (4300 km<sup>2</sup>)**

Les données obtenues accusent parfois des interruptions pour beaucoup de stations. Dans notre travail, nous avons considéré trois stations pour mener les simulations. Ces dernières sont : Ambohimambola, Anosizato et Anjeva. Les données horaires obtenues ont été

converties en données journalières. Nous avons pris seulement le débit pour la suite de notre travail.

### **3.2. Réseaux de neurones statiques**

Dans ce chapitre, nous allons appliquer la théorie des RNA à la prévision du débit de quelques rivières et de tester la performance des réseaux. Tout au long de ce chapitre, les codes de calculs ont été développé sous Matlab dans la boite à outils «Netlab» [29]. Nous avons proposé deux modèles basés sur l'architecture statique.

#### **3.2. 1. Présentation du modèle neuronal classique**

Le premier modèle est du type classique basé sur le perceptron, Les données utilisées représentent l'année 2005 pour la phase d'apprentissage et de l'année 2006 de ces mêmes stations pour la phase de test.

#### **3.2.2. Organigramme du réseau classique**

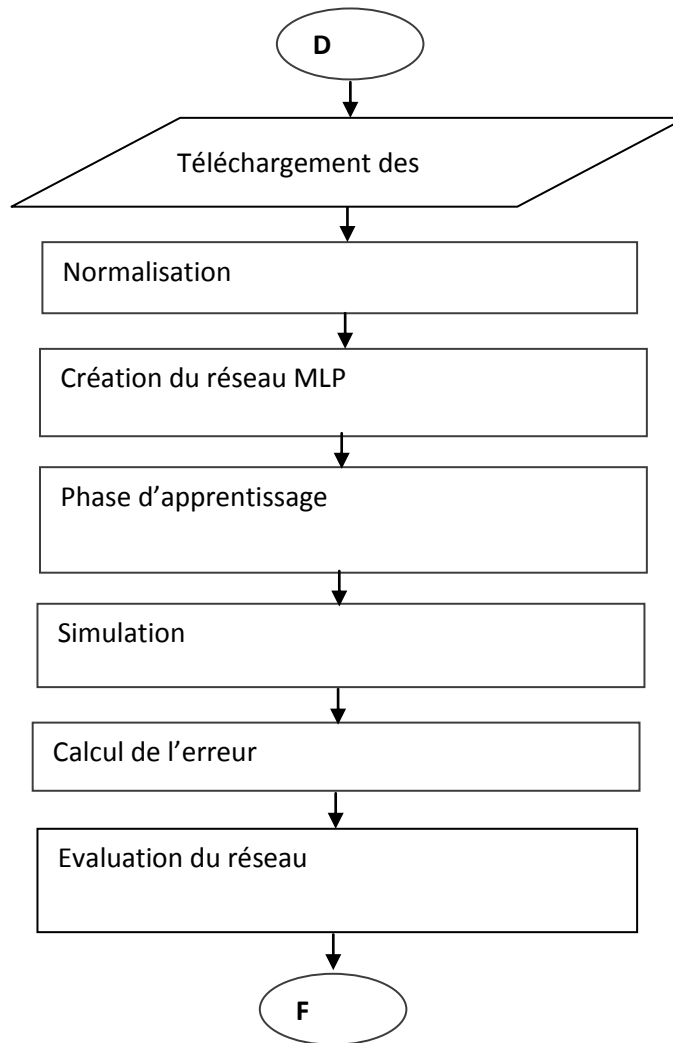
La figure 3.2 représente l'organigramme que nous avons suivi lors de l'apprentissage et de multicouche (PMC) dont le nombre initial de neurones est estimé par la formule [30]:

$$Hidden\_neuron = \frac{1}{2} ( (Inputs + Outputs) + \sqrt{Sample\_training} ) \quad (3.1)$$

Nous avons pris les bases de données d'une manière indépendante dans toute l'année pour les différentes stations.

- La base d'apprentissage représentant les données de l'année 2005 pour les trois stations considérées.

L'évaluation du modèle.



**Figure 3.2: organigramme du réseau classique**

Le principe de ce modèle était de faire la prévision du débit jour par jour pendant toute l'année, la sortie de tous les modèles étant le débit de la rivière au jour  $j$ .

Ses variables d'entrée utilisées, dans ce cas, sont consignées dans le tableau 3.1:

**Tableau 3.1: les entrées du modèle**

N°	1	2	3	4
Paramètre	$Q(j-4)$	$Q(j-3)$	$Q(j-2)$	$Q(j-1)$

### 3.2.3. Résultats avec le modèle classique

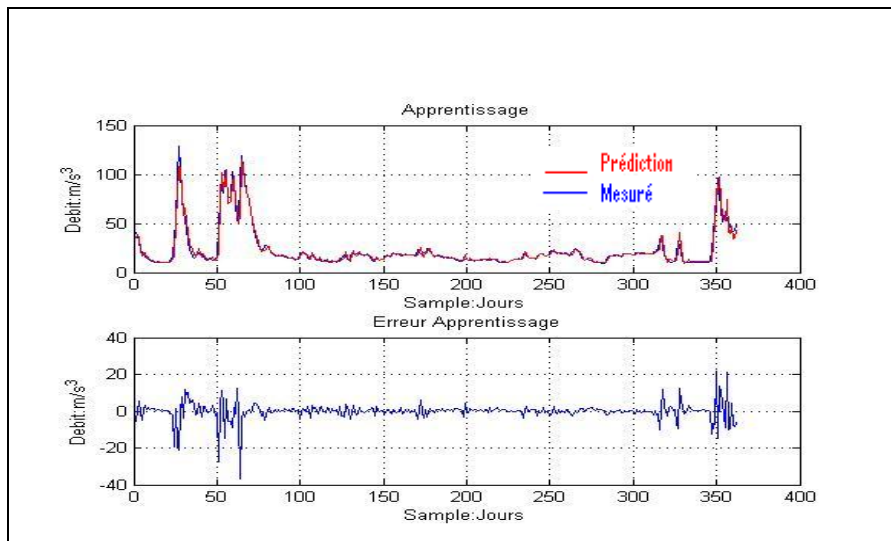
Nous présentons les résultats, soit à travers des tableaux ; soit à travers des figures où l'axe de  $x$  indique le jour et sur l'axe de  $y$ , est portée les sorties (sortie du modèle et sortie souhaitée).

**Tableau 3. 2: Evaluation des critères de performance pour le modèle classique**

PMC avec une seule couche cachée de 10 neurones				
	RMSE Training (m <sup>3</sup> /s)	RMSE Test (m <sup>3</sup> /s)	MAPE Training (%)	MAPE Test (%)
Anjeva	0,1949	0,7405	0,5052	0,3775
Anosizato	0,1620	0,3629	0,0915	0,0710
Ambohimanambola	0,2302	0.5207	0.1799	0.0852

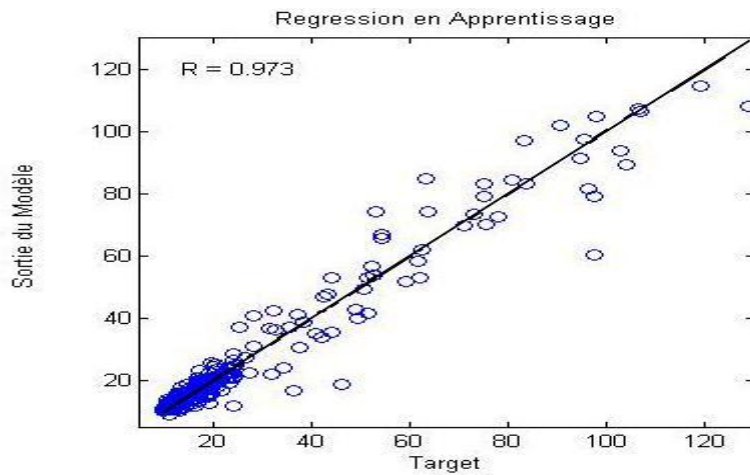
**3.2.3.1. La station d’Anosizato**

Les résultats montrent qu’en phase d’apprentissage, le réseau est satisfaisant (voir figure 3.3). Les simulations ont donné comme RMSE enregistré pour la station d’Anosizato la valeur de 0,1620m<sup>3</sup>/s et comme MAPE la valeur de 0,92%.



**Figure 3. 3: MLP en phase d’apprentissage**

La figure 3.3 montre aussi l’erreur de l’apprentissage, et cette erreur est confirmée par la considération de la droite de régression de la figure 3.4 qui donne la comparaison entre la sortie du modèle et les points de mesure.



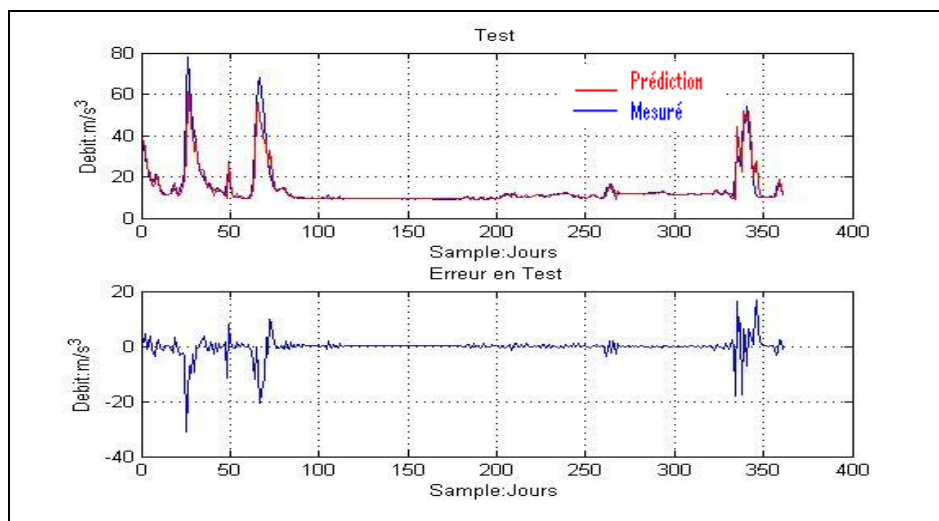
**Figure 3. 4: droite de régression MLP en apprentissage**

La droite régression :

$$A = 0.973(T) + 0.0915 \quad (3.2)$$

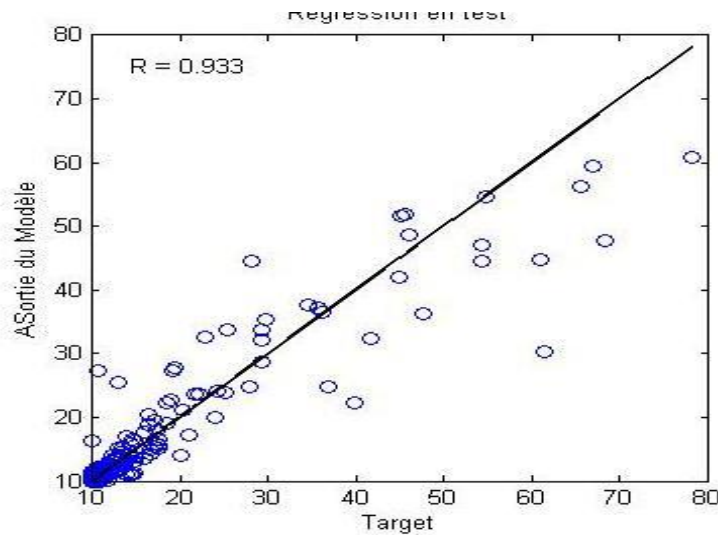
Avec : A et T sont successivement la sortie prédite et la sortie souhaitée.

En considérant la performance du modèle, et en évaluant sur la base de test, il est clair que le résultat est satisfaisant selon le critère RMSE enregistré qui est égal à  $0.3629\text{m}^3/\text{s}$ .



**Figure 3. 5: MLP en phase de test**

La figure 3.5 montre l'erreur de l'apprentissage, et cette erreur est confirmée par la considération de la droite de régression de la figure 3.6 qui donne la comparaison entre la sortie du modèle et les points de mesure.



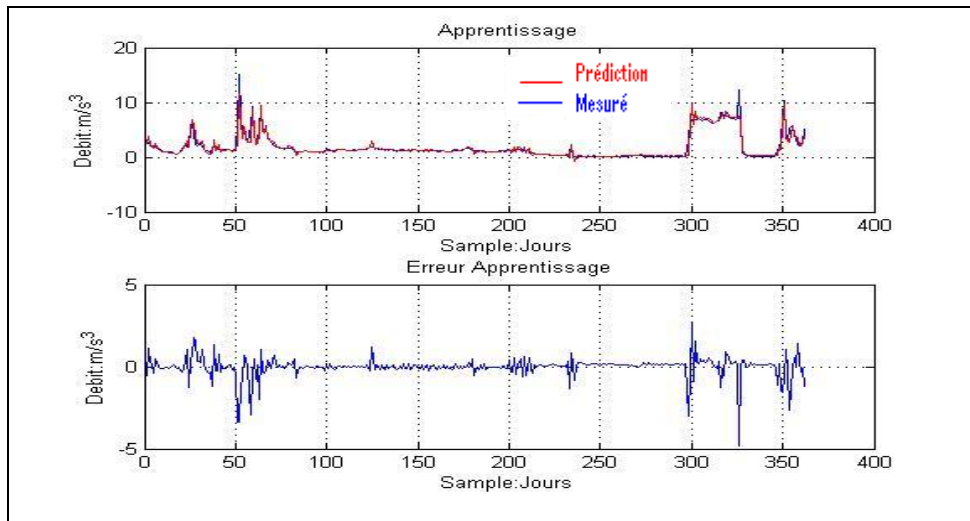
**Figure 3. 6: Droite de régression MLP en phase de test**

La droite régression :

$$A = 0.933(T) + 0.0710 \quad (3.3)$$

### 3.2.3.2. La station d’Anjeva

Les résultats montrent qu’en phase d’apprentissage, le réseau est satisfaisant (voir figure 3.7). Les simulations ont donné comme RMSE enregistré pour la station d’Anjeva la valeur de 0,1949m<sup>3</sup>/s et comme MAPE la valeur de 0 ,50%.



**Figure 3. 7: MLP en phase d’apprentissage**

La figure 3.7 montre aussi l’erreur de l’apprentissage, et cette erreur est confirmée par la considération de la droite de régression de la figure 3.8 qui donne la comparaison entre la sortie du modèle et les points de mesure.

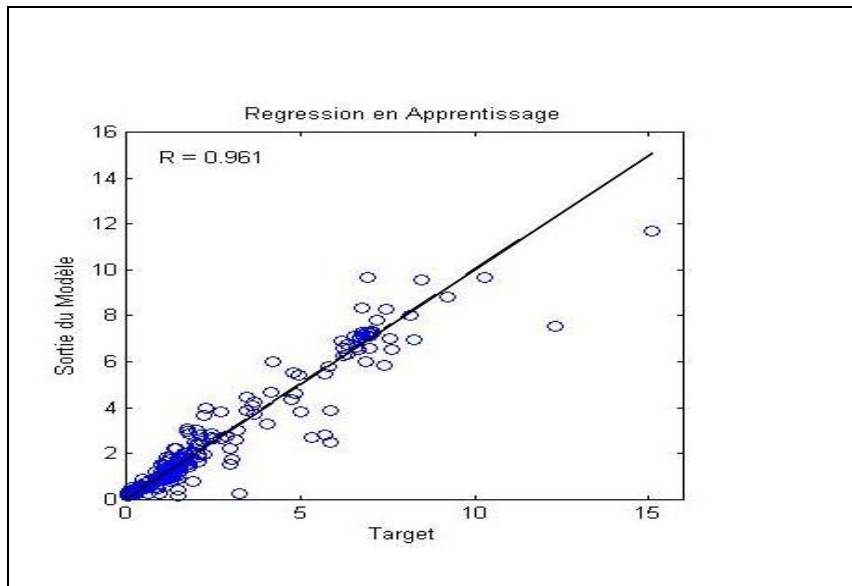


Figure 3. 8: droite de régression MLP en apprentissage

La droite régression :

$$A = 0.961(T) + 0.5052 \quad (3.4)$$

Avec : A et T sont successivement la sortie prédite et la sortie souhaitée.

En considérant la performance du modèle, et en évaluant sur la base de test, il est clair que le résultat est satisfaisant selon le critère RMSE enregistré qui est égal à  $0.7402\text{m}^3/\text{s}$ .

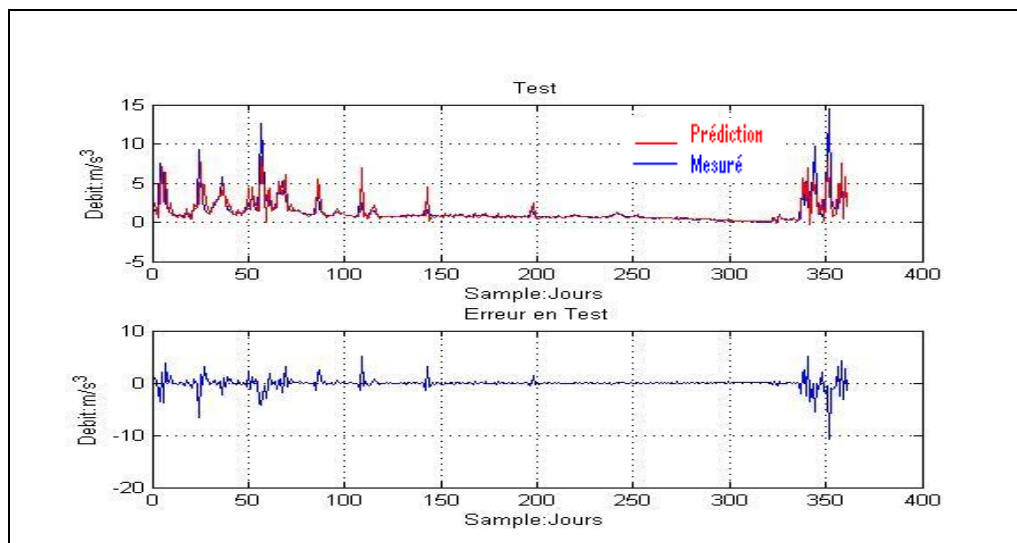
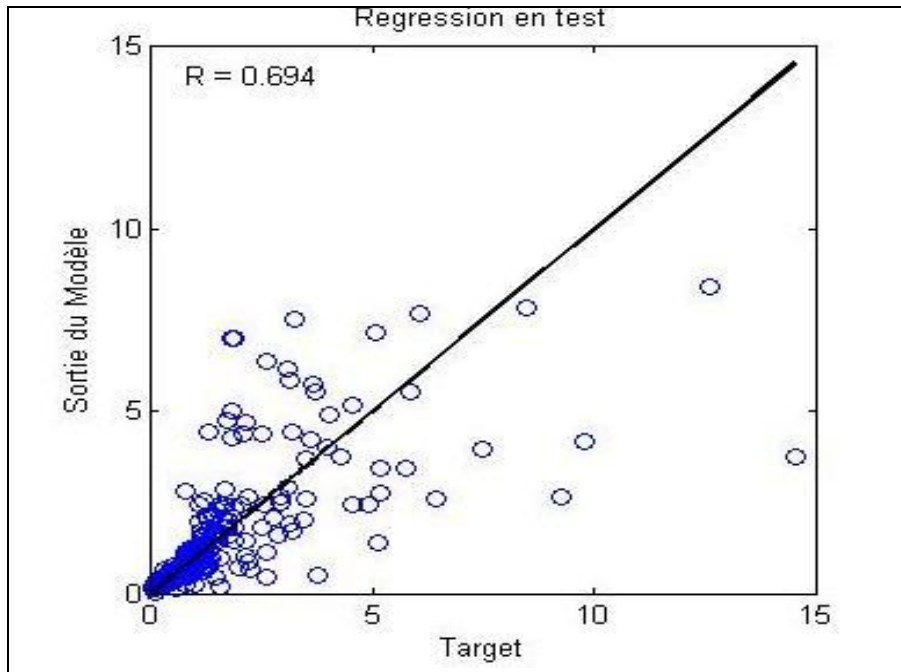


Figure 3. 9: MLP en phase de test

La figure 3.9 montre aussi l'erreur de l'apprentissage, et cette erreur est confirmée par la considération de la droite de régression de la figure 3.10 qui donne la comparaison entre la sortie du modèle et les points de mesure.



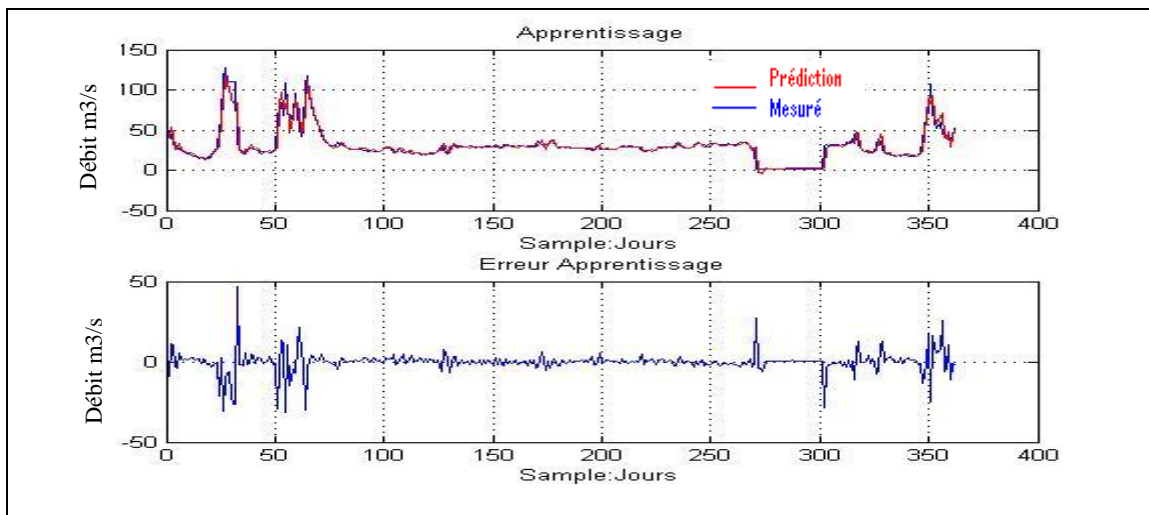
**Figure 3. 10: Droite de régression MLP en phase de test**

La droite régression :

$$A = 0.6940 (T) + 0.3775 \quad (3.5)$$

### 3.2.3.3. La station d'Ambohimambola

Les résultats montrent qu'en phase d'apprentissage, le réseau est satisfaisant (voir figure 3.11). Les simulations ont donné comme RMSE enregistré pour la station d'Anjeva la valeur de 0,2302 m<sup>3</sup>/s et comme MAPE la valeur de 0,18 %.



**Figure 3. 11: MLP en phase d'apprentissage**

L'erreur de l'apprentissage est confirmée par la considération de la droite de régression qui donne la comparaison entre la sortie du modèle et les points de mesure.

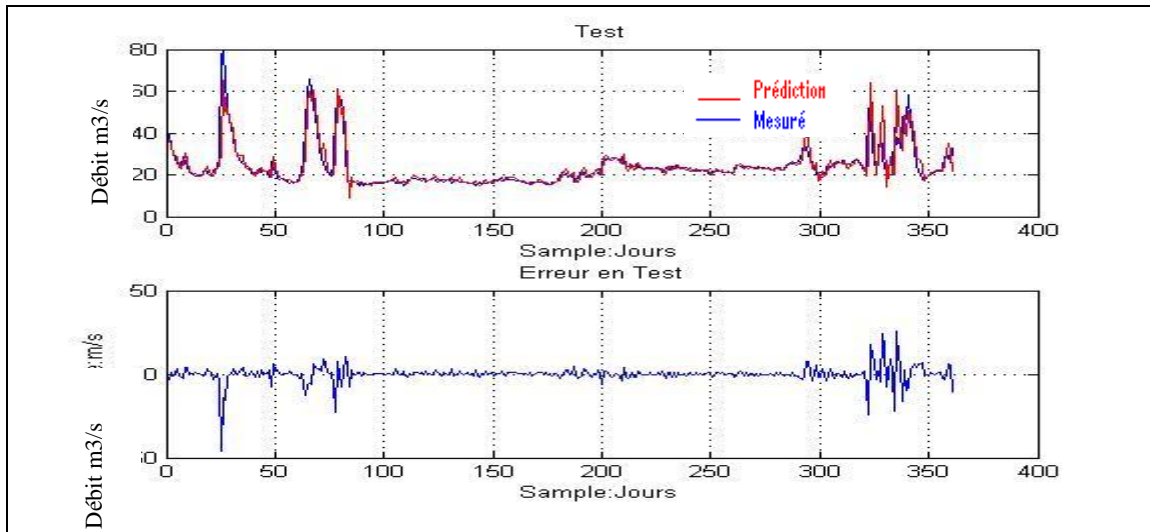
Cette droite est donnée dans la formule suivante :



$$A = 0,9460 (T) + 0,1799 \quad (3.6)$$

Avec : A et T sont successivement la sortie prédite et la sortie souhaitée.

En considérant la performance du modèle, et en évaluant sur la base de test, il est clair que le résultat est satisfaisant selon le critère RMSE enregistré qui est égal à  $0.7402\text{m}^3/\text{s}$ .



**Figure 3. 12: MLP en phase de test**

De même l'erreur de l'apprentissage est confirmée par la considération de la droite de régression qui donne la comparaison entre la sortie du modèle et les points de mesure.

$$A = 0,885 (T) + 0,852 \quad (3.7)$$

### 3.3. Modélisation avec l'approche bayésienne

La construction d'un RNA par l'approche bayésienne, il suffit de déterminer, par un processus itératif, les valeurs optimales des hyperparamètres  $\alpha$  et  $\beta$  et celles des paramètres neuronaux (poids  $w_i$  et biais  $b_0$ ). Pour cette deuxième partie de simulations, nous avons repris le même programme qu'avec le réseau classique. Les hyperparamètres  $\alpha$  et  $\beta$  ont été choisis arbitrairement. Deux cycles ont apparus dans la procédure adoptée :

- Un cycle externe qui a permis l'optimisation des poids et biais
- Un cycle interne grâce auquel l'optimisation des hyper paramètres a été faite.

#### 3.3.1. Organigramme de l'apprentissage bayésien

La figure 3.4 illustre l'organigramme de l'apprentissage bayésien que nous allons suivre. Nous remarquons les deux cycles dans la procédure.

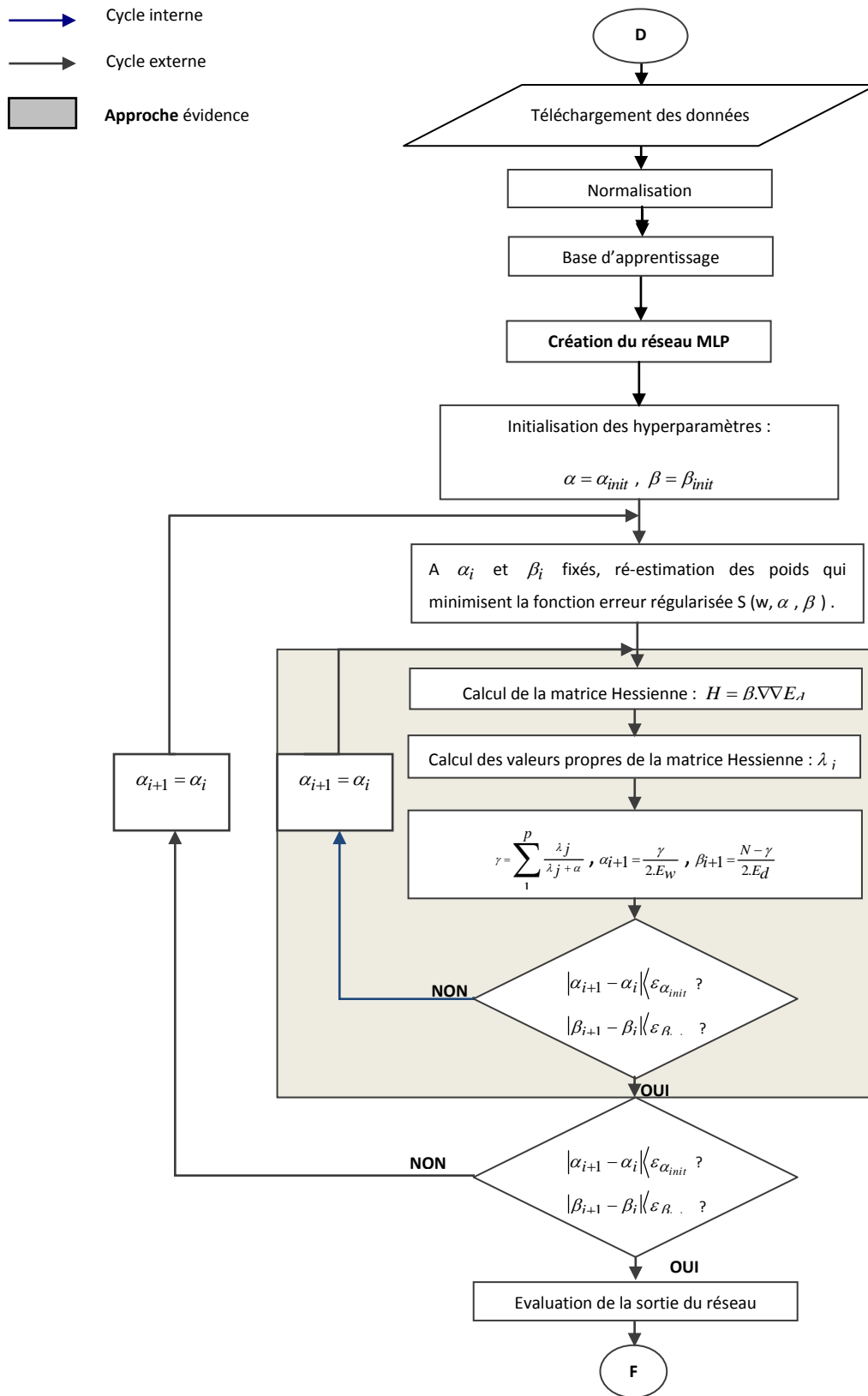


Figure 3. 13: Organigramme suivi lors de l'apprentissage bayésien.

### 3.3.2. Résultats avec l'approche bayésienne

Nous présentons les résultats, soit à travers des tableaux, soit à travers des figures.

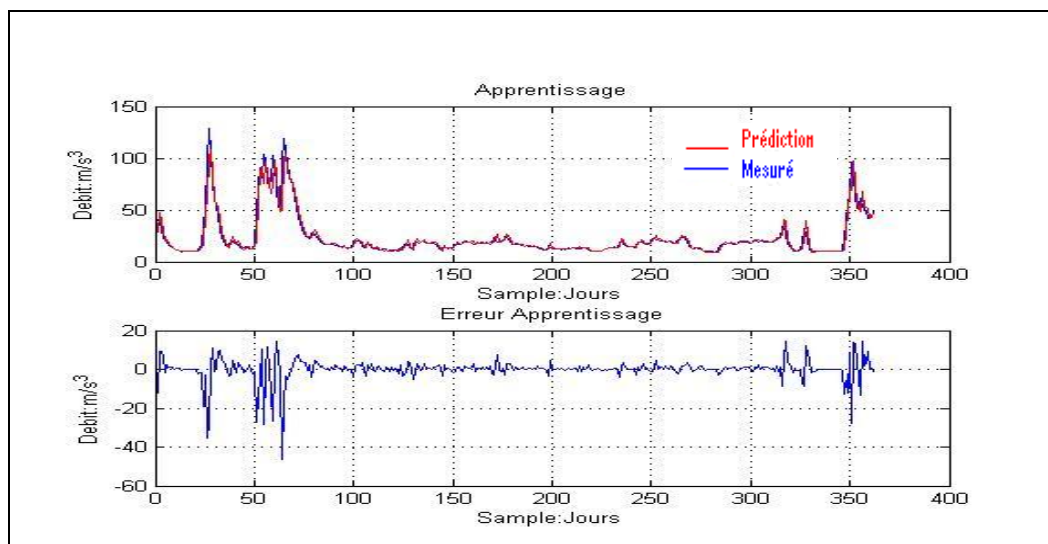
Le tableau N° 3.2 nous montre la performance pour le modèle bayésienne dans les différentes stations concernées.

**Tableau 3. 3: Evaluation des critères de performance pour le modèle bayésienne**

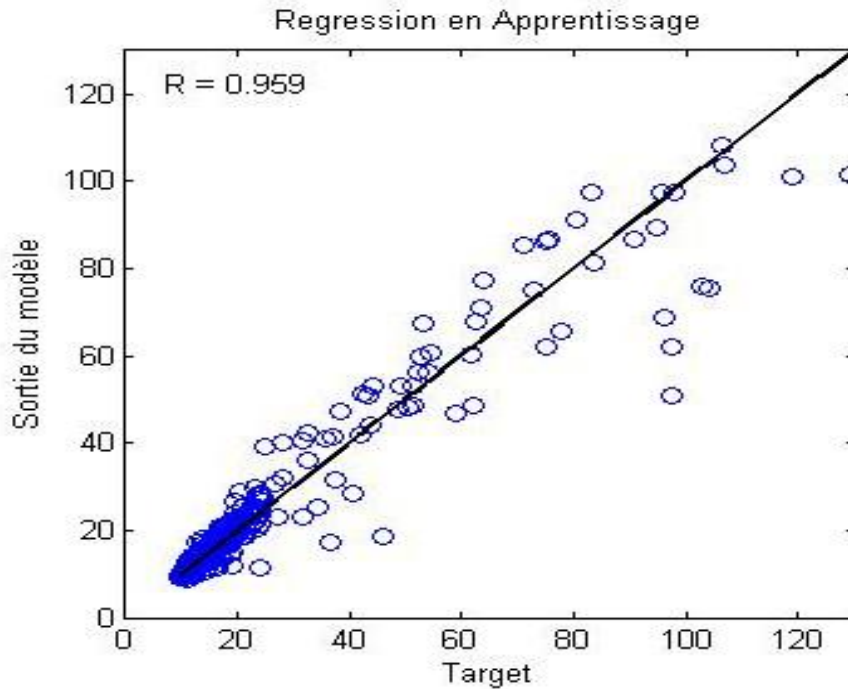
BAYES avec une seule couche cachée de 10 neurones				
	RMSE Training (m <sup>3</sup> /s)	RMSE Test (m <sup>3</sup> /s)	MAPE Training (%)	MAPE Test (%)
Anjeva	0,2316	0,6109	0,2965	0,2928
Anosizato	0,2015	0,3243	0,0914	0,0606
Ambohimanambola	0,2694	0.4759	0.2207	0.0757

### 3.3.2. La station d'Anosizato

En évaluant sa performance, le modèle n'enregistre que 0,2015 m<sup>3</sup>/s pour le critère RMSE et comme MAPE 0.9 % pendant l'apprentissage. Ce comportement est appuyé par les figures 3.14 et 3.15 qui confirment la très bonne approximation du modèle en phase d'apprentissage.



**Figure 3. 14: Approche bayésienne en phase d'apprentissage**

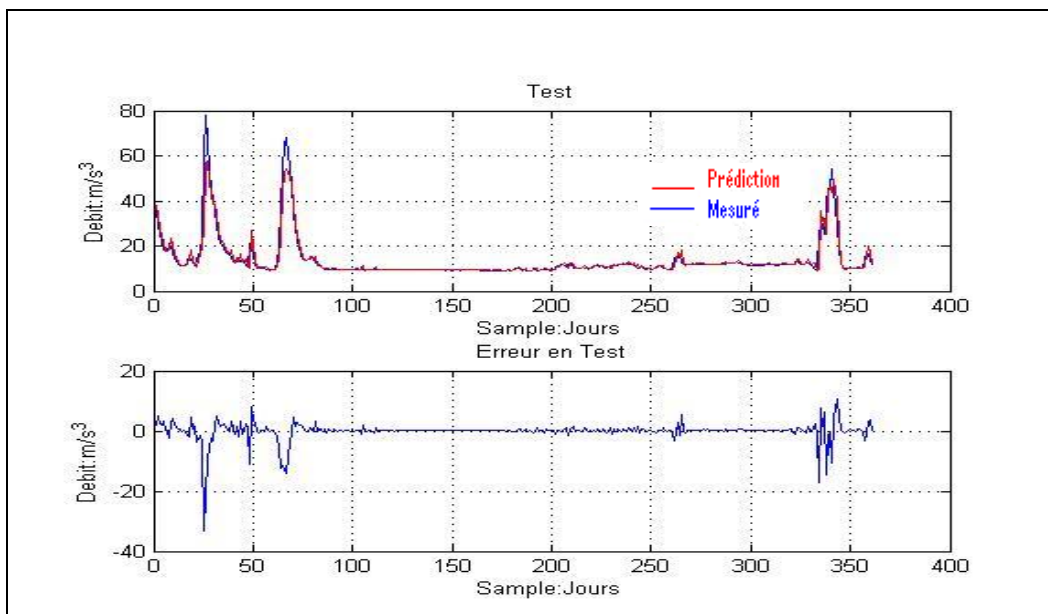


**Figure 3. 14: Droite de régression approche bayésienne en apprentissage**

Droite de régression:

$$A = 0.98 (T) + ( 0.0877) \quad (3.8)$$

Concernant la performance du modèle en phase de test, nous avons trouvé comme critère RMSE la valeur de  $0,3243\text{m}^3/\text{s}$ , ce qui confirme la bonne approximation du prédicateur par rapport aux données mesurées. Les figures 3.16 et 3.17 nous décrivent la phase de test du modèle.



**Figure 3. 15: Approche bayésienne en phase test**

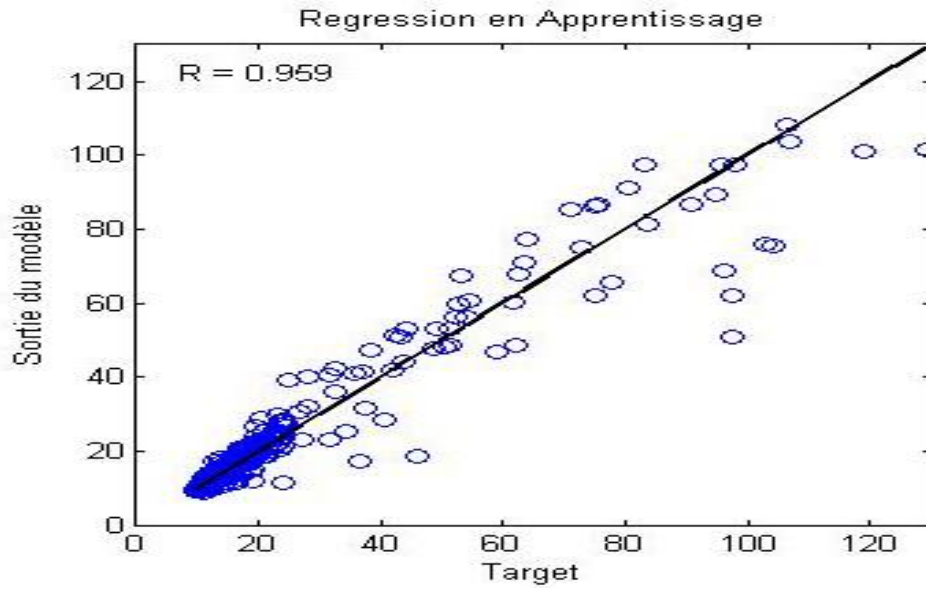


Figure 3. 16: Droite de régression approche bayésienne en phase test

Droite de régression:

$$A = 0,959 (T) +( 0,0606) \tag{3.9}$$

### 3.3.4. La station d’Anjeva

De même, en évaluant sa performance, le modèle n’enregistre que 0,2316 m<sup>3</sup>/s pour le critère RMSE pendant l’apprentissage. Ce comportement est appuyé par les figures 3.18 et 3.19 qui confirment la très bonne approximation du modèle en phase d’Apprentissage.

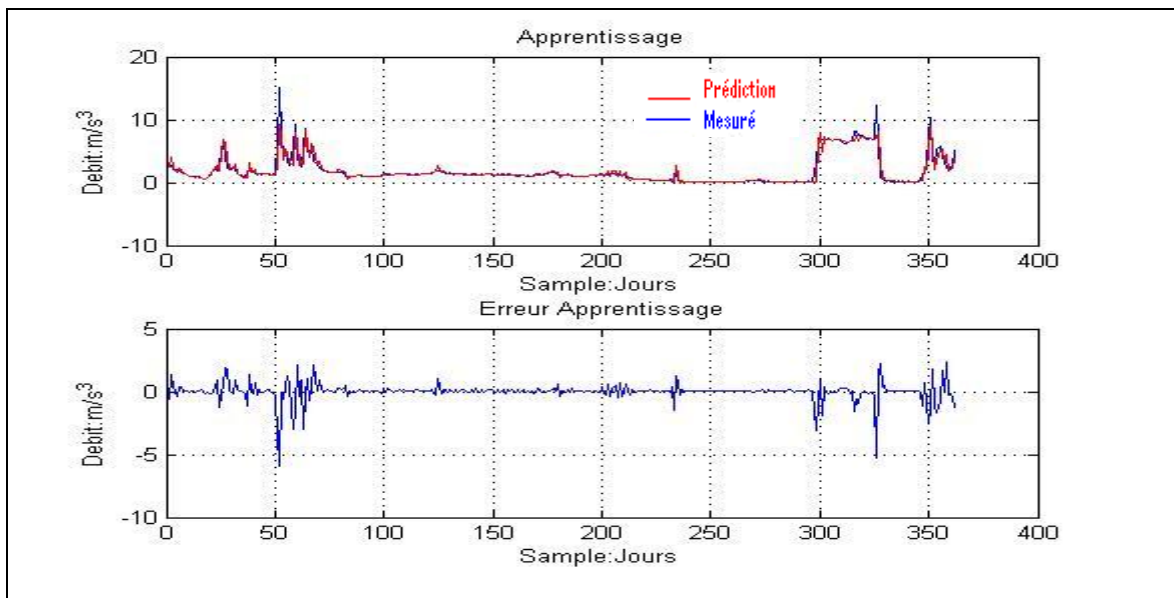
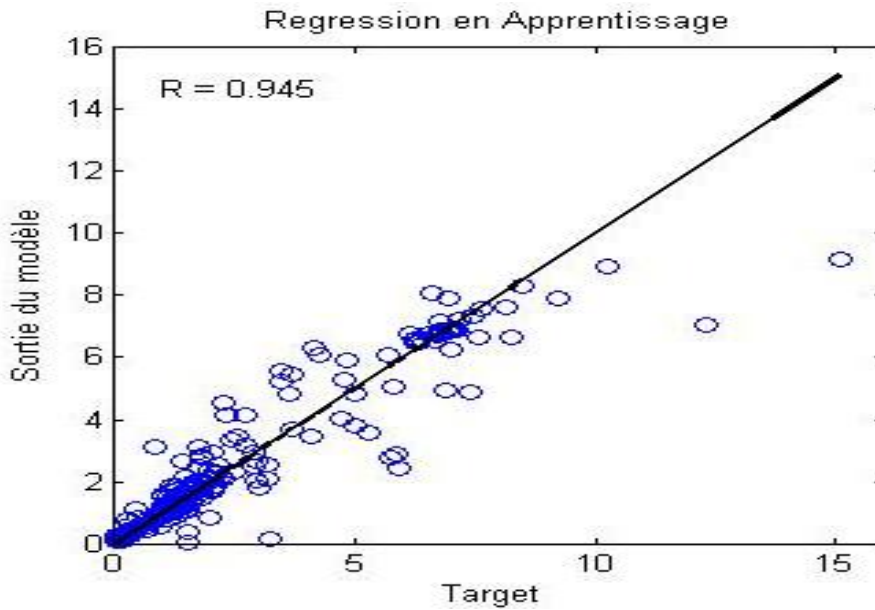


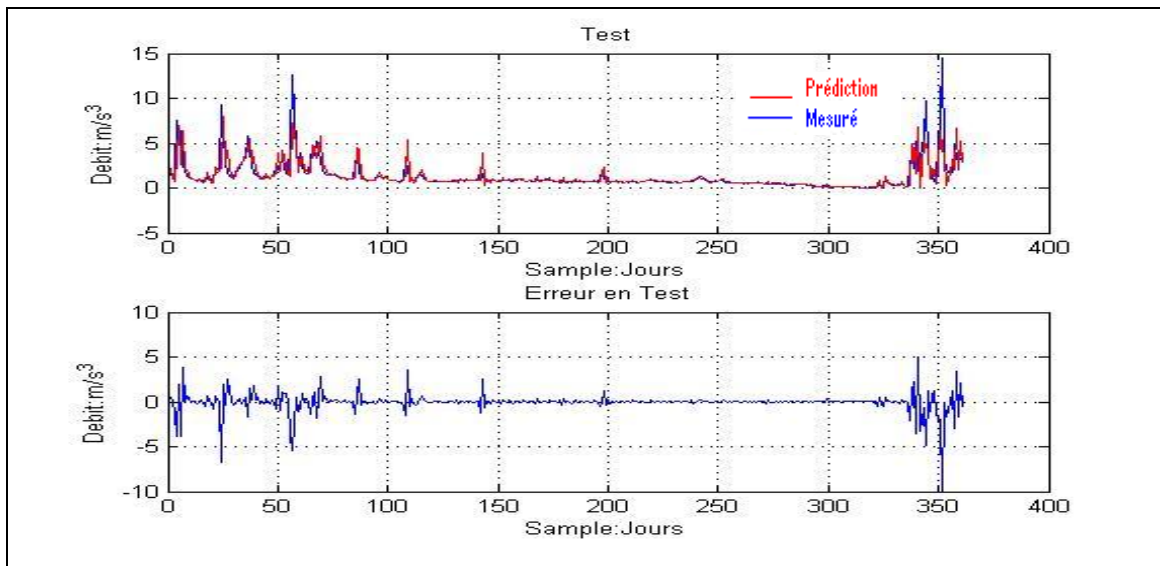
Figure 3. 17: Approche bayésienne en phase d’apprentissage



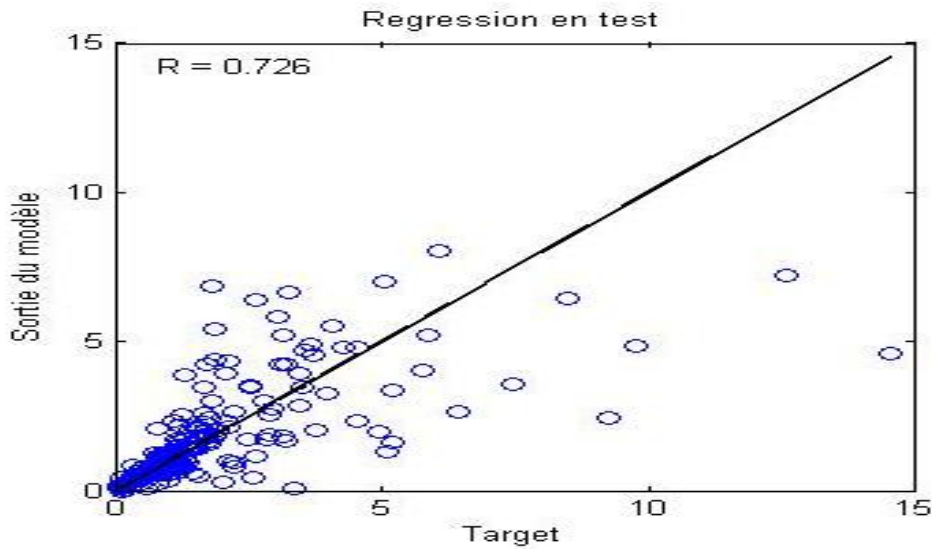
**Figure 3. 18: Droite de régression approche bayesienne en apprentissage**  
Droite de régression:

$$A = 0,945 (T) +( 0,2601) \tag{3.10}$$

Concernant la performance du modèle en phase de test, nous avons trouvé comme critère RMSE la valeur de 0,6109 m<sup>3</sup>/s, ce qui confirme la bonne approximation du prédicateur par rapport aux données mesurées. Les figures 3.19 et 3.20 nous décrivent la phase de test du modèle.



**Figure 3. 19: Approche bayesienne en phase test**



**Figure 3. 20: Droite de régression approche bayésienne en test**

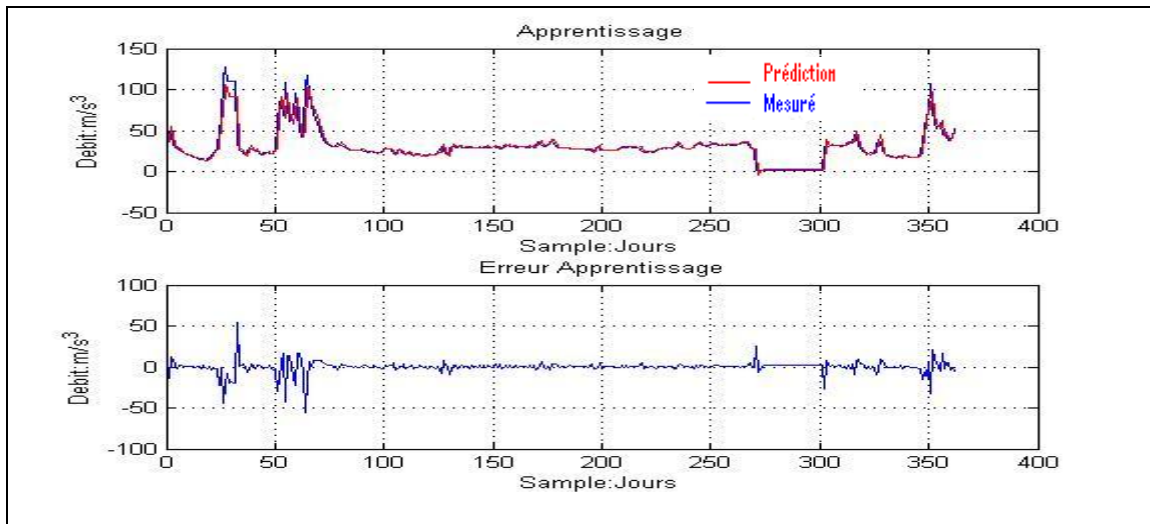
Droite de régression:

$$A = 0,726 (T) + 0,2901$$

(3.11)

### 3.3.4. La station d’Ambohimambola

En évaluant sa performance, le modèle n’enregistre que 0,2695 m<sup>3</sup>/s pour le critère RMSE et comme MAPE 0,22 % pendant l’apprentissage. Les figures 3.22 nous décrivent la phase d’Apprentissage du modèle.



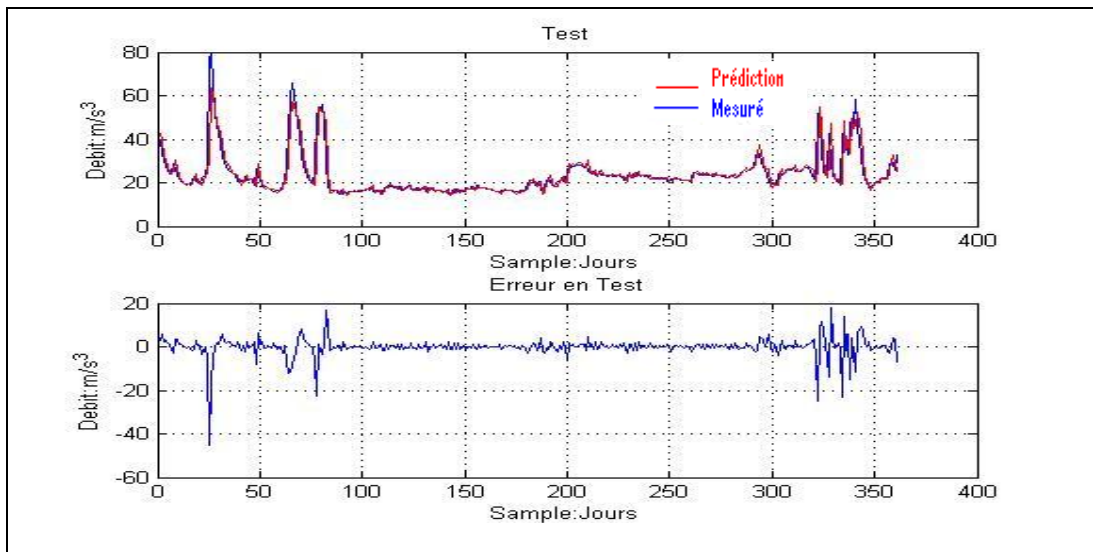
**Figure 3. 21: Approche bayésienne en phase d’apprentissage**

L’erreur de l’apprentissage est confirmée par la considération de la droite de régression qui donne la comparaison entre la sortie du modèle et les points de mesure.

Cette droite est donnée dans la formule suivante :

$$A = 0,925(T) + 0,2207 \tag{3.12}$$

En considérant la performance du modèle, et en évaluant sur la base de test, il est clair que le résultat est satisfaisant selon le critère RMSE enregistré qui est égal à 0.4740m<sup>3</sup>/s.



**Figure 3. 22: Approche bayésienne en phase test**

De même l'erreur de l'apprentissage est confirmée par la considération de la droite de régression qui donne la comparaison entre la sortie du modèle et les points de mesure.

$$A = 0,875 (T) + 0,0752 \quad (3.13)$$

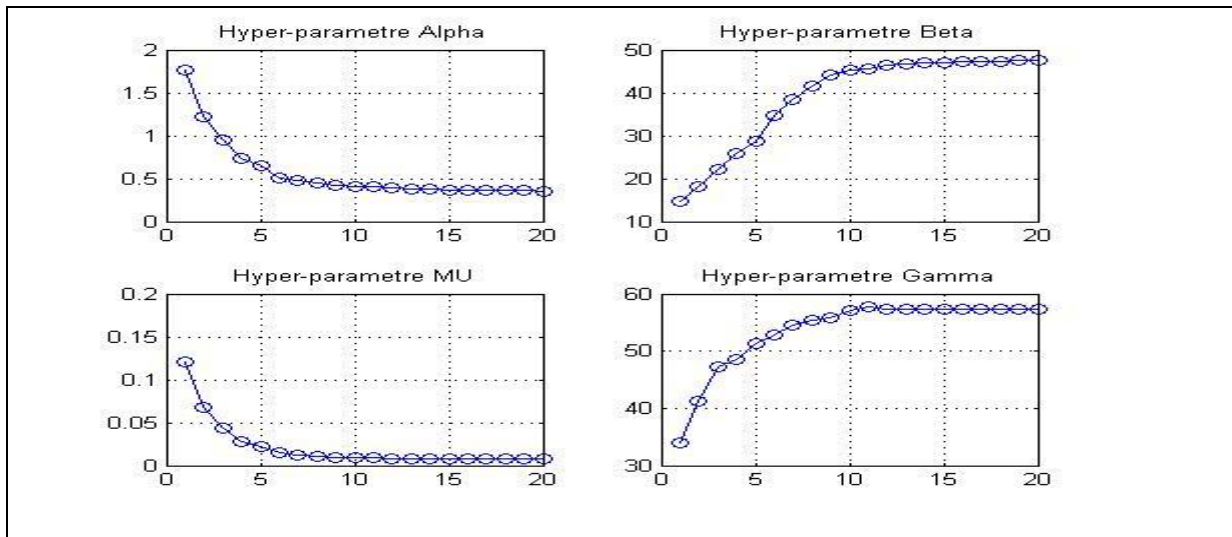
### 3.2.4.3 Recherche des modèles parcimonieux

#### 3.2.4.3.1. Station d'Anjeva

Rappelons que l'approche bayésienne consiste à déterminer les valeurs optimales des hyperparamètres  $\alpha_{MP}$  et  $\beta_{MP}$ . Comme nous avons vu au § 2.8.7 que ces valeurs correspondent au maximum du terme évidence qui est donné en fonction du nombre de neurones dans la couche cachée. Ainsi, nous avons fait varier le nombre de neurones cachés de 1 à 20 comme l'indique la figure 3.20. De plus cette même figure nous donne les valeurs des hyperparamètres  $\alpha$  et  $\beta$  les plus probables ( $\alpha_{MP}$  et  $\beta_{MP}$ ).

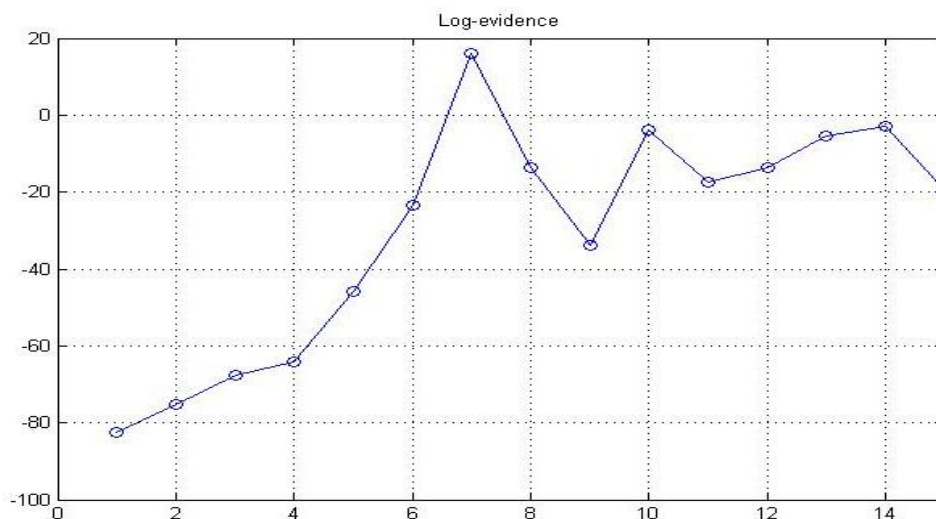
Nous avons aussi déterminé le nombre de paramètres effectifs pendant la même phase d'apprentissage qui exprime le nombre effectif des paramètres neuronaux dans le réseau. Nous constatons une convergence de sa valeur aux alentours de 48.





**Figure 3. 23: la détermination des hyper-parametres**

A partir de la courbe indiquant le Logarithme de l'évidence sur la figure 3.21, le nombre optimal de neurones dans la couche cachée pour le modèle est de 6, ce qui correspond à la valeur maximale de l'évidence.



**Figure 3. 24: Nombre de neurones cachés**

Avec les données de cette figure et la valeur calculée en utilisant la relation §.2.50, nous avons obtenu le nombre optimal de neurones cachés égal à 6 pour ce modèle.

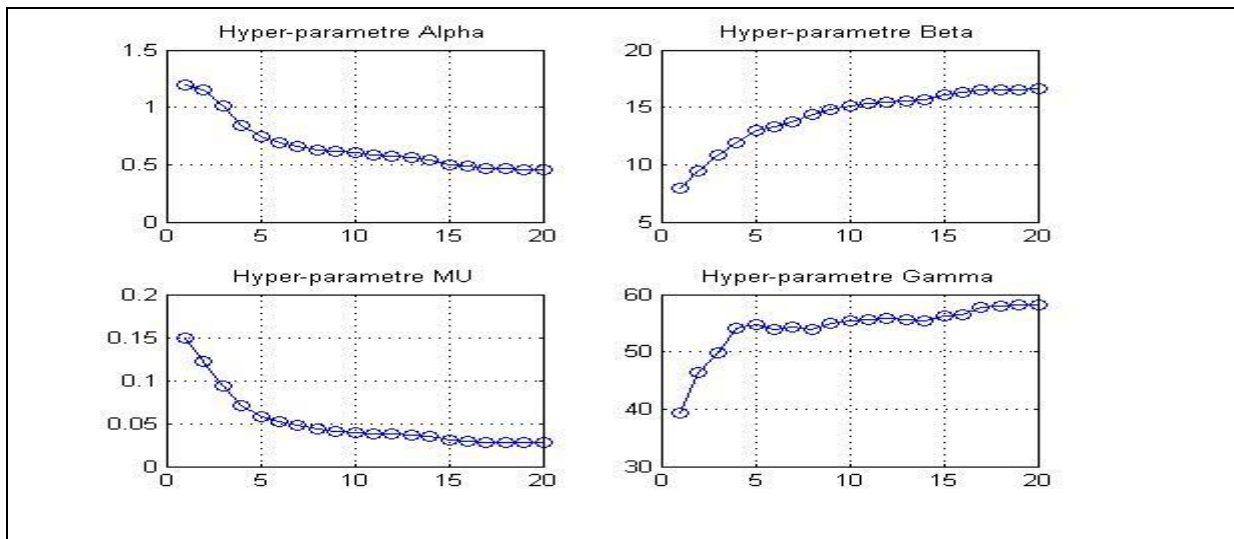
Nous en déduisons que 6 ou 7 neurones dans la couche cachée sont suffisants pour l'architecture de notre réseau construit à partir du modèle PMC, ce qui indique que le nombre de 10 que nous avons pris auparavant se situe dans le cas de surapprentissage.

### 3.2.4.3.1. Station d'Ambohimambola

Comme précédemment, nous avons suivi la même démarche pour la détermination des valeurs optimales des hyperparamètres  $\alpha_{MP}$  et  $\beta_{MP}$ . Ainsi, nous avons fait varier le nombre de

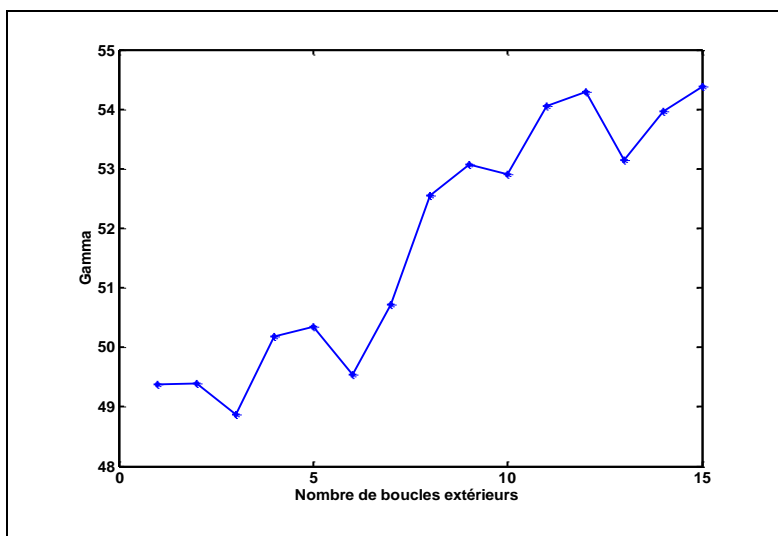
neurones cachés de 1 à 20 comme l'indique la figure 3.22. De plus cette même figure nous donne les valeurs des hyperparamètres  $\alpha$  et  $\beta$  les plus probables ( $\alpha_{MP}$  et  $\beta_{MP}$ ).

Nous constatons une convergence de sa valeur aux alentours de 58.



**Figure 3. 25: la détermination des hyper-parametres**

A partir de la courbe indiquant le logarithme de l'évidence sur la figure 3.23, le nombre optimal de neurones dans la couche cachée pour le modèle est de 6, ce qui correspond à la valeur maximale de l'évidence.



**Figure 3. 26: Evidence en fonction du nombre de neurones cachés.**

Nous en déduisons que 5 ou 6 neurones dans la couche cachée sont suffisants pour l'architecture de notre réseau construit à partir du modèle PMC, ce qui indique que le nombre de 10 que nous avons pris auparavant peut se situer dans le cas de surapprentissage.

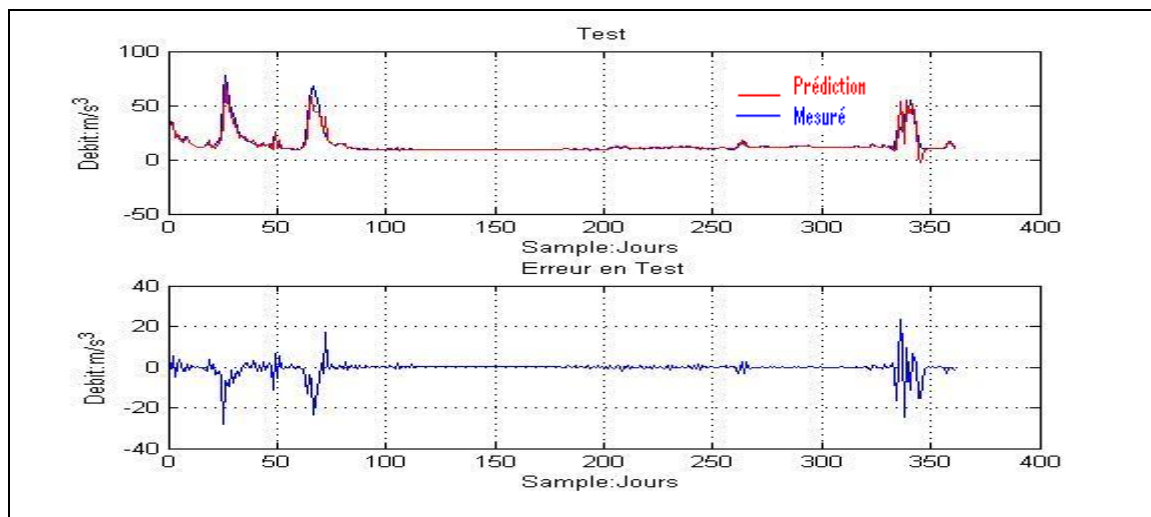
### 3.2.5.1. Comparaison des performances des modèle bayésien et optimisé

Dans cette partie, nous allons considérer seulement la station d’Anjeva. Le tableau 3.3 nous renseigne sur l’erreur de sortie pour les trois modèles adoptés.

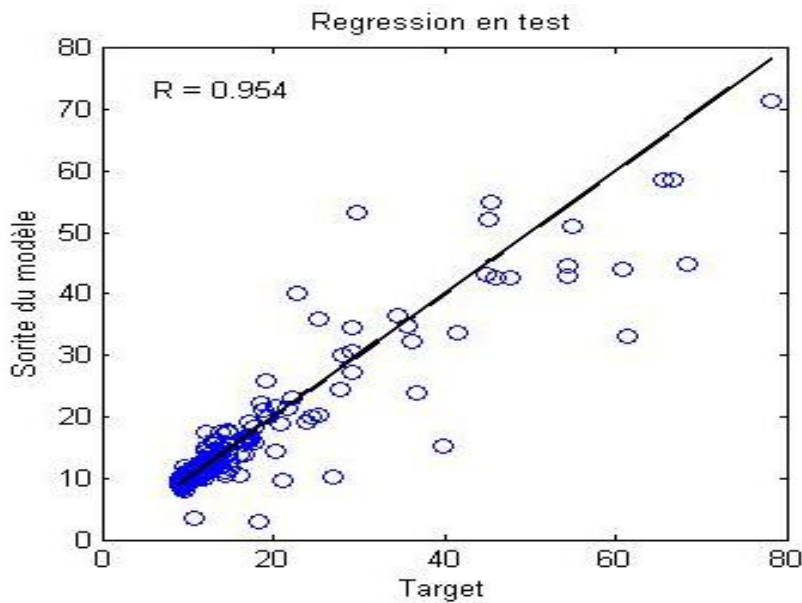
**Tableau 3. 4: Comparaison des erreurs des modèles**

Modèle	RMSE Training	RMSE Test	MAPE Training	MAPE Test
PMC classique (4 entrées, 1 sortie)	0.1620m <sup>3</sup> /s	0.3629m <sup>3</sup> /s	9.15%	7.10%
Couplage PMC - technique bayésienne (4 entrées, 1 sortie)	02015m <sup>3</sup> /s	03243m <sup>3</sup> /s	9.14%	6.06%
Couplage PMC - technique optimisée (4 entrées, 1 sortie)	01432m <sup>3</sup> /s	01827m <sup>3</sup> /s	8.71%	5.32%

Le modèle en apprentissage est illustré par les figure 3.24 et 3.25.



**Figure 3. 27: Approche bayésienne optimisé en phase test**



**Figure 3. 28: Droite de régression approche bayésienne optimisée en apprentissage**  
Droite de régression:

$$A = 0,95 (T) + 0,187$$

### 3.3. Interprétations

En évaluant la performance du réseau, en ce qui concerne les modèles à architecture statique, l'inférence bayésienne dans la modélisation est plus performante que l'utilisation du PMC classique. En effet, le modèle bayésien engendre moins d'erreurs. Cependant, le modèle classique peut traiter efficacement la modélisation du débit dans notre cas.

En introduisant le terme évidence, les critères de performances se sont encore diminués du fait que nous avons déterminé les valeurs plus probables des hyperparamètres (voir tableau N°3.4).

## CONCLUSION GENERALE

La mise en place d'une centrale hydroélectrique nécessite de nombreuses études hydrologiques et géologiques.

La théorie des réseaux de neurones nous a permis d'avoir des modèles performants capables d'optimiser la prévision du débit d'une rivière. Les réseaux de neurones présentent plusieurs avantages. Ils peuvent être vus comme un complément à l'Intelligence Artificielle classique. Ils disposent d'un apprentissage, tout en étant tolérant aux erreurs. Les réseaux de neurones sont robustes car ils reconnaissent les données bruitées ou incomplètes. Ils peuvent généraliser des modèles similaires et trouvent des solutions aux problèmes non linéaires.

Nous avons élaboré deux modèles. Dans un premier temps, grâce à l'utilisation de l'approche neuronale classique, nous avons pu voir la performance du modèle. Malgré sa complexité, des améliorations ont été apportées par d'autres techniques, ce qui nous a amené au deuxième modèle construit à partir de l'approche bayésienne, une méthode plus efficace comparée à l'approche neuronale classique.

L'utilisation des réseaux de neurones artificiels nous permis d'avoir des modèles globaux capables de générer les régimes moyens interannuels des cours d'eau aboutissant à des meilleures gestions et évaluations des potentiels hydroélectrique des sites d'implantation.

Après avoir modélisé les données hydrologiques des cours d'eau, nous avons pu arriver à des résultats concluants. L'utilisation de l'approche bayésienne à la prévision nous donne des résultats satisfaisants.

Les modèles étant applicables à toute zone des hautes terres centrales et partout dans Madagascar, moyennant quelques modifications.

Parmi les perspectives du présent travail, nous pouvons citer, entre autres, l'application des RNA à la modélisation des systèmes non linéaires. En effet, il est possible d'étendre notre étude à un niveau plus élevé de la météorologie et de la climatologie, les sources des énergies renouvelables étant fortement liées aux données météorologiques et climatiques du lieu d'étude.

---

## REFERENCES

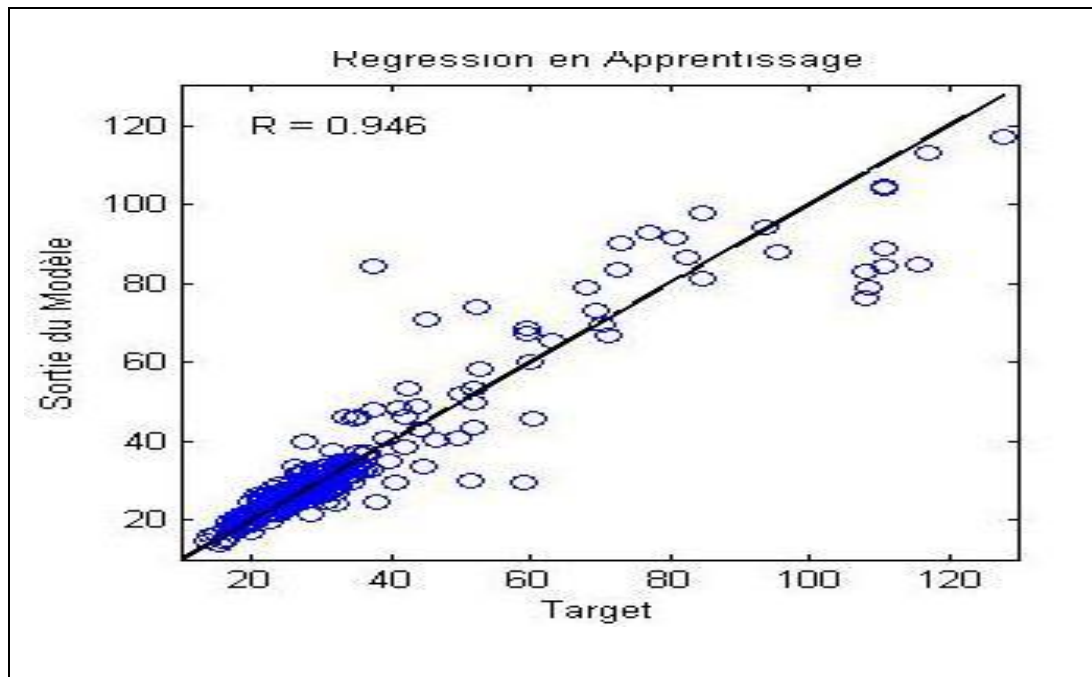
- [1] RANAIVOSON Narilala Daniel, «méthodologie de conception de réseaux de neurones approche modulaire», P 2
- [2] Vania ANDRONOVA, « Utilisation des données météo et des réseaux de neurones pour la prédiction de la vitesse du vent », juillet 2006.
- [3] W. S. McCulloch et W. Pitts, « A logical calculus of the ideas imminent in the nervous activity », Bulletin of Mathematical Biophysics, no 5, pp 115-133, 1943.
- [4] <http://fr.wikipedia.org/wiki/Neurone>
- [5] « Artificiel Neurone », <http://encyclopedia.thefreedictionary.com/McCulloch-Pitts+neuron>, juin 2006.
- [6] Hebb D.O, «The Organization of Behavior: A neuropsychological theory Wiley», New York, 1949
- [7] Rosenblatt, F. 1958, « The perceptron: a probabilistic model for information storage and organization in the brain », Psychological Review, 65: 386–408.
- [8] Minsky, M.L., Papert, S.A., 1969, « Perceptrons », MIT Press, Cambridge, MA.
- [9] Hopfield, J.J., 1982, « Neural networks and physical systems with emergent collective computational abilities » Proceedings of the National Academy of Sciences of the United States of American 79 (8), 2254–2258
- [10] Himour Yassine Djeghader Reda,2004,«Commande de systèmes non-linéaires dynamiques par réseaux de neurones»,
- [11] Hopfield, J.J., 1982, « Neural networks and physical systems with emergent collective computational abilities » Proceedings of the National Academy of Sciences of the United States of America 79 (8), 2254–2258
- [12] RANDRIANARIVONY Rija Nirina, 2007, « Modélisation du régime annuel des petits cours d'eau a faibles données hydrologiques par les réseaux de neurones : détermination des paramètres influents en vue de la mise en place d'une microcentrale hydroélectrique », p 57.
- [13] Hafiane Mohamed Lamine, « Conception d'un capteur de pression intelligent », Alger 2005.
- [14] Hornik K., Stinchcombe. White H. «Multilayer feedforward networks are universal approximators», Neural Networks 2: PP 359-366. (1989)

- 
- [15] HORNIK K., STINCHCOMBE M., WHITE H. & AUER P. "Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives», *Neural computation* vol 6, 199
- [16] J. L. McClelland and D. E. Rumelhart, *Explorations in Parallel Distributed Processing, a Handbook of Models, programs, and Examples*, MIT Press, Cambridge, 1988.
- [17] Rumelhart D., G. Hinton & Williams R., "Learning internal representations by error propagation," *Parallel Distributed Processing*, Vol. 1. D. Rumelhart and J. McClelland Eds. Cambridge: MIT Press, pp. 318-362, 1986.
- [18] Broyden C.G., 1970, "The convergence of a class of double-rank minimization algorithms 2: the new algorithm" *Journal Institute of Mathematics and its Applications* 6, pp. 222-231
- [19] Bayes T., 1763, « An essay towards solving a problem in the doctrine of chances », *Phil. Trans. Roy. Soc.*, 53, pp.370-418.
- [20] Bretthorst G. L., 1990, « An introduction to parameter estimation using Bayesian Probability, Maximum Entropy and Bayesian Methods », P. Fougere, eds., Kluwer Academic Publishers, Dordrecht the Netherlands, pp. 53-79.
- [21] MacKay D. J. C., 2003, « Information Theory, Inference and Learning Algorithms », Cambridge University Press, Cambridge, U.K.
- [22] Jaynes E. T., 1986, « Bayesian methods: an introductory tutorial, Maximum Entropy and Bayesian Methods in applied statistics », Cambridge University Press, Cambridge, U.K.
- [23] Bishop C. M., 1995, « Neural network for pattern recognition », Oxford University Press, Oxford, UK.
- [25] W. S. McCulloch et W. Pitts, « A logical calculus of the ideas imminent in the nervous activity », *Bulletin of Mathematical Biophysics*, no 5, pp 115-133, 1943.
- [27] Hebb D.O, «The Organization of Behavior: A neuropsychological theory Wiley», New York, 1949
- [28] Rosenblatt, F. 1958, « The perceptron: a probabilistic model for information storage and organization in the brain », *Psychological Review*, 65: 386–408.
- [29] Minsky, M.L., Papert, S.A., 1969, « Perceptrons », MIT Press, Cambridge, MA.
- [30] Hopfield, J.J., 1982, « Neural networks and physical systems with emergent collective computational abilities » *Proceedings of the National Academy of Sciences of the United States of America* 79 (8), 2254–2258

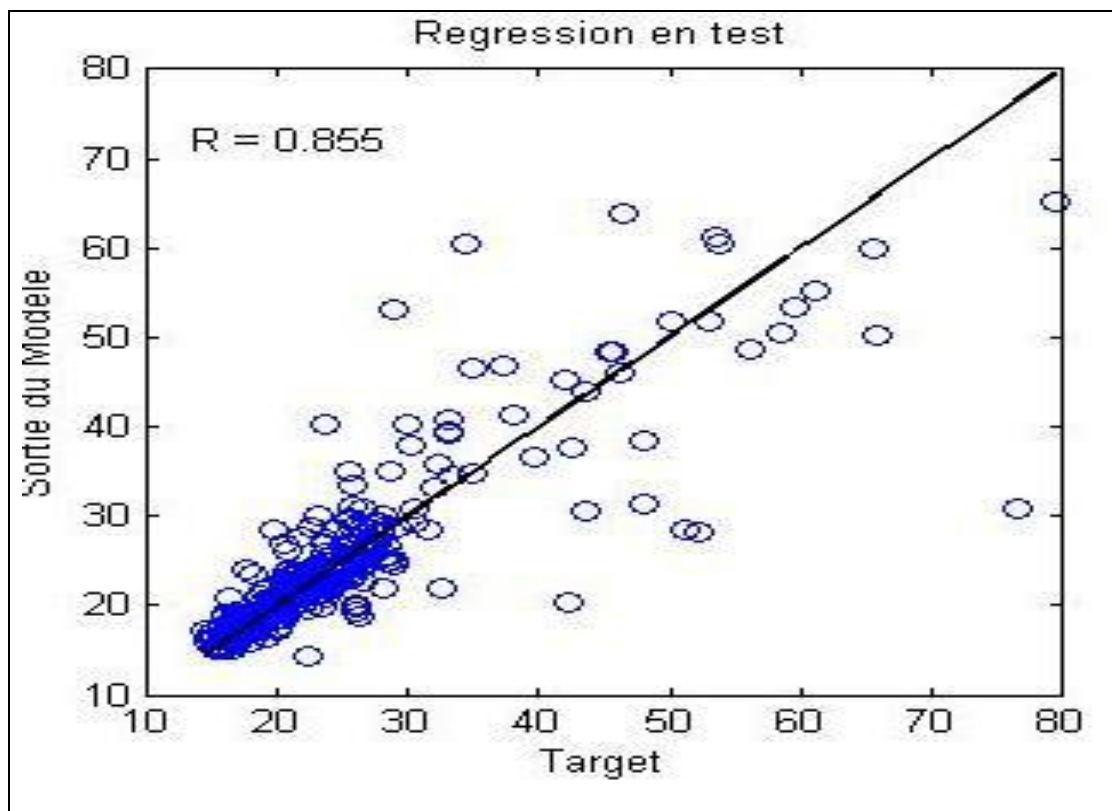
**ANNEXE :**

**RESULTATS OBTENUS AVEC LES DONNEES D'AMBOHIMANAMBOLA**

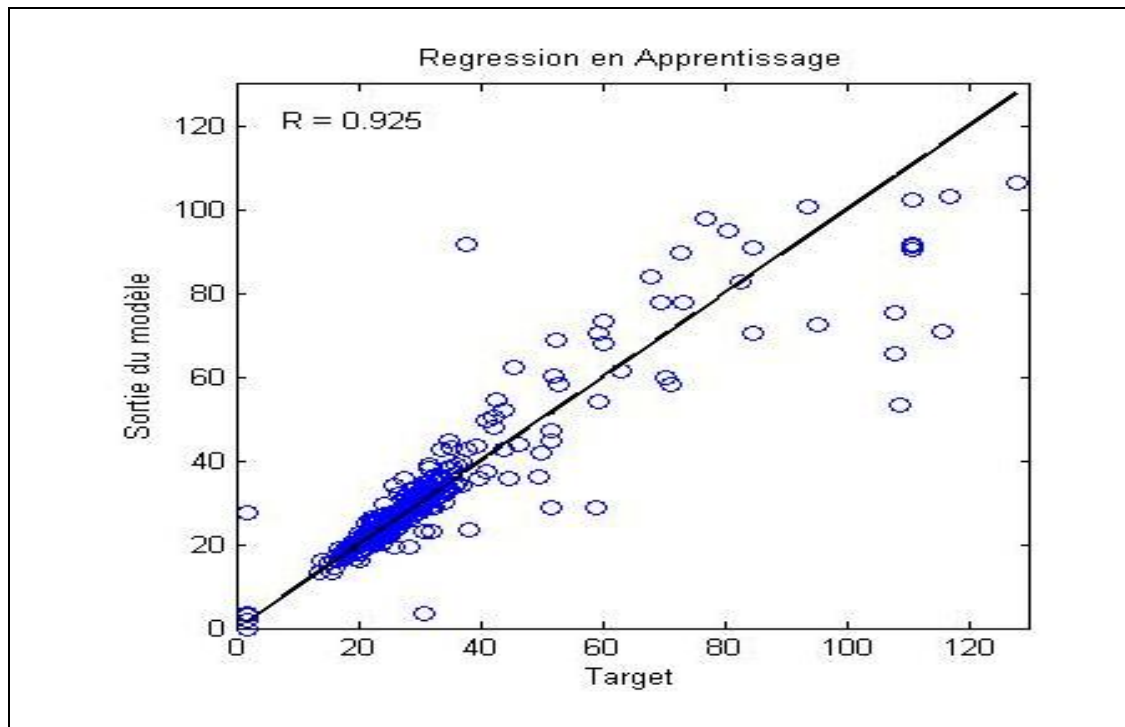




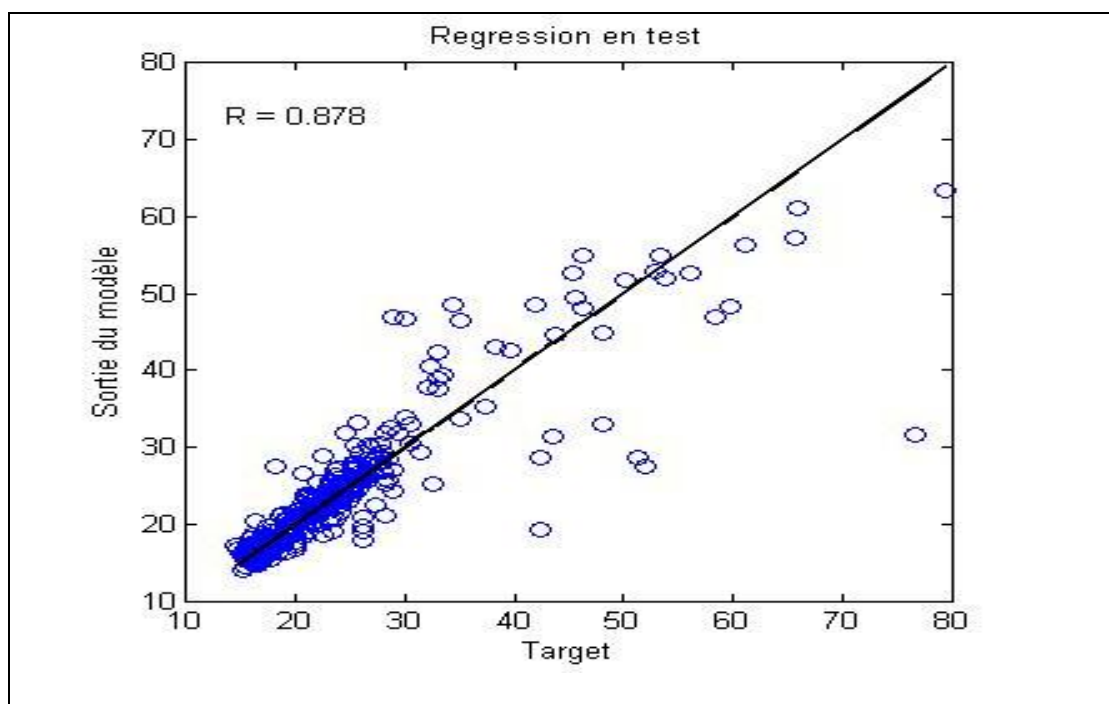
**Figure A1** : Droite de régression pour la phase d'apprentissage en utilisant les bases de données de la station d'Ambohimambola.



**Figure A2** : Droite de régression pour la prévision en utilisant les bases de données de la station d'Ambohimambola.



**Figure A3** : Droite de régression pour la phase d'apprentissage en utilisant les bases de données de la station d'Ambohimambola selon la théorie bayésienne.



**Figure A3** : Droite de régression pour la p en utilisant les bases de données de la station d'Ambohimambola selon la théorie bayésienne.

# OPTIMISATION D'ESTIMATION DU DEBIT D'UNE RIVIERE PAR LA THEORIE DE RESEAUX DE NEURONES ARTIFICIELS

## RESUME

Ce travail consiste à estimer avec optimisation le débit d'une rivière avant la mise en place d'une MCH, cette dernière étant confrontée à de nombreux problèmes tels que le manque de données hydrologiques.

Pour y remédier, les réseaux de neurones artificiels ont été utilisés. Nous avons utilisé les perceptrons multicouches et l'inférence bayésienne pour obtenir des modèles optimisés permettant la prévision d'une base de données manquante concernant le débit d'une rivière, moyennant la disposition d'une partie de cette base. En appliquant ces méthodes, nous avons obtenu des résultats satisfaisants.

**Mots clés** : Réseaux de neurones, Prévision, Microcentrale hydroélectrique, Débit.

## OPTIMIZATION OF ESTIMATE OF THE FLOW OF A RIVER BY THE THEORY OF RESEAUX OF ARTIFICIAL NEURONS

### ABSTRACT:

This work consists to estimate with optimization the flow of a river before the installation of a *MCH* which is confronted with many problems such as the lack of hydrological data.

To solve it, the artificial neurons networks are used. In order to obtain the optimized models allowing the forecast of a missing data base relating to the flow of a river, realising the provision of part of this base, the multi-layer and the inference Bayesian were used. By applying these methods, we obtained the satisfactory results.

**Key words**: neural networks, Forecast, Micro hydroelectric power, Flow.

Encadreur : Pr. RANDRIAMANANTANY Zely Arivelo	Impétrant : ASSAANDI Saindou. +261 34 01 464 73 Email : <a href="mailto:assaandi6@yahoo.fr">assaandi6@yahoo.fr</a> Lot VT 85 HBL Andohanimandroseza
--	---