

République Algérienne Démocratique et Populaire

**Ministère de l'Enseignement Supérieur et de
La Recherche Scientifique
Ecole Nationale Polytechnique
DEPARTEMENT D'HYDRAULIQUE**



**Mémoire de fin d'étude pour l'obtention du diplôme
d'ingénieur d'état en Hydraulique**

**Prévision des variables climatologiques
aux pas de temps mensuel et décadaire
par les réseaux de neurones et
l'Analyse en Composantes principales**

Proposé et dirigé par :

Pr. N. DECHEMI

Réalisé par :

M^{lle}. S. AMROUCHE

Promotion juin 2006

DEDICACES

Je dédie ce travail :

A mes très chers parents, que Dieu les protège ;

A mes sœurs et mes frères ;

A mes nièces Amel et Wafa, et mon neveu Ayoub ;

A toute ma famille ;

A mes chères amies et sœurs ;

A Khalti Malika.

Soumia

REMERCIEMENTS

*Avant toute personne, je dois remercier **Dieu** qui m'a donné la foi, le courage, la volonté et tout le bien.*

Mes sincères remerciements s'adressent à mes très chers parents pour leurs sacrifices, prières, et encouragements...

Je dois avouer que ce travail n'a pas de valeur sans le suivi, les conseils, et les orientations du

***Pr. DECHEMI**, mon promoteur et le directeur de l'équipe **GESTOP**.*

*Merci **Mr. DECHEMI**.*

*Ma gratitude et reconnaissance s'adressent à **Mr BERMAD**, **Mr BENKACI**, et tous les membres de l'équipe **GESTOP**.*

Je remercie les membres de jury pour leur acceptation d'examiner et juger mon travail.

*Les services de l'**ONM** trouvent ici toute ma reconnaissance pour leur aide en matière de données.*

Je tiens à remercier tous les enseignants qui ont contribué à ma formation.

Merci à mes frères, à mes sœurs, et à mes amies qui n'ont pas cessé de me soutenir, encourager et aider.

Merci à tous les gentils gens que j'ai connu dans ma vie.

Merci à ceux qui m'ont appris quelque chose dans ma vie.

Résumé :

La prévision des pluies et des évapotranspirations potentielles est nécessaire, pour une bonne gestion des ressources hydriques et l'alerte en cas de sécheresse ou inondation.

Ce présent travail propose des modèles de prévision des variables climatologiques cyclique et aléatoire. Ces modèles sont basés sur deux techniques différentes : Les réseaux de neurones artificiels et l'analyse en composantes principales.

Mots clés : pluie, évapotranspiration potentielle, prévision, réseaux de neurones, analyse en composantes principales.

Abstract :

The prediction of rains and potential evapotranspirations is necessary for a good management of water resources, and alarm in case of drought or flood.

This work proposes modals of cyclical and random climatological variables prediction. These modals are based on two different techniques: The artificial neural networks and the principal components analysis.

Keywords: Rain, potential, evapotranspiration, neural networks, principal components analysis.

SOMMAIRE

Introduction générale.....	1
Chapitre I : Modélisation.....	2
I-1-Définition de la modélisation.....	2
I-2-Les modèles hydrologiques.....	2
I-2-1- Les modèles d'interpolation ou de reconstitution.....	2
I-2-2- Les modèles de simulation.....	2
I-2-3- Les modèles de prédétermination.....	3
I-2-4- Les modèles de prévision.....	3
I-3- La prévision hydrologique.....	3
I-4- Etude bibliographique.....	4
I-4-1 Prévision par la famille des modèles ARMA.....	4
I-4-2- Prévision par les modèles Markoviens.....	4
I-4-3- Prévision par les réseaux de neurones artificiels.....	4
I-4-4- Prévision par la logique floue.....	5
Chapitre II : Réseaux de neurones artificiels.....	6
II-1-Introduction.....	6
II-2- Réseaux de neurones biologiques.....	6
II-3- Réseaux de neurones artificiels.....	8
II-3-1 Historique.....	8
II-3-2 Le premier neurone artificiel : modèle de Mc Culloch et Pitts.....	9
II-3-3- Modèles mathématiques.....	10
II-4- Propriétés de réseaux de neurones.....	12
II-5- Apprentissage des réseaux de neurones artificiels.....	13
II-6- Architecture des réseaux de neurones.....	15
II-6-1- Les réseaux feedforward.....	16
II-6-2 Réseaux récurrents.....	17
II-7- Quelques types de réseaux de neurones.....	18
II-7-1- Les réseaux monocouches.....	18
II-7-2- Les réseaux multicouches.....	20
II-7-3- Les réseaux à fonction de base radiale (RBF).....	21
II-7-4- Les réseaux récurrents.....	21
II-7-5- Les cartes auto organisatrices de Kohonen.....	23
II-8-Utilisation des réseaux de neurones.....	23
II -9- La mise en œuvre des réseaux de neurones.....	23
II-10- Les avantages des réseaux de neurones.....	24
Chapitre III : Analyse en Composantes Principales.....	25
III-1- Introduction.....	25
III-2 Définitions.....	26
III-2-1 Notions d'individu et de caractère.....	26
III-2-2 Définition algébrique.....	26
III-2-3 Définition géométrique.....	27
III-3- Formulation mathématique du problème.....	28

III-4- Procédé d'application de l'ACP.....	29
III-4-1 Calcul de la matrice des covariances.....	29
III-4-2 Calcul des composantes principales.....	32
III-5- Principaux résultats.....	32
III-6- Représentation graphique.....	33
III-7- Application de l'ACP.....	33
III-8-Influence de la transformation des variables.....	34
Chapitre VI : Collecte et critique des données.....	38
IV- 1- Définition et intérêt d'une série temporelle.....	38
IV-2- Présentation des données.....	39
IV -2-1 Calcul de l'évapotranspiration potentielle.....	40
IV-2-2 La formule de Penman.....	40
IV-3- Critique des données.....	40
IV-3-1 Méthodes de comblement des lacunes.....	41
IV-3-2 Détection des valeurs singulières.....	43
IV-3-3 Traitement statistique des données.....	44
IV-3-4 Test d'homogénéité.....	44
IV-3-5 Test d'indépendance.....	45
Chapitre V : Elaboration des modèles de prévision.....	48
V-1-Le perceptron multicouches.....	48
V-1-1 L'architecture.....	48
V-1-2 L'apprentissage.....	49
V-1-3 Programmation du modèle.....	50
V-2- L'Analyse en Composantes Principales.....	50
V-3- Modèle combiné (Réseaux de neurones +ACP).....	51
V-4 -Les critères de validation.....	51
Chapitre VI : Résultats et interprétations.....	53
VI-1- Modèle des réseaux de neurones.....	53
VI-2-Prévision de la pluie et de l'ETP.....	55
VI-2-1-Prévision de l'ETP.....	55
VI-2-1-1- Prévision de l'ETP au pas de temps mensuel.....	55
VI- 2-1-2- Prévision de l'ETP décadaire.....	60
VI-2-1-3-Conclusion.....	63
VI-2-2- Prévision de la pluie.....	63
VI-2-2-1- Prévision de la pluie mensuelle.....	64
VI-2-2-2 Prévision de la pluie décadaire.....	69
VI-2-2-3 Conclusion.....	71
Conclusion générale.....	73
Bibliographie.....	75
Annexes.....	78

Introduction générale

Les déficits et les excédents de précipitations (sécheresse-inondations), les fortes pluies d'orage dans les régions très urbanisées ont, sur les activités humaines, des conséquences souvent dommageables. Les organismes qui ont en charge la sécurité civile, l'assainissement pluvial, l'annonce des crues, la protection des végétaux, la sécurité routière, attendent des services de la météorologie des informations qui leur permettent de mettre en œuvre, en temps opportun, les moyens de prévention et d'intervention dont ils ont la maîtrise (BERMAD, 2004).

Par ailleurs, une bonne gestion des ressources en eau, le dimensionnement des ouvrages de retenue, et l'optimisation des volumes d'eau destinés à l'irrigation sont des problèmes d'influence majeure sur le développement économique et social d'un pays.

Le changement climatique, est un autre souci qui menace le globe terrestre, la vie et le bien de l'homme.

En fait, une prévision des pluies et des ETP, devient dans ces cas, une exigence nécessitant la compréhension de ces phénomènes et la connaissance des facteurs qui les engendrent. Plusieurs paramètres du régime hydrologique, comme les indicateurs de l'atmosphère (température, pression, vent humidité), les caractéristiques physiques du bassin versant, l'état du couvert végétal et le degré de saturation des sols, doivent être étudiés et parfois quantifiés afin de bien étudier et prévoir les pluies et les ETP.

L'objectif de ce projet de fin d'étude, est de chercher un modèle adéquat qui permet la prévision des variables cycliques et aléatoires pour deux pas de temps différents.

La pluie, étant variable aléatoire, reste un phénomène très complexe dont l'étude du processus physique et sa prévision sont des tâches très délicates face aux chercheurs.

En effet, afin de prévoir la pluie et l'ETP de la station d'Alger, nous avons opté pour deux modèles basés sur des approches différentes : les réseaux de neurones, et l'analyse en composante principale. Nous avons limité notre étude sur deux pas de temps à savoir le mensuel et le décadaire.

Notre travail est organisé de la manière suivante :

- Le chapitre I : fait un rappel sur la modélisation et l'étude bibliographique sur la prévision hydrologique ;
- Le chapitre II : détaille la théorie des réseaux de neurones ;
- La théorie de l'analyse en composantes principales est présentée dans le chapitre III ;
- La collecte et la critique des données sont données dans le chapitre IV ;
- Le chapitre V s'intéresse à l'élaboration des modèles utilisés ;
- Les résultats et leurs interprétations sont donnés dans le chapitre VI ;
- Enfin, ce travail est clôturé par une conclusion générale.

Chapitre I : Modélisation

I- 1- Définition de la modélisation

La *modélisation* est la simplification du phénomène réel, à des échelles de temps et d'espace différents du réel. Il s'agit généralement de trouver des équations reliant les différents paramètres du phénomène entre eux.

Gérard Dreyfus en 2002 de son côté, définit le modèle, dans son livre *réseaux de neurones méthodologie et application*, comme suit : (Un modèle est une représentation de la réalité visible ou observable. Selon Jean Perrin, *modéliser*, c'est « remplacer du visible compliqué par l'invisible simple »).

Il classe les modèles en trois catégories : les modèles *boîtes noires*, de *connaissance*, et semi physique.

- Les modèles de connaissance consistent à mettre en œuvre des équations décrivant les phénomènes étudiés (biologiques, physiques, chimiques...) en se basant sur des connaissances théoriques sur ces derniers. Ils ont une valeur prédictive et explicative.
- Par contre, les modèles boîte noire, qui ne nécessitent que la possession des mesures, n'ont qu'une valeur prédictive sur un domaine de validation.
- Entre ces deux types de modèles se situent les modèles semi physique. C'est l'élaboration des équations basées sur une théorie, et des équations purement empiriques qui résultent d'une modélisation de type boîte noire.

I- 2- Les modèles hydrologiques

En hydrologie, on peut classer les modèles selon leurs objectifs en quatre classes distinctes :

I-2-1- Les modèles d'interpolation ou de reconstitution:

Ils aident à mesurer, observer, constater, et estimer. Qu'il s'agisse d'acquérir l'information la plus «représentative» pour un effort de mesure donné, et donc de sélectionner les futurs lieux d'observation, ou bien d'évaluer une grandeur intégratrice d'une grande diversité spatiale, ou encore de compléter une information momentanément manquante, ils servent donc autant à décrire qu'à décider. (J. Miquel 2005).

Ils peuvent être utilisés pour *combler* les valeurs manquantes d'une variable hydrologique, à cause d'une panne momentanée du dispositif de mesure. Ainsi, ils peuvent aider à *estimer* les valeurs d'une variable hydrologique, sur des points du bassin versant, autres que les points de mesure. C'est donc une sorte d'*interpolation*. Leur principe est le même : reconstituer aux mieux la réalisation d'un processus dont on en connaît une partie.

I-2-2- Les modèles de simulation

Il s'agit de produire des scénarios possibles dans le futur d'un phénomène donné, en se basant sur des informations issues de l'historique.

Leur principe est d'utiliser des générateurs de variables aléatoires afin de décrire l'évolution temporelle, autant que spatiale, du phénomène.

Ils permettent la projection (la planification) pour le futur, comme ils aident à prendre des décisions, surtout en cas de risques.

I-2-3- Les modèles de prédétermination

La prédétermination d'un phénomène donné consiste à chercher la probabilité d'occurrence de ce dernier, sans s'intéresser à préciser le moment où il se produirait.

Les modèles de prédétermination sont utiles pour la projection, l'évaluation des risques ainsi que pour le dimensionnement des ouvrages soumis aux aléas hydrométéorologiques : ponts, barrages,...

I-2-4- Les modèles de prévision

Ils consistent à chercher la valeur d'une variable au temps $t+\Delta t$, en se basant sur l'historique de celle-ci, connu jusqu'au temps t .

Leur finalité est purement opérationnelle : alerte en période de crue, anticipation en période de sécheresse, plus généralement aide à la gestion de la ressource en eau d'un bassin.

Vu que notre domaine d'intérêt dans ce travail est la prévision des (trois) paramètres, à savoir la pluie, l'évapotranspiration potentielle et la température maximale moyenne, et ceci pour le pas de temps mensuel et décadaire, on doit détailler des notions requises dans le domaine de prévision hydrologique.

I-3- La prévision hydrologique

La prévision hydrologique est définie comme l'estimation des conditions futures des phénomènes hydrologiques pour une période donnée, à partir des observations passées et actuelles (WMO 1994). Son objectif général est de fournir

les meilleures estimations de ce qui peut arriver en un point donné à une date future précise (Hipel et McLeod 1994), contrairement à la prédiction qui vise l'estimation des conditions futures sans référence à un temps spécifique (Lettenmaier et Wood 1993).

On peut noter que les éléments caractéristiques de la prévision sont :

- La variable à prévoir et les variables explicatives (c'est-à-dire l'ensemble des données disponibles) ;
- L'horizon de prévision ;
- Les méthodes de calcul ou d'estimation ;
- l'objectif de la prévision (alerte de crue, planification de l'opération des réservoirs, des projets d'irrigation ou de navigation) ;
- Le type de résultats désirés (valeurs numériques, graphiques, ou distribution de probabilités).

Du point de vue mathématique, deux types d'approches ont été utilisés dans la prévision hydrologique : *l'approche déterministe* (ou conceptuelle) et *l'approche empirique* (ou stochastique).

- **L'approche déterministe** est basée sur la simulation physique du système. Le modèle déterministe suppose qu'un calcul exact des paramètres est possible. Dans de nombreux cas, cette hypothèse s'avère peu réaliste, car les phénomènes naturels sont très aléatoires. Les modèles déterministes sont finalement tous limités par le grand nombre de paramètres à mesurer et par les limites des connaissances actuelles des systèmes naturels complexes.

- **L'approche stochastique** permet de passer outre les limites des connaissances physiques du système. Elle prend seulement en compte l'aspect aléatoire des phénomènes. Contrairement aux modèles déterministes, les modèles empiriques fonctionnent comme des « boîtes noires », c'est à dire sans aucune considération de la structure interne du système. (P. Coulibaly 1999)

I-4- Etude bibliographique

I-4-1 Prévision par la famille des modèles ARMA

Parmi les méthodes de prévision, les modèles linéaires sont de loin les plus employés, en hydrologie, en particulier par les services d'annonce de crues. Ces modèles appelés ARMA (Auto Regressive Moving Average), développés initialement par Box et al. , (1970) et (1976), consistent à mettre en relation linéaire le débit que l'on souhaite prévoir avec des variables explicatives (débit au pas de temps précédents mesurés à la même station, débit amont, pluie, évaporation) :

$$Q_t = \sum_{j=k}^n a_j Q_{t-j} + \sum_{j=k}^n \sum_{i=1}^m b_{i,j} X_{i,t-j} + \varepsilon_t \quad (\text{I-1})$$

Q : Débit mesuré à la station où l'on souhaite faire des prévisions ;

X_i : Autres variables explicatives retenues ;

K : Horizon de prévision ;

ε : Bruit blanc ;

t : Indice de temps.

Ce modèle est capable de modéliser le relation mathématique pluie-débit sans prendre en considération le processus physique (Hsu, et al. ,1995).

Son principal inconvénient, est le nombre élevé de paramètres à estimer.

Les modèles linéaires ARMAX (Auto Regressive Moving Average with eXternal data), qui supposent que les pluies et les débits sont liés par une relation linéaire invariante dans le temps, n'ont pas abouti aux attentes des hydrologues, en effet d'après Shamseldin (1997), ces modèles ne sont pas considérés comme étant fiables et opérationnels, mais seulement pour offrir une assise de comparaison pour les autres modèles plus performants.

I-4-2- Prévision par les modèles Markoviens

Seidou et al. , (2002) ont proposé une méthode pour modéliser l'incertitude sur les apports à court terme en fonction de la période de l'année. Pour cela, un modèle paramétrique markovien à temps continu, basé sur le processus de Poisson filtré, est construit et estimé grâce à des observations régulièrement espacées dans le temps puis validé sur trois stations s hydrométriques de Québec (Canada). Les points étudiés sont sa capacité à représenter correctement les transitions de débits et ses performances en tant que prédicteur de la valeur à venir du débit. Ce dernier aspect est comparé à celui du modèle ARMA saisonnier.

I-4-3- Prévision par les réseaux de neurones artificiels

Ces dernières années une nouvelle génération de modèles est apparue, basée sur le concept des réseaux de neurones humains (Jodouin, 1994). French et al. , (1992) l'ont utilisé pour la première fois dans la prévision spatiale et temporelle des précipitations, alors que Zealand et

al. , (1999), Compolo et al. , (1999) l'ont appliqué pour la prévision de l'écoulement des rivières.

Ces modèles ont trouvés une vaste utilisation dans la simulation et la prévision des séries hydrologiques et notamment dans la modélisation pluie-débit (Lek et al. 1996, Hsu et al. 1995, Minns et al. 1996, Sajikumar et al. 1999, Shamseldin 1997, Coulibaly et al. 1999).

Les réseaux de neurones artificiels constituent une nouvelle méthode d'approximation des systèmes complexes, particulièrement utiles lorsque ses systèmes sont difficilement modélisables à l'aide des méthodes statistiques classiques.

I-4-4- Prévision par la logique floue

La régression floue est un autre outil récent issu des théories des sous ensembles, elle a été étudié d'une manière détaillés par Bardossy en 1990. En effet, ce dernier l'a utilisé, pour quantifier la prévision de la demande en eau en tenant compte des prévisions météorologiques, et pour prévoir la température moyenne journalière.

La régression linéaire floue repose sur l'hypothèse d'un modèle linéaire dont les coefficients sont flous :

$$Y = A_0 + \sum_{k=1}^K A_k (x_k - \bar{x}_k) \quad (I-2)$$

$A_k, (k=1,2,\dots,k)$: est le nombre flou ;

$x = \{x_1, x_2, \dots, x_k\}$: est un point de l'espace choisi, pour lequel on estime que le modèle est plus précis.

Au cours de ces dix dernières années, les systèmes à base de règles floues se sont avérés des outils souples et puissants pour la modélisation de l'incertain dans un très grand nombre de domaines, et surtout dans la prévision des séries temporelles (Fiordaliso, 1999).

L'estimation des paramètres d'une régression linéaire floue est un problème de programmation linéaire. Les avantages de ce type de modèle proviennent du fait qu'il est basé sur des hypothèses moins restrictives que celle de la régression linéaire classique.

La régression linéaire floue est cependant très sensible aux valeurs extrêmes, qui doivent être validées avec précaution. Un autre inconvénient de cette méthode est la complexité de son algorithme d'estimation des paramètres ; pour résoudre 2^K systèmes d'inéquations à $2(K+1)$ inconnues et $2n$ contraintes, où n est le nombre d'observations, la méthode devient impraticable lorsque K et n sont grands mais avantageuse quand ces derniers sont petits.

Malgré ses avantages la méthode et ses fondements de base ont été remis en question par plusieurs scientifiques (Bermad, 2004).

Chapitre II : Les réseaux de neurones artificiels

II-1-Introduction

Depuis plus de cinquante ans, les chercheurs s'inspirent de la biologie dans le but de réaliser des structures capables de résoudre divers problèmes complexes. Le cerveau, constitué d'une centaine de milliards de neurones (connectés chacun à parfois plus de dix mille autres de manière tridimensionnelle) a de quoi attirer l'attention.

En effet, le cerveau est la machine de calcul la plus sophistiquée que nous connaissons. Même si l'être humain est incapable de réaliser des calculs avec la rapidité et la précision d'un ordinateur, il est autrement bien adapté que ce dernier à des problèmes de traitement de l'information dans son ensemble comme : la perception, le traitement du signal, l'extraction des caractéristiques pertinentes dans un monde complexe, et la décision, qui sont tous des tâches que nous accomplissons couramment (Yahiaoui et Ariat, 2002).

Ces tâches sont accomplies avec une grande rapidité, et le cerveau adapte son comportement aux situations nouvelles, principalement selon deux processus :

- **L'adaptation rapide et automatique (réflexe)** : c'est le cas de notre pupille, par exemple, qui adapte son diamètre en fonction de la quantité de lumière,

- **l'apprentissage** : cet apprentissage autorise une adaptation lente d'un individu à l'exécution d'une tâche nouvelle, par exemple, la conduite automobile, ou l'utilisation d'une nouvelle théorie des mathématiques.

Toutes ces constatations ont constitué un ensemble de motivations important poussant à analyser le fonctionnement de cette machine. Les biologistes ayant isolé son élément de base, à savoir le **neurone**, une approche mécaniste de la pensée a alors vu le jour : en effet si on modélise le fonctionnement de l'élément de base et si l'on reproduit les interactions des éléments de base entre eux, alors en est certainement capable de reproduire artificiellement la pensée.

II-2- Réseaux de neurones biologiques

Le système nerveux est principalement composé de deux types de cellules : les cellules gliales et les neurones. Les premières servent de support et de protection pour les neurones qu'elles entourent. Ces derniers sont chargés de transporter l'information qui arrive ou part du système nerveux central (cerveau et moelle épinière) sous la forme d'influx électriques. Ils interviennent dans le contrôle des muscles ainsi que dans le raisonnement. Le cerveau humain comporte de 10^{11} à 10^{12} neurones interconnectés par un nombre encore plus grand de synapses (de 10^{14} à 10^{15}). Ces dernières évoluent en fonction de l'environnement : leur formation continue à se faire en fonction des informations reçues après que le cerveau ait terminé de se former. Il y a trois états possibles pour une synapse : elle est soit labile, stable ou bien évolue vers la dégénérescence. Une synapse labile peut devenir stable si elle est fonctionnelle (c'est à dire si elle est sollicitée) ou dans le cas contraire dégénérer : on parle de stabilisation sélective des synapses. Cette évolution traduit l'apprentissage de la mémoire qui ne s'arrête pas avant quinze voire vingt ans. Voyons à présent plus en détail ce qu'est un **neurone biologique**.

Un **neurone** (figure II-1) est une cellule comprenant une partie centrale (le corps cellulaire, soma ou péricaryon) qui renferme le noyau, un prolongement appelé **axone**, dont la longueur varie chez l'Homme entre 1 millimètre et 1 mètre, et enfin de petits diverticules ramifiés appelées **dendrites** (Le Fablec, 1999).. Les entrées du neurone sont donc **les dendrites** et sa sortie est **l'axone**. Les dendrites réceptionnent **les signaux d'entrée** provenant d'autres neurones et les propagent au corps du neurone où ils seront transformés en **impulsions sortantes**. Ces dernières sont conduites par les axones et se dirigent du corps cellulaire vers les cellules suivantes (un autre neurone ou une cellule musculaire par exemple).

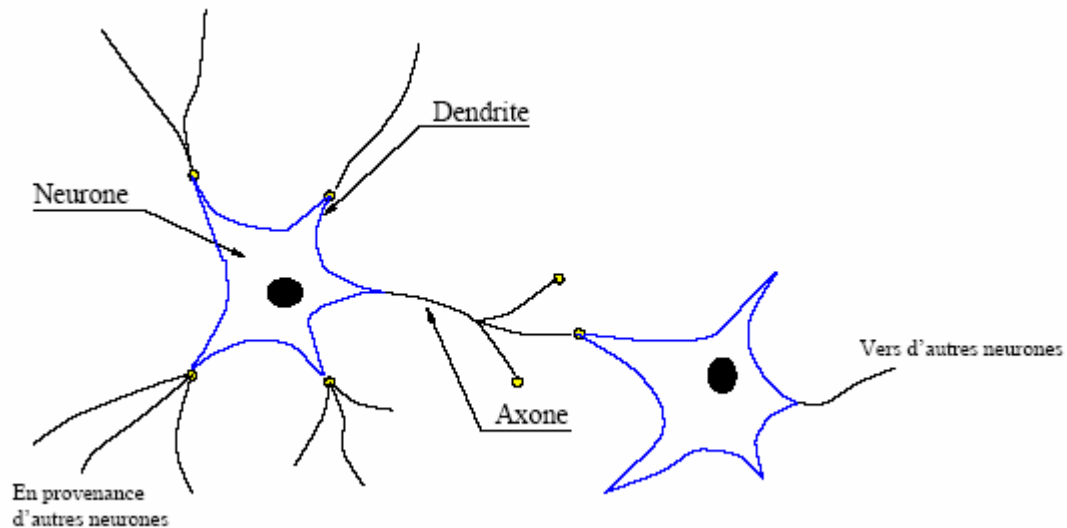


Figure II-1 : Neurone biologique

Dans chaque neurone, un traitement de l'information est effectué : lorsque la somme temporelle, sur une durée d'intégration de l'ordre du milliseconde, et/ou la somme spatiale de l'énergie de ses sollicitations dépasse un certain seuil, alors le neurone propage un signal de sortie, impulsion appelée **spike**, le long de son axone. Lorsque cette somme reste en deçà du seuil, alors l'axone ne propage rien.

L'axone d'un neurone peut se partager pour venir activer les entrées, les dendrites et le corps cellulaire des autres neurones. La connexion ainsi réalisée entre deux cellules est appelée **synapse**. Lorsque l'axone touche l'autre neurone, il le fait en séparant son influx en plusieurs boutons appelés boutons synaptiques. Plus il y a de bouton, et plus la valeur de l'entrée renforcée. Les boutons synaptiques peuvent être inhibiteurs ou excitateurs (Yahiaoui et Ariat, 2002).

En fait, **la synapse** est un lieu d'interactions fort complexes où le signal électrique de l'impulsion nerveuse est converti en un signal biochimique. Des substances biochimiques appelées *neurotransmetteurs* sont libérées par le terminal axonal présynaptique du neurone émetteur et traversent rapidement la courte distance qui les sépare de la membrane cellulaire *post-synaptique* du neurone récepteur. Un des effets possibles des neurotransmetteurs est d'*exciter* les neurones qui la reçoivent ; si suffisamment d'impulsion excitatrices convergent en un temps suffisamment court sur un même neurone, celui-ci produira généralement à son tour une impulsion nerveuse, et affectera ses voisins. L'effet inverse existe aussi : *l'inhibition*. Un neurone réduira sa tendance à produire des impulsions nerveuses.

Les principaux avantages d'un réseau de neurones biologique sont les suivants : (Le Fablec, 1999)

- Il est capable d'apprendre en adaptant ses poids synaptiques en fonction de l'environnement ;
- Il peut utiliser des informations probabilistes et imprécises ;
- A partir de cas d'apprentissage particuliers, il sait généraliser ;
- Il est très résistant aux "pannes".

L'idée de copier l'organisation des cellules du cerveau humain est venue très tôt. Le but était de permettre à des machines de "raisonner" et de reproduire les avantages des réseaux de neurones biologiques.

II-3- Réseaux de neurones artificiels

II-3-1 Historique

Bien que le concept de la mémoire associative était introduite pour la première fois par le célèbre psychologue américain **W. James** en **1890**, qui propose ce qui deviendra une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones connues plus tard sous le nom de **la loi de Hebb**, l'étude des réseaux de neurones, proprement dite, date des années 1940-50. Héritière de la révolution scientifique qui mobilisait les esprits à cette période, cette discipline trouve ses sources à la fois dans les connaissances neurobiologiques de l'époque et dans un nouvel outil permettant d'explorer par simulation le comportement des modèles théoriques proposés, *l'ordinateur* (Jodouin, 1994).

Cette décennie vit aussi plusieurs travaux importants (en particulier les réflexions sur la nature du raisonnement humain de **John Von Neumann**, d'**Alan Turing**, de **Alonzo Church**, d'**Alan Newell** et de **Herbert Simon**) qui galvanisèrent les recherches sur le cerveau et l'esprit humain et posèrent les fondations de ce qui allait devenir *l'intelligence artificielle* (Jodouin, 1994).

Alan Turing et **John Von Neumann** étaient parmi les premiers à s'être intéressés au cerveau humain. Ils ont l'idée d'une machine basée sur le modèle de neurone artificiel proposé en **1943** par le neurophysiologiste **Warren McCulloch** et le logicien **Walter Pitts** qu'on décrira plus loin. Bien qu'il soit essentiellement connu pour la machine qui porte son nom, **Alan Turing** est aussi un des pères des réseaux de neurones : dès **1948** il les étudie dans un article intitulé **Intelligent Machinery** (publié en 1968, 14 ans après sa mort) où il décrit les principes du connexionnisme. Il introduit une machine constituée de neurones artificiels associée à des dispositifs permettant de modifier leurs connexions. La publication tardive de ces travaux fait que l'on ne redécouvre que plus tard les réseaux de neurones.

En **1949**, **Hebb** souligna l'importance du couplage synaptique dans le processus d'apprentissage en se fondant son argumentation sur le comportement psychophysiologique: première règle d'apprentissage.

Ainsi, les premiers réseaux de neurones artificiels basés sur la vision des systèmes biologiques apparaissent sous le terme **perceptron** proposé en **1958** par **Frank Rosenblatt**. Inspiré des idées de Hebb, Mc Culloch et Pitts, le perceptron était en mesure d'apprendre à calculer certaines fonctions logiques par l'exemple en modifiant ses connexions synaptiques.

Le **perceptron**, qui est le réseau le plus simple possible puisqu'il est composé d'un seul neurone, a été à l'origine développé pour la reconnaissance de formes. Par la suite, **Rosenblatt** en 1960, donne la preuve de convergence du **perceptron**.

Dans la même époque, des réseaux similaires appelés **ADALINE (Adaptive Linear Neuron networks)** furent inventés (Widrow 1960).

Malgré cela, dans les années 60, on a trouvé que les perceptrons simples ne peuvent pas résoudre beaucoup de problèmes. De plus, en 1969, **M.L Minsky** et **S.A Papert** montrent les limites du perceptron monocouche et indiquent qu'utiliser **des perceptrons multicouches** n'arrange a priori rien. Ils décrivent en effet une série de problèmes, dits d'apprentissage difficile, de reconnaissance de formes que les perceptrons ne peuvent résoudre (problème du ou exclusif (XOR), qui sera détaillé plus loin). Cela a pour effet de réduire drastiquement la recherche sur les réseaux de neurones artificiels qui sont alors délaissés pendant une dizaine d'années.

Cependant, certains tels **Wilkie**, **Stonham** et **Aleksander** persistent et parviennent en 1980 à résoudre les problèmes d'apprentissage difficiles en utilisant, malgré tout, des réseaux de neurones multicouches. D'autres études telles celles de **Grossberg** ((Grossberg1972), (Grossberg1976a) et (Grossberg1976b)) et **Kohonen** en 1982 contribuent elles aussi au regain d'intérêt pour les réseaux de neurones formels et leurs dérivés. Dans la même période Werbos invente l'idée de rétropropagation en 1974.

Par la suite se développent de nouveaux types de réseaux de neurones formels (**Hopfield**, 1982, [Hopfield84]) et des méthodes d'apprentissage (algorithme de rétro propagation de l'erreur).

Destiné aux réseaux avec des connexions non bouclées, l'algorithme de rétro propagation a été développé indépendamment par **Le Cun** (1985 et 1988), **Rumelhurt** (1986b et 1986a), et **Parker** (1987).

Cela permettra alors la diffusion de l'utilisation des réseaux de neurones dans de nombreuses applications (reconnaissance de caractères, de la parole.

Dans les années 90, d'autres modèles et algorithmes ad hoc ont été développés avec l'attention d'étendre les possibilités des réseaux statiques non bouclés: introduction progressive d'une mémoire interne et de la récurrence au sein du réseau.

II-3-2 Le premier neurone artificiel : modèle de Mc Culloch et Pitts (le neurone formel)

Les constatations faites pendant l'analyse du neurone biologique ont conduit au schéma de fonctionnement du neurone artificiel de **Mc Culloch et Pitts (Figure II-2)**.

Ce neurone artificiel est une simplification extrême du neurone biologique dont le comportement temporel, notamment, est fort complexe (importance des temps de propagation, importance des dispersions temporelles des entrées dans le comportement du neurone, évolution dans le temps du seuil lorsque le neurone a fonctionné, etc.).

Un neurone formel fait une somme pondérée des potentiels d'action qui lui parviennent (chacun de ces potentiels est une valeur numérique qui représente l'état du neurone qui l'a émis), puis s'active suivant la valeur de cette sommation pondérée. Si cette somme dépasse un certain seuil, le neurone est activé et transmet une réponse (sous forme de potentiel d'action) dont la valeur est celle de son activation. Si le neurone n'est pas activé, il ne transmet rien.

Dans ce modèle, les poids synaptiques (qui remplacent les boutons synaptiques du neurone biologique) pondèrent les n entrées x_i . Le sommateur envoie donc dans la fonction seuil une combinaison linéaire des n entrées. Les excitations sont simulées par les poids synaptiques positifs, et les inhibitions par les poids synaptiques négatifs (Yahiaoui et Ariat, 2002).

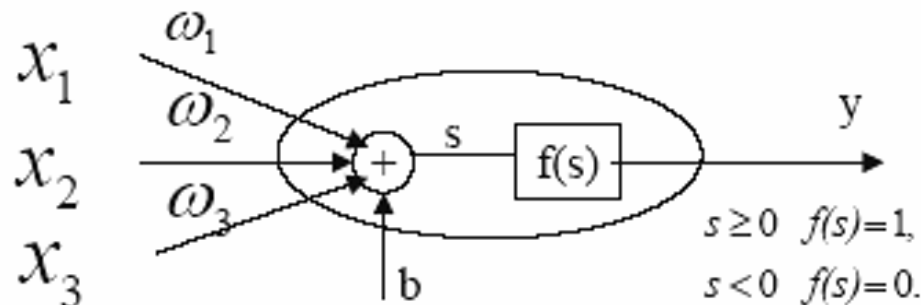


Figure II-2 : Neurone formel de Mc Culloch et Pitts

II-3-3- Modèles mathématiques

Les réseaux de neurones biologiques réalisent facilement un certain nombre d'applications telles que la reconnaissance de formes, le traitement du signal, l'apprentissage par l'exemple, la mémorisation, la généralisation. Ces applications sont pourtant, malgré tous les efforts déployés en algorithmique et en intelligence artificielle, à la limite des possibilités actuelles. C'est à partir de l'hypothèse que le comportement intelligent émerge de la structure et du comportement des éléments de base du cerveau que les réseaux de neurones artificiels se sont développés. Les réseaux de neurones artificiels sont des modèles, à ce titre ils peuvent être décrits par leurs composants, leurs variables descriptives et les interactions des composants (Touzet, 1992).

Le modèle mathématique le plus souvent utilisé est celui de Mc Culloch et Pitts représenté sur la **figure II-2**. On y reconnaît les divers éléments associés à un neurone biologique: (Le Fablec)

- Des synapses que nous appellerons par la suite connexions sur lesquelles arrivent les entrées x_i . Elles sont caractérisées par un poids : chaque signal x_i passant sur une connexion sera multiplié par le coefficient de pondération W_i ;
- Un mécanisme de sommation permettant de faire une combinaison linéaire des entrées pondérées précédentes ;
- Un biais b qui représente le seuil au dessus duquel le neurone sera excité. Il sert à centrer la zone d'action du neurone ;
- Une sortie y qui peut être utilisée comme entrée pour d'autres neurones ;
- Une fonction d'activation f qui sert à limiter sa sortie et permet d'agir sur la forme de la zone sur laquelle agit le neurone (ainsi, si l'on considère qu'un neurone agit sur un certain espace, le biais représente l'endroit où le neurone se trouve et la fonction d'activation indique comment il intervient sur la zone de l'espace qui se trouve à proximité). (**Figure II-3**)

En général, on utilise une fonction de type sigmoïde, c'est cette fonction qui introduit la non linéarité dans le neurone.

Un neurone réel produit une impulsion à partir d'influx afférents provenant d'autres neurones. De la même manière, le neurone artificiel produit une sortie associée à un ensemble d'entrées : la sortie y est calculée à partir des entrées x_i avec :

$$f = ((\sum w_i \times x_i) - b) \tag{II-1}$$

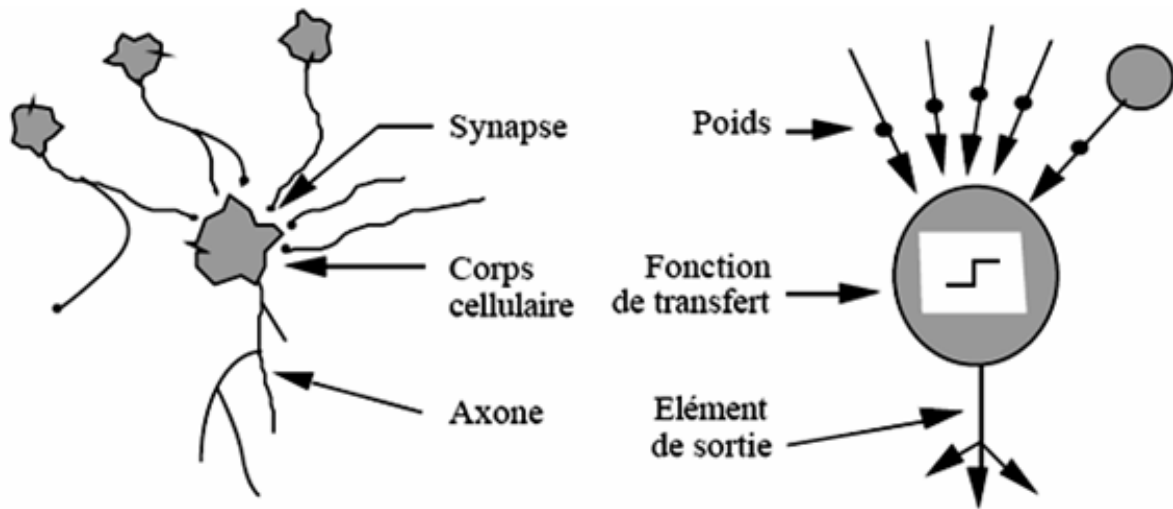


Figure II-3 : La correspondance entre le neurone biologique et le neurone artificiel

➤ **Comportement** (Touzet, 1992)

On distingue deux phases. La première est habituellement le calcul de la somme pondérée des entrées (a) selon l'expression suivante :

$$\tag{II-2}$$

A partir de cette valeur, une fonction de transfert calcule la valeur de l'état du neurone. C'est cette valeur qui sera transmise aux neurones avals. Il existe de nombreuses formes possibles pour la fonction de transfert. Les plus courantes sont présentées sur la figure 4. On remarquera qu'à la différence des neurones biologiques dont l'état est binaire, la plupart des fonctions de transfert sont continues, offrant une infinité de valeurs possibles comprises dans l'intervalle $[0, +1]$ (ou $[-1, +1]$).

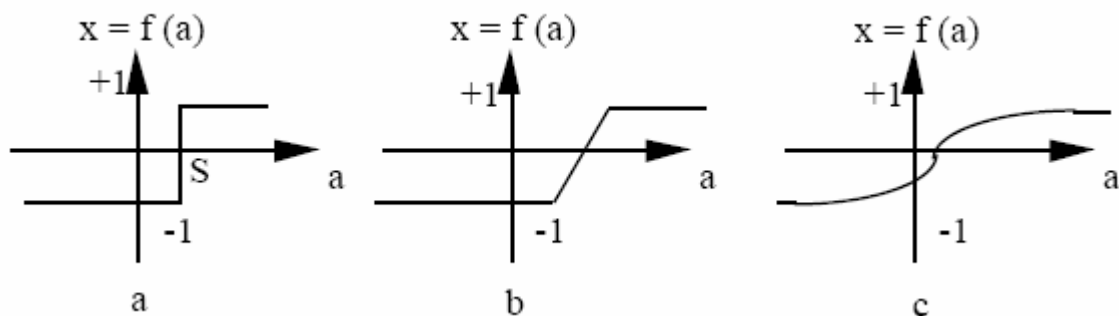


Figure II-4 : Différents types de fonctions de transfert pour le neurone Artificiel

- a : fonction à seuil (S, la valeur du seuil) ;
 b : linéaire par morceaux ;
 c : sigmoïde.

➤ Limites du neurone formel

Ce modèle n'a pas pour rôle de représenter toutes les caractéristiques du neurone biologique:

- Le processus de transport par l'axone et la modulation synaptique sont cruellement réduit à une simple multiplication de deux vecteurs $\omega \cdot x$;
- La dynamique électrochimique complexe, à la source de la propagation du signal, est ramenée à sa plus simple expression par une fonction statique, f ;
- Il n'y a pas de dynamique interne au sein de ce modèle; pour des poids synaptiques fixes, le neurone formel accomplit une transformation statique.

Dans un sens, ce modèle ne capture que les propriétés spatiales du neurone biologique et passe outre les relations temporelles.

La capacité représentative du neurone formel est fortement restreinte en raison de la forme rudimentaire de la fonction f utilisée (seuil), (Yahiaoui et Ariat, 2002) .

En résumé, les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique de réseaux est évidemment un réseau. La capacité de traitement d'un réseau est stockée sous forme de **poids d'interconnexions** obtenus par un processus d'**adaptation** ou d'**apprentissage** à partir d'un ensemble d'exemples d'apprentissage.

II-4- Propriétés de réseaux de neurones

Un réseau de neurones se compose de neurones qui sont interconnectés de façon à ce que la sortie d'un neurone puisse être l'entrée d'un ou plusieurs autres neurones. Ensuite il y a des entrées de l'extérieur et des sorties vers l'extérieur [Rumelhurt et al. 1986].

Rumelhurt et al. donnent huit composants principaux d'un réseau de neurones

- Un *ensemble de neurones*
- Un *état d'activation* pour chaque neurone (actif, inactif, ...)
- Une *fonction de sortie* pour chaque neurone ($f(S)$)
- Un *modèle de connectivité* entre les neurones (chaque neurone est connecté à tous les autres, par exemple)
- Une *règle de propagation* pour propager les valeurs d'entrée à travers le réseau vers les sorties
- Une *règle d'activation* pour combiner les entrées d'un neurone (très souvent une somme pondérée)

- Une règle d'apprentissage
- Un environnement d'opération (le système d'exploitation, par exemple)

Le comportement d'un réseau et les possibilités d'application dépendent complètement de ces huit facteurs et le changement d'un seul d'entre eux peut changer le comportement du réseau complètement.

Les réseaux de neurones sont souvent appelés des "**boîtes noires**" car la fonction mathématique qui est représentée devient vite trop complexe pour l'analyser et la comprendre directement. Cela est notamment le cas si le réseau développe des représentations distribuées [Rumelhart et al. 1986], c'est à dire que plusieurs neurones sont plus ou moins actifs et contribuent à une décision. Une autre possibilité est d'avoir des représentations localisées, ce qui permet d'identifier le rôle de chaque neurone plus facilement. Les réseaux de neurones ont quand même une tendance à produire des représentations distribuées (Framling, 1992).

➤ La non linéarité

Cette notion est extrêmement importante puisque le fonctionnement des réseaux neuronaux ressemble beaucoup au fonctionnement d'autre système, si l'on élimine la fonction d'amplification non linéaire de chaque neurone. Cette non linéarité intrinsèque, et parfois récurrente, autorise les réseaux neuronaux à modéliser des associations très complexes.

La non linéarité de type seuil est génératrice de capacité de généralisation : cela est simple à concevoir intuitivement, puisque la sortie seuillée d'un neurone ne peut prendre que deux valeurs alors que cette même sortie non seuillée est multivaluée. En vocabulaire mathématique, on l'exprime par le fait que le neurone de **Mc Culloch et Pitts** sépare l'espace d'entrée par un hyper plan. Cela est une capacité intrinsèque de généralisation (Yahiaoui et Ariat, 2002).

II-5- Apprentissage des réseaux de neurones artificiels

Dans le cas d'un cerveau humain, l'utilisation correcte des neurones nécessite au moins une phase d'apprentissage qui permettra d'avoir le fonctionnement escompté. Cet apprentissage consiste à jouer sur les synapses qui relient les neurones. De la même manière, les réseaux de neurones artificiels que nous venons de décrire ont besoin d'apprendre (Le Fablec, 1999).

L'apprentissage est une phase du développement d'un réseau de neurones. C'est un changement du comportement du réseau de façon à lui permettre de se rapprocher d'un but défini. Ce but est normalement l'approximation d'un ensemble d'exemples ou l'optimisation de l'état du réseau en fonction de ses poids pour atteindre l'optimum d'une fonction économique fixée a priori.

Il faudra Là aussi agir sur les synapses. Pour cela, on devra régler les valeurs des poids de chaque connexion entre neurones ainsi que les seuils (ou biais).

L'apprentissage des réseaux de neurones n'est donc qu'une procédure qui consiste à estimer les paramètres des neurones d'un réseau, afin que celui-ci remplisse au mieux la tâche qui lui est affecté.

Il existe trois types d'apprentissages principaux. Ce sont l'apprentissage supervisé, non supervisé, et semi supervisé.

II-5-a- Apprentissage supervisé

Il consiste à apprendre une association particulière entre deux ensembles de vecteurs. Durant une première phase, appelée *phase d'apprentissage*, on présente au réseau des vecteurs d'entrées en lui imposant, comme des contraintes, les vecteurs de sortie que l'on désire lui faire associer à ces entrées. Pendant une seconde phase, appelée *phase de test*, on présente au réseau de nouveaux vecteurs qui n'ont pas servi à l'apprentissage et l'on observe ses réponses. Les performances du réseau sont alors évaluées par le pourcentage d'erreurs.

L'intérêt des réseaux neuronaux réside précisément dans leur capacité intrinsèque à la généralisation, c'est-à-dire leur aptitude à bien se comporter sur des vecteurs qu'ils n'ont pas appris (Yahiaoui et Ariat, 2002).

Pour ce type d'apprentissage la règle la plus utilisée est celle de **Widrow-Hoff**. D'autres règles d'apprentissage existent telle que la règle de Hebb, la règle du perceptron, la règle de Grossberg etc.

➤ La loi de Hebb :

Si deux unités connectées sont actives simultanément, le poids de leur connexion est augmenté ou diminué. R est une constante positive qui représente la force d'apprentissage (*learning rate*).

$$\begin{aligned} \Delta w_{ij} &= R a_i a_j \\ a_i &= -1 \quad a_j = 1 \\ a_j &= -1 \quad \Delta W_{ij} = R \quad \Delta W_{ij} = -R \\ a_j &= 1 \quad \Delta W_{ij} = -R \quad \Delta W_{ij} = R \end{aligned}$$

➤ La loi de Widrow-Hoff

La règle d'apprentissage de Widrow-Hoff est une règle qui permet d'ajuster les poids d'un réseau de neurones pour diminuer à chaque étape l'erreur commise par le réseau de neurones (à condition que le facteur d'apprentissage soit bien choisi).

Un poids est modifié en utilisant la formule suivante:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \alpha \delta_k \mathbf{x}_k$$

où:

- \mathbf{w}_k est le poids à l'instant k ,
- \mathbf{w}_{k+1} le poids à l'instant $k+1$,
- α est le facteur d'apprentissage,
- δ_k caractérise la différence entre la sortie attendue et la sortie effective du neurone à l'instant k .
- \mathbf{x}_k la valeur de l'entrée avec laquelle le poids \mathbf{w} est associé à l'instant k .

Ainsi, si δ_k et \mathbf{x}_k sont positifs tous les deux, alors le poids doit être augmenté. La grandeur du changement dépend avant tout de la grandeur de δ_k mais aussi de celle de \mathbf{x}_k . Le coefficient α sert à diminuer les changements pour éviter qu'ils deviennent trop grands, ce qui peut entraîner des oscillations du poids.

Deux versions améliorées de cet apprentissage existent, la version "par lots" et la version "par inertie" (momentum en anglais), dont l'une utilise plusieurs exemples pour calculer la moyenne des changements requis avant de modifier le poids et l'autre empêche que le changement du poids au moment k ne devienne beaucoup plus grand qu'au moment $k-1$ (Framling, 1992).

➤ **Loi de Grossberg :**

On augmente les poids qui entrent sur l'unité gagnante a_i s'ils sont trop faibles, pour les rapprocher du vecteur d'entrée a_j . C'est la règle d'apprentissage utilisée dans les cartes auto organisatrices de Kohonen.

$$\Delta w_{ij} = R(a_j - w_{ij})a_i$$

II-5-b Apprentissage non supervisé

Il consiste à découper l'ensemble des vecteurs d'entrées en classes d'équivalence, sans qu'il soit nécessaire de donner au réseau neuronal les noms de classes pour chaque exemple. Cette classification automatique ressemble à certaines méthodes utilisées en analyse de données.

La séparation en classes d'équivalence s'opère par mesure de ressemblance entre les vecteurs proposés. La contrainte que l'on doit imposer au réseau neuronal est le nombre de classes d'équivalence.

Pendant la phase d'apprentissage, le réseau neuronal construit une topologie de l'espace des vecteurs d'entrées.

L'intérêt de cette approche est que l'on n'a pas besoin de disposer d'exemples de problèmes résolus (c'est-à-dire de vecteurs de sortie correspondant à chaque vecteur d'entrée utilisé pour l'apprentissage).

L'inconvénient est que le critère de mesure de ressemblance ne correspond pas forcément avec ce que l'on veut (Yahiaoui et Ariat, 2002).

II-5-c Apprentissage semi supervisé

C'est un apprentissage de type essai-erreur où le réseau donne une solution et est seulement alimenté avec une information indiquant si la réponse était correcte ou si elle était au moins meilleure que la dernière fois.

II-6- Architecture des réseaux de neurones

Définir l'architecture d'un RNA revient à faire un choix judicieux du vecteur d'entrée, de la taille du réseau (ou nombre total de couches et de neurones), de la structure du réseau (c.-à-d. le type d'interconnexion entre les couches de neurones) et des fonctions de transfert, sans disposer de procédure standard (à présent). Le choix de la taille du réseau incombe à l'utilisateur. Il n'existe pas dans la littérature d'éléments suffisants pour déterminer objectivement le nombre de neurones requis pour résoudre un problème donné. Les approches heuristiques proposées par différents auteurs (Hecht-Nielsen 1987; Baum et Haussler 1989; Widrow et Lehr 1990; Zurada 1992; Murata et al. 1994; Kohavi 1995; Jacobs et al. 1997; Rudolph 1997) sont toutes expérimentales et ne s'appliquent qu'à des cas particuliers. La

définition de la taille du réseau et de sa structure sont deux étapes difficiles de la construction d'un RNA (Rumelhurt et al. 1986; Hassoum 1995), (Coulibaly et al 1999).

Selon leur structure, on peut classer les réseaux de neurones en deux grandes catégories :

- Les réseaux « feedforward » (statiques)
- Les réseaux « feedback » (dynamiques ou récurrents)

II-6-1- Les réseaux feedforward

On dit d'un réseau qu'il est feedforward, ou encore à sens unique, lorsqu'il ne contient pas de boucle interne, c'est à dire lorsque l'information qui le traverse ne circule que de l'entrée vers la sortie. Les neurones sont organisés en couches successives, les neurones d'une couche étant reliés à ceux de la suivante. Parmi ce type de réseaux, il existe les réseaux à deux couches. Ils sont composés d'un ensemble de neurones répartis sur deux couches distinctes appelées couche d'entrée et de sortie, les neurones de la première étant connectés à ceux de la seconde (**figure II-6**).

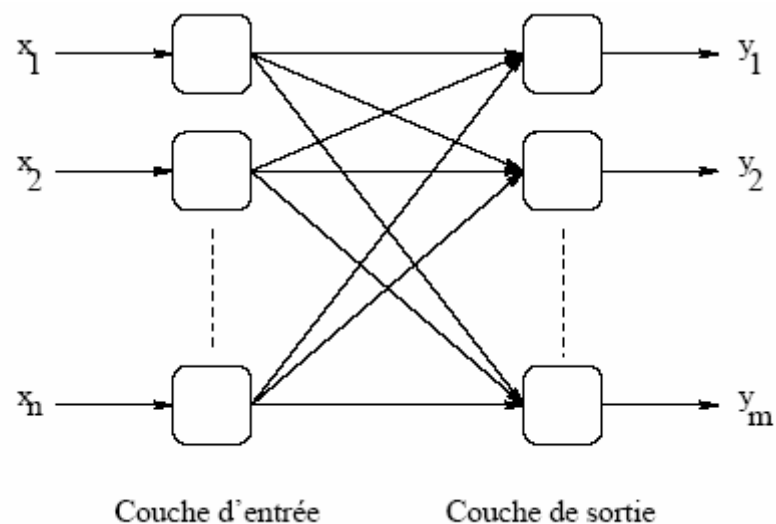


Figure II-6 réseau feedforward à deux couches

Si l'on intercale entre la sortie et l'entrée une ou plusieurs couches (qui seront alors appelées couches cachées) (**figure II-7**). On parle alors de réseau feedforward multicouches (perceptron multicouches).

L'intérêt de rajouter au moins une couche cachée est d'augmenter le nombre de connexions, ce qui accroît la capacité d'un réseau à extraire l'information des données fournies en entrée. (Le Fablec 1999)

Les perceptrons multicouches sont appelés souvent des approximateurs universels et ceci est grâce à leur capacité de faire l'approximation de n'importe quelle fonction mathématique.

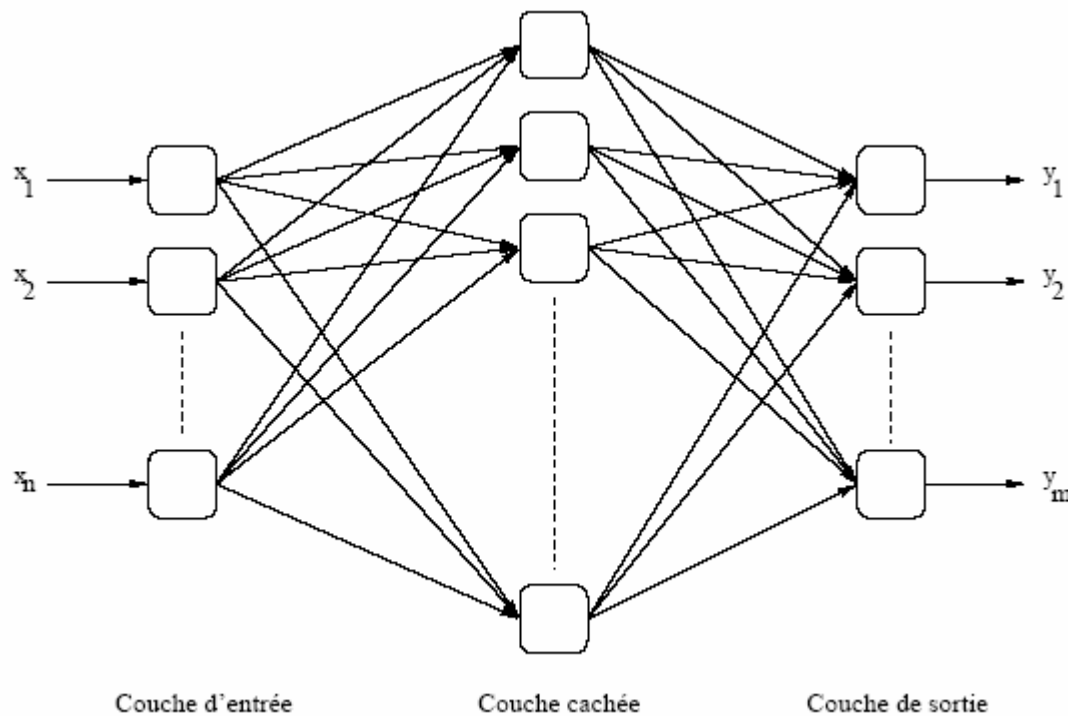


Figure II-7 : réseau feedforward à trois couches

II-6-2 Réseaux récurrents

On parle de réseau de neurones récurrent lorsqu'il existe une boucle au moins dans l'ensemble des connexions. Il y a alors au minimum un feedback d'une couche sur la précédente. La figure II-8 en donne un exemple très simple : un neurone de la couche de sortie renvoie sa sortie sur la couche d'entrée avec un délai. Cela permet d'utiliser une des sorties à l'instant t comme entrée à l'instant $t + 1$. (Le Fablec 1999)

Les réseaux récurrents sont souvent utilisés pour la modélisation et la prévision des systèmes dynamiques ainsi que les séries temporelles.

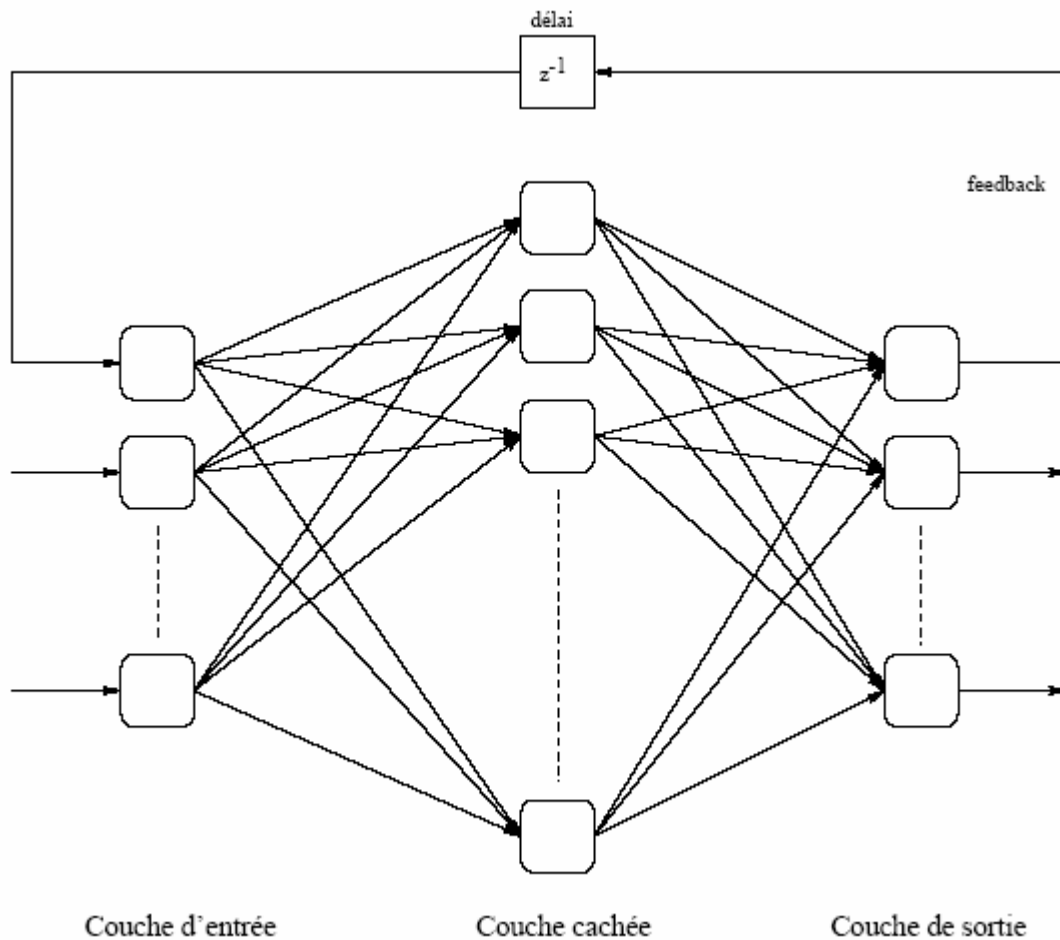


Figure II-8 : réseau feedback simple

Certains auteurs ajoutent à ces deux architectures une troisième : les *réseaux auto organisés*.

- Les réseaux de neurones auto organisés sont des réseaux qui changent leurs structures internes pendant l'apprentissage. Ainsi les neurones se regroupent topologiquement suivant la représentation des exemples. Ces réseaux sont des dérivées des modèles de Kohonen. (Mokhnache 2004)

II-7- Quelques types de réseaux de neurones

II-7-1- Les réseaux monocouches

➤ Le perceptron

Introduit par Rosenbalt en 1958, sur la base du modèle du neurone formel de Mc Culloch et Pitts en 1943 ainsi de la première loi de Hebb en 1949, le perceptron représente le premier réseau de neurones artificiels. Il est très utilisé pour le classement comme il est introduit dans les réseaux traitant la décision.

Il se compose de deux couches de neurones formels appelées « rétine » et « couche de sortie ». La couche rétine est une contraction de la couche d'entrée et de la couche de traitement des

informations. Cette contraction a été effectuée pour des raisons de simplification, car la couche d'entrée est une couche inerte qui ne participe pas directement au traitement de l'information. Les cellules de la rétine, lorsqu'elles sont excitées, répondent par « oui » ou par « non » (1 ou 0) à leur sortie, car elles sont régies par une fonction binaire à seuil (Davalò, 1991).

La fonction de transfert utilisée dans ce type de réseau est la fonction à seuil (*hardlimit*). La modification des poids du perceptron est effectuée avec un apprentissage de type supervisé en exploitant la loi d'apprentissage de Widrow-Hoff (1960), décrite dans le paragraphe II-5-b. L'apprentissage supervisé consiste à présenter au réseau les vecteurs d'entrée et de sortie couplés.

L'apprentissage se fait en trois étapes :

1. Présentation du vecteur d'entrées E ;
2. Calcul de l'erreur Δ en sortie par la formule suivante :

$$\Delta = T - \sum (W_j \cdot E_j) \quad (\text{II- 3})$$

Où

Δ : est l'erreur ;

$\sum (W_j \cdot E_j)$: Entrée pondérée du neurone ;

T : Vecteur de la sortie désirée.

3. Modification des poids de connexions :

$$\Delta W_{ij} = K \cdot \Delta \cdot E_j \quad (\text{II- 4})$$

Où

ΔW_{ij} : Modification à apporter aux poids ;

K : Constante numérique aléatoire ;

Δ : Erreur calculée en sortie ;

E_j : Activation du neurone j.

Limite du perceptron

L'inconvénient du perceptron est qu'il n'est efficace que pour les problèmes linéairement séparables, car l'algorithme d'apprentissage ne peut pas converger dans le cas contraire. C'est-à-dire que les vecteurs d'entrées doivent être classés en deux classes. Si W_0 est le seuil du neurone, on aura donc :

La réponse = 1 si $\sum (W_j \cdot E_j) > W_0$

La réponse = 0 si $\sum (W_j \cdot E_j) < W_0$

C'est pourquoi le perceptron ne peut pas résoudre le problème du ou exclusif (XOR).

Table II-1 : Table de la loi XOR :

X	Y	XOR(X, Y)
1	1	0
1	0	1
0	1	1
0	0	0

Un réseau de neurone à seuil égale à 1, devrait répondre par 0 à l'entrée {0,0} (et c'est ce qu'il fait) et par 1 à l'entrée {1,1}. Mais cela ne répond pas à la réalité.

La séparation linéaire obtenue par le réseau serait :

Les entrées {1,0} et {0,1} dans la classe 1 → ayant une réponse 1 ;

Les entrées {0,0} et {1,1} dans la classe 2 → ayant une réponse 0. (Lahsen 2005)

Cette problématique a limité le champ d'application du perceptron monocouche. Par conséquent, la recherche dans le domaine des réseaux de neurone a été arrêtée pendant 20 ans, pour revoir la lumière dans les années 80 après le développement de l'algorithme de rétropropagation de l'erreur.

➤ Adaline et Madaline

Introduit par Widrow et M. Hoff en 1960 [Widrow 60], le modèle d'Adaline (Adaptive Linear Element) a la même architecture que le perceptron. Une couche d'entrée est connectée à une plusieurs sorties. Dans ce dernier cas, il est appelé Madaline (Multiple Adaptive Linear Element). La différence entre ce réseau et le perceptron, est qu'il est destiné à l'approximation des fonctions. Sa sortie est linéaire. (Mokhnache 2004)

II-7-2- Les réseaux multicouches

Contrairement aux deux modèles de réseaux de neurones précédents, les réseaux multicouches sont constitués de plusieurs couches disposées l'une après l'autre. La fonction d'activation utilisée dans ce type de réseau est généralement de type sigmoïde, ce qui les donne un caractère non linéaire. L'apprentissage dans ce cas consiste à modifier des connexions ayant un nombre très grand, du fait de la multiplicité des couches ainsi que celle des neurones dans chaque couche. Après l'intégration de l'algorithme d'apprentissage de la rétropropagation (back-propagation) dans les réseaux de neurones, la tâche du traitement de problèmes très complexes est devenue moins délicate.

La topologie de ce type de réseaux est telle qu'avant d'arriver à une sortie, chaque unité d'entrée est suivie par une succession de couches cachées et chaque unité d'une couche donnée reçoit son entrée à partir de la couche précédente seulement (Mokhnache 2004).

➤ Algorithme de back-propagation

Cet algorithme est une généralisation aux réseaux de neurones formels multicouches du modèle de Widrow-Hoff. Il a été mis au point par deux équipes simultanément : une équipe française (Fogelman-Soulié, Gallinari, Lecun) et une équipe américaine (Rumelhurt, Hinton, Williams). Son idée consiste en l'utilisation d'une fonction sigmoïde dérivable en remplacement de la fonction à seuil couramment usité. La fonction sigmoïde prise en général est (Davallo, 1991) :

$$f(x) = \frac{1}{1 + e^x} \quad (\text{II-5})$$

L'algorithme permet de réinjecter l'erreur, calculée en sortie, vers les couches internes, en suivant le chemin inverse du signal provenant de la couche d'entrée et allant vers la couche de sortie.

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

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

b- Block adaptive learning (Batch learning)

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

L'algorithme de base de la back-propagation est celui de la descente du gradient (Mokhnache, 2004).

II-7-3- Les réseaux à fonction de base radiale (RBF)

Ce sont des réseaux à une seule couche cachée. Ils sont très utilisés pour la classification et l'approximation de fonctions, à travers une combinaison linéaire de fonctions non linéaires à base radiale (Mokhnache 2004). Parmi ces fonctions, la fonction gaussienne, ayant comme formule celle de II-6, est la plus utilisée.

$$f(x) = \exp\left(\frac{-x^2}{k^2}\right) \quad (\text{II-6})$$

Les modèles précédents sont de type « feedforward », c'est-à-dire qu'ils ne contiennent aucun rebouclage.

II-7-4- Les réseaux récurrents

Deux modèles très connus dans cette catégorie de réseau et qui sont : le réseau d'Elman et celui de Hopfield :

➤ Le réseau d'Elman

Il s'agit d'une structure multicouche (figure II-9) où les boucles de rétro-action relient la couche cachée avec les cellules d'entrée. La couche d'entrée consiste en deux ensembles de cellules, les cellules de contexte et les cellules d'entrée. Il y a autant de cellules cachées que de cellules de contexte. Chaque cellule cachée est reliée à une seule cellule de contexte par une connexion fixe, de poids égal à 1. L'état interne est défini par l'ensemble des cellules de la couche cachée.

Comportement du réseau: La couche cachée a la tâche d'établir, à la fois, les associations de la fonction de sortie et les associations de la fonction de transition.

A chaque instant, une sortie est calculée sur la base de l'état interne (contexte) et de l'entrée présente. Il y a copie du vecteur d'activation de la couche cachée sur la couche de contexte. De cette façon, le résultat intermédiaire du traitement à la date précédente ($t-1$) peut influencer le traitement à la date actuelle (t).

L'algorithme d'apprentissage : C'est la rétropropagation de gradient. Les vecteurs d'activation de la couche cachée développés par le réseau sont significatifs du problème soumis. Ces vecteurs d'activation sont les représentations internes construites par le réseau. (Touzet 1992)

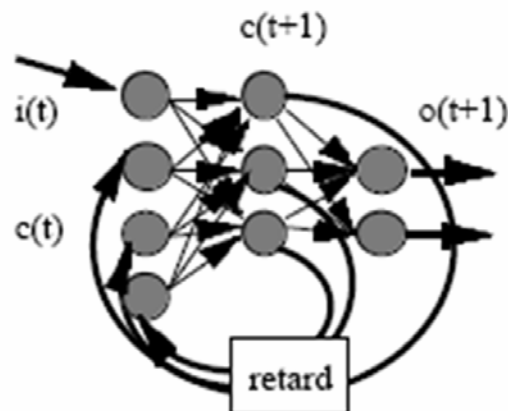


Figure II-9 : Le modèle d'Elman

L'algorithme d'apprentissage du réseau d'Elman peut s'effectuer comme suit

1. Initialiser ($e=0$) les unités de contexte à 0,
2. Présenter l'exemple X_e , effectuer le calcul en avant (forward),
3. Appliquer l'algorithme de rétro propagation,
4. Poser $e=e+1$; revenir à 2 jusqu'à ce qu'une convergence soit obtenue.

Les unités de contexte à l'étape e possèdent toujours la valeur d'activation à l'étape $e-1$ des unités cachées.

➤ Le réseau de Hopfield (mémoire associative)

Le réseau de Hopfield est un réseau avec des sorties binaires où tous les neurones sont interconnectés avec des poids symétriques, c'est à dire que le poids du neurone N_i au neurone N_j est égal au poids du neurone N_j au neurone N_i . Les poids sont donnés par l'utilisateur. (Framling 1992)

A partir des représentations partielles ou bruitées, les réseaux de Hopfield trouvent le modèle le plus proche de la réalité parmi les modèles stockés : c'est le principe de la mémoire associative.

L'apprentissage appliqué dans ce type de réseau est non supervisé.

Ce modèle est souvent utilisé pour la résolution des problèmes d'optimisation.

II-7-5- Les cartes auto organisatrices de Kohonen

Kohonen s'est inspiré d'observations biologiques pour mettre au point son architecture. En effet, dans le cortex existe une carte topologique de notre corps, où chaque groupe de neurones correspond à une partie du corps (Yahiaoui 2002)

Un réseau de Kohonen est constitué d'une couche compétitive unique de neurones de sortie dont la compétition est basée sur le principe « tout au vainqueur ». L'apprentissage dans la couche est non supervisé et est caractérisé par une généralisation au voisinage du neurone élu. L'algorithme de Kohonen représente une simplification algorithmique de la loi d'activation compétitive dans une couche de neurones totalement interconnectés (Mokhnache 2004).

II-8-Utilisation des réseaux de neurones

Il est avantageux de mettre en œuvre les réseaux de neurones pour toute application nécessitant de trouver, par des méthodes **statistiques**, une relation **non linéaire** entre des données numériques.

Le problème posé par l'ingénieur doit satisfaire certaines **conditions** afin de le pouvoir résoudre à l'aide des réseaux de neurones :

- La première condition, nécessaire et non suffisante: il faut disposer d'un échantillon de taille suffisamment grande, et bien représentatifs.
- En l'absence de toute connaissance a priori sur l'intérêt d'un modèle non linéaire, l'ingénieur doit d'abord utiliser les méthodes simples et éprouvées d'élaboration d'un modèle linéaire (qui est simple et moins coûteux en temps de calcul que celui d'un réseau de neurone). S'il s'avère que la précision du modèle est insuffisante bien que tous les facteurs pertinents soient présents dans les entrées du modèle, alors le concepteur de modèle doit envisager la mise en œuvre de modèles non linéaires tels que les réseaux de neurones.
- Si les données sont disponibles et s'il on s'est assuré qu'un modèle non linéaire est utile, il faut s'interroger sur l'opportunité d'utiliser un réseau de neurone de préférence à une autre famille de fonction non linéaire, les polynômes par exemple. Il est prouvé que les réseaux de neurones sont plus avantageux que le nombre de variables est « *grand* », c'est-à-dire qu'il est supérieur ou égal à trois.

En résumé, les réseaux de neurones peuvent être utiles pour résoudre un problème à régression non linéaire possédant au moins trois variables, sachant que les données numériques disponibles sont suffisamment nombreuses et représentatives (Dreyfus, 2002).

Si les données ne sont pas numériques (mais linguistique par exemple), il faut avoir recours à des prétraitements permettant de « *quantifier* » ces données (par exemple à l'aide de techniques issues de la théorie des ensembles flous).

II -9- La mise en œuvre des réseaux de neurones

Les réseaux de neurones réalisent des fonctions non linéaires paramétrées ; ils peuvent approcher toute fonction de régression si l'on trouve un réseau de neurones qui rend minimum la fonction de coût des moindres carrés (somme des carrés des différences entre les valeurs calculées par le réseau de neurones et les valeurs mesurées). La mise en oeuvre d'un réseau de neurones nécessite donc de

- Déterminer les entrées pertinentes, c'est-à-dire les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser,

- Collecter les données nécessaires à l'apprentissage et à l'évaluation des performances du réseau de neurones,
- Trouver le nombre de neurones cachés nécessaires pour obtenir une approximation satisfaisante,
- Estimer les valeurs des paramètres correspondant à un minimum de la fonction de coût, c'est-à-dire d'effectuer un apprentissage,
- Evaluer les performances du réseau de neurones à l'issue de l'apprentissage.

En fonction des résultats obtenus, il peut être nécessaire d'effectuer plusieurs itérations de tout ou partie de cette procédure.

II-10- Les avantages des réseaux de neurones

Un des intérêts majeurs des réseaux de neurones artificiels (informatiques) est qu'ils sont organisés sous la forme d'une structure massivement parallèle. De plus, on peut citer aussi les avantages suivants (similaires à ceux des réseaux de neurones biologiques) :

- Les neurones sont des composants non linéaires, par conséquent le réseau aussi : ceci est une propriété très importante car cela leur permet de traiter des problèmes physiques non linéaires ;
- Ils sont capables d'apprendre et de généraliser : après une période d'apprentissage qui consiste à leur présenter des entrées ainsi que les sorties que l'on voudrait qu'ils trouvent, ils sont capables de produire des sorties "raisonnables" sur des cas qu'ils n'ont pas appris ;
- S'il a appris à travailler dans un environnement précis, on peut réapprendre un réseau pour qu'il s'adapte à un autre environnement semblable ;
- Dans le cas d'un problème de classification un réseau de neurones ne se contente pas de faire un choix : il fournit aussi un niveau de confiance sur sa décision. Cela permet de rejeter des motifs ambigus et d'avoir une idée sur la précision du résultat ;
- Si l'on considère un réseau connecté physiquement qui subit un dommage qui n'est pas trop important alors les réponses ne sont pas trop dégradées. Le réseau est donc résistant car il continue à être utilisable bien qu'il soit endommagé (Le Fablec, 1999).

Chapitre III : L'analyse en composantes principales

:

III-1- Introduction :

L'analyse des données a vu au cours de ces dernières années son domaine d'application s'étendre considérablement, et une effervescence bien légitime règne dans le champ de recherche qu'elle a couvert.

Les premières analyses à plusieurs variables remontent au tout début du siècle dernier (Pearson 1901, Spearman 1904, Fisher 1915). Les bases théoriques de presque toutes celles utilisées actuellement ont été établies, de façon quasi définitive vers 1930 (Hotelling 1933 et 1936, Mahalanobis 1936) avec des compléments importants après 1950 (Burt 1950, Anderson 1958, Rao 1964).

Cependant faute de moyens de calcul suffisants, ces méthodes n'ont pu être utilisées que de façon extrêmement limitée.

Jusqu'aux années soixante, elles étaient perfectionnées et s'enrichissaient de variantes, mais toutes restaient inabordables pour les praticiens, car elles nécessitaient une masse considérable de calcul (Benzecri 1982).

Elles ne sont devenues d'une application courante, que grâce au développement de l'informatique. Dès leur apparition en effet, les ordinateurs ont permis l'existence et l'exploitation de grands fichiers de données ainsi que la résolution de problèmes numériques devant lesquels l'homme seul est impuissant.

Ces analyses ont fréquemment le caractère d'une démarche exploratoire du statisticien et du chercheur. Il s'agit le plus souvent d'obtenir une description statistique pour un certain phénomène qui a donné lieu au recueil de mesures ou d'observations trop nombreuses et dépendantes les unes des autres pour être interprétables en première lecture. La simplification de grands tableaux d'informations ou de données s'avère un outil de synthèse remarquable (Jambu, 2000).

Enfin, cette analyse est devenue une technique de représentation des données, ayant un caractère optimal selon certains critères algébriques et géométriques spécifiés, que l'on utilise en général sans référence à des hypothèses statistiques ou à un modèle particulier (Delagarde 1983).

La plus simple et la plus connue des techniques d'analyse de données, c'est l'Analyse en Composantes Principales (ACP), qui a pour objectif de remplacer les variables initiales (numériques) par de nouvelles variables (également numériques) dites "Composantes Principales".

III-2 Définitions :

L'étude des phénomènes climatologiques est une tâche très laborieuse vue la masse importante de données à traiter, la démarche basée sur l'analyse en composantes principales (ACP), qui est une technique d'analyse des données, permet d'extraire l'essentiel de l'information contenue dans les tableaux des données, et fournit une représentation imagée se prêtant plus aisément à l'interprétation. Celle-ci a trois objectifs principaux :

- Le premier consiste à remplacer les variables initiales qui sont généralement plus ou moins liées entre elles, par des variables non corrélées, appelées Composantes Principales, (CP).
- Le second revient à réduire autant que possible le nombre des nouvelles variables (CP) à prendre en considération, toute en conservant la quasi totalité de la variance de l'ensemble des variables initiales (Duband 1989).
- Le troisième est de donner une interprétation particulière aux CP les plus importantes parmi les variables initiales.

III-2-1 Notions d'individu et de caractère :

Dans une série de données on distingue généralement deux ensembles : les individus et les caractères relatifs à ces individus. Le terme "individu" peut désigner selon les cas : une année d'observation ou une autre unité de temps. L'ensemble des individus peut provenir d'un échantillonnage dans une population ou il peut s'agir de la population toute entière (cas rare surtout en hydrologie).

- L'individu "i" est décrit par le vecteur appartenant à \mathbb{R}^P :

$$X_i = \{X_{ij} / j = 1 \text{ à } P\}$$

Sur un individu, on relève un certain nombre de caractères (dits aussi variables) désignant en général un paramètre intervenant dans un phénomène complexe à étudier (exemple, la répartition spatiale ou temporelle des précipitations dans un bassin versant).

- Le caractère (ou variable) "j" est décrit par le vecteur de \mathbb{R}^N :

$$X_j = \{X_{ij} / i = 1 \text{ à } N\}$$

Le terme X_{ij} est un nombre réel qui représente la mesure de la variable X_j sur l'individu i .

Si l'ensemble des individus doit être homogène, l'ensemble des variables peut être hétérogène.

III-2-2 Définition algébrique :

L'ACP considère P variables pour lesquelles on dispose de N individus, donc la matrice $[X]$ résultant du croisement « $N \times P$ » est la matrice de données $[X]$ tel que :

$$[X] = \begin{pmatrix} X_{11} & X_{12} & \dots & \dots & X_{1j} & \dots & \dots & X_{1P} \\ X_{21} & X_{22} & \dots & \dots & X_{2j} & \dots & \dots & X_{2P} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ X_{i1} & X_{i2} & \dots & \dots & X_{ij} & \dots & \dots & X_{iP} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ X_{N1} & X_{N2} & \dots & \dots & X_{Nj} & \dots & \dots & X_{NP} \end{pmatrix}$$

Pour chaque variable, nous pouvons évaluer sa moyenne \bar{X}_j et son écart type S_j :

$$\bar{X}_j = \frac{1}{N} \sum_{i=1}^N X_{ij} \tag{III-1}$$

$$S_j = \left[\frac{1}{N} \sum_{i=1}^N (X_{ij} - \bar{X}_j)^2 \right]^{1/2} \tag{III-2}$$

Le coefficient de covariance entre les variables X_j et X_k est donné par :

$$Cov(X_j, X_k) = \frac{1}{N} \sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k) \tag{III-3}$$

Le coefficient de corrélation entre les variables X_j et X_k est donné par :

$$Cor(X_j, X_k) = \frac{Cov(X_j, X_k)}{S_j * S_k} = \frac{\sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k)}{\left[\sum_{i=1}^N (X_{ij} - \bar{X}_j)^2 * \sum_{i=1}^N (X_{ik} - \bar{X}_k)^2 \right]^{1/2}} \tag{III-4}$$

Le tableau [X] est remplacé par un tableau [Y] (individus x nouvelles variables) en réduisant le nombre de variables nécessaires pour décrire les individus, avec une perte minimale d'informations, ces nouvelles variables sont appelées composantes principales (ou CP).

Calculer les composantes principales notées C_j revient à déterminer N relations linéaires entre les variables X_j :

$$\begin{cases} C_1 = a_{10} + a_{11} X_1 + \dots + a_{1N} X_N \\ C_2 = a_{20} + a_{21} X_1 + \dots + a_{2N} X_N \\ \dots \\ C_i = a_{i0} + a_{i1} X_1 + \dots + a_{iN} X_N \\ \dots \\ C_N = a_{N0} + a_{N1} X_1 + \dots + a_{NN} X_N \end{cases}$$

C_j : j^{ème} Composante Principale ;

X_j : Vecteur variable initiale ;

a_{ji} : Coefficient du système.

Notons au passage que les termes a_{j0} désignent le vecteur permettant la translation de l'origine de l'ancien repère vers le centre de gravité du nuage de points. Un centrage des données initiales annule les coefficients a_{j0} .

III-2-3 Définition géométrique :

L'Analyse en Composantes Principales est puissante par son support géométrique, la méthode consiste à rechercher un premier axe qui soit le plus proche possible de tous les points au sens des moindres carrés, tel que la somme des carrés des distances des N points à cet axe serait minimale, ou encore la projection de ces derniers sur cet axe possède une dispersion maximale. Cet axe est appelé "axe factoriel".

Un second axe est obtenu après projection des N points sur un hyperplan orthogonal au premier axe, tel que la dispersion des projections des N points sur celui-ci serait toujours maximale, et le procédé se réitère P fois.

Nous obtenons ainsi un nouveau système d'axes défini par les nouvelles variables dites composantes principales.

III-3 Formulation mathématique du problème :

La recherche des composantes principales est faite sous deux contraintes :

- Elles doivent être indépendantes, c'est à dire, prise deux à deux, elles présentent obligatoirement des corrélations nulles.
- Les axes factoriels doivent être déterminés par ordre d'importance décroissante, le premier axe expliquera le maximum de la variance totale tandis que le second expliquera le maximum de la variance résiduelle non expliquée par le premier, jusqu'au dernier axe. Mais l'expérience a montré qu'un nombre Q d'axes nettement inférieur à P suffit pour donner le maximum d'informations.

a) Formulation matricielle :

L'objectif de l'ACP étant de maximiser la variance, la formulation matricielle du problème est la suivante :

Soit les matrices colonnes V_1, V_2, \dots, V_Q de dimension (N x 1) représentant l'hyperplan formé par les axes principaux vérifiant les conditions de normalité et d'orthogonalité :

$$\begin{cases} V_i^p \cdot V_i^p = 1 \\ V_i^p \cdot V_j^p = 0 \end{cases} \rightarrow \begin{cases} i = 1 \rightarrow Q \\ j = 1 \rightarrow Q \end{cases} \quad (\text{III-5})$$

On veut maximiser la quantité :

$$\sum_{j=1}^Q \text{Var}(C_j)$$

$$\text{Sachant que : } \sum_{j=1}^Q \text{Var}(C_j) = V_j^t \cdot [M] \cdot [R] \cdot [M] \cdot V_j \quad (\text{III-6})$$

C_j : Composante principale d'ordre j ;

$[R]$: Matrice des covariances des variables (X_1, X_2, \dots, X_p) ;

$[M]$: Métrique définissant le produit scalaire sur l'espace \mathbb{R}^p .

b) Choix de la métrique :

La métrique $[M]$ possède deux options classiques :

* $[M]=1$: Matrice identité.

La covariance sera utilisée afin de quantifier les relations inter variables, on parlera alors d'une *ACP canonique*.

* $[M]=D_{\frac{1}{\sigma^2}}$

Tel que :

$$D_{\frac{1}{\sigma^2}} = \begin{vmatrix} \frac{1}{\sigma_1^2} & 0 & & & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & & \\ & 0 & \cdot & & \\ & & 0 & \frac{1}{\sigma_i^2} & 0 \\ & & & \cdot & \\ & & & \cdot & 0 \\ 0 & & & 0 & \frac{1}{\sigma_p^2} \end{vmatrix}$$

On utilise généralement cette métrique pour pallier au problème de l'hétérogénéité des caractères (variables), et éviter l'influence du choix d'unité des variables. Dans ce cas on parlera d'ACP normée, elle est équivalente à une ACP canonique effectuée sur des variables centrées réduites.

Les données ainsi transformées se présentent sous forme d'une matrice dont toutes les variables sont de moyenne nulle et d'écart type unité.

III-4 Procédé d'application de l'ACP :

III-4-1 Calcul de la matrice des covariances :

La matrice des covariances, notée $[R]$ est la base de l'ACP, elle est obtenue en appliquant la relation suivante :

$$[R] = (1/n)[X]^t \cdot [M] \cdot [X] \quad (\text{III-7})$$

$[R]$: Matrice de covariance de dimension $(p \times p)$;

$[X]$: Matrice de données ;

$[X]^t$: Matrice transposée de $[X]$;

$[M]$: Métrique.

III-1-4-2 Recherche des axes principaux :

Le but est de construire un nouveau système d'axes avec un minimum de variables assurant un maximum de variance.

a) Recherche du premier axe :

Telle que mentionnée précédemment, la contribution maximale est donnée par le premier axe principal, nous devons donc maximiser la variance relative à celui-ci. La recherche du premier axe principal consiste à résoudre le problème

$$\begin{cases} \text{Max} \text{Var}(C_1) \\ V_1' \cdot [M] \cdot V_1 = 1 \end{cases} \quad (\text{III-8})$$

Nous pouvons exprimer la variance de C_1 à l'aide de la matrice des covariances $[R]$ du vecteur aléatoire $X = (X_1, X_2, \dots, X_j, \dots, X_p)$:

$$\text{Var}(C_1) = V_1' \cdot [M] \cdot [R] \cdot [M] \cdot V_1 \quad (\text{III-9})$$

En utilisant la méthode du multiplicateur de LAGRANGE nous pouvons écrire :

$$L = V_1' [M] \cdot [R] \cdot [M] \cdot V_1 - \lambda_1 (V_1' \cdot [M] \cdot V_1 - 1) \quad (\text{III-10})$$

La dérivée par rapport à V est nécessairement nulle :

$$\partial L / \partial V_1 = 2 \cdot [M] \cdot [R] \cdot [M] \cdot V_1 - 2 \lambda_1 \cdot [M] \cdot V_1 = 0 \quad (\text{III-11})$$

Puisque la matrice $[M]$ est inversible :

$$[R] \cdot [M] \cdot V_1 = \lambda_1 \cdot V_1 \quad (\text{III-12})$$

Donc V_1 est le vecteur propre de la matrice $[R] \cdot [M]$. Il suffit de choisir comme vecteur V_1 le vecteur propre associé à la plus grande valeur propre λ_1 de la matrice $[R] \cdot [M]$ pour maximiser la variance de C_1 .

b) Recherche du second axe :

Nous cherchons à déterminer le vecteur unitaire V_2 tel que la composante C_2 soit de variance maximale et non corrélée à C_1 .

Sachant que :

$$\text{Var}(C_2) = V_2' \cdot [M] \cdot [R] \cdot [M] \cdot V_2 \quad (\text{III-13})$$

$$V_2' \cdot [M] \cdot V_2 = 1 \quad (\text{III-14})$$

$$COV(C_1, C_2) = 0 \quad (\text{III-15})$$

L'expression de $COV(C_1, C_2)$ est donnée par :

$$COV(C_1, C_2) = V_1' \cdot [M] \cdot [R] \cdot [M] \cdot V_2 \quad (\text{III-16})$$

Comme la covariance ne tient pas compte de l'ordre nous pouvons écrire :

$$COV(C_1, C_2) = COV(C_2, C_1) = V_2' [M] [R] [M] V_1 = 0 \quad (\text{III-17})$$

Or nous savons que V_1 est un vecteur propre de $[R] \cdot [M]$ associé à la valeur propre λ_1 . Nous en déduisons que :

$$COV(C_1, C_2) = \lambda_1 V_2' \cdot V_1 = 0 \quad (\text{III-18})$$

Une covariance nulle entre C_1 et C_2 est équivalente à l'orthogonalité des vecteurs V_1 et V_2 :

$$COV(C_1, C_2) = 0 \Leftrightarrow V_1 \cdot V_2 = 0 \quad (\text{III-19})$$

En appliquant la même méthode pour la recherche du deuxième axe, nous aurons :

$$L = V_2' [M] [R] [M] V_2 - \lambda_2 (V_2' [M] V_2 - 1) - \mu_2 (V_2' [M] V_1) \quad (\text{III-20})$$

$$\partial L / \partial V_2 = 2 [M] [R] [M] V_2 - 2 \lambda_2 [M] V_2 - \mu_2 [M] V_1 = 0 \quad (\text{III-21})$$

En simplifiant par $[M]$ nous obtenons :

$$2 [R] [M] V_2 - 2 \lambda_2 V_2 - \mu_2 V_1 = 0 \quad (\text{III-22})$$

Nous multiplions à gauche par $V_1' \cdot [M]$ l'équation (III-22) s'écrit alors :

$$2 \cdot V_1' [M] [R] [M] V_2 - 2 \lambda_2 V_1' [M] V_2 - \mu_2 V_1' [M] V_1 = 0 \quad (\text{III-23})$$

$$\text{Or (par hypothèse)} \quad V_1' \cdot [M] \cdot V_2 = 0 \quad (\text{III-24})$$

Donc :

$$V_1' \cdot [M] \cdot [R] \cdot V_2 = V_2' \cdot [M] \cdot [R] \cdot [M] \cdot V_1 = \lambda_1 \cdot V_2' \cdot [M] \cdot V_1 = 0 \quad (\text{III-25})$$

Puisque le vecteur V_1 est unitaire:

$$V_1' \cdot [M] V_1 = 1 \quad (\text{III-26})$$

Le multiplicateur de Lagrange μ_2 est donc nul, et nous sommes ramené au problème précédent. Nous pouvons donc énoncer la définition suivante : le second axe est défini par le

vecteur V_2 , vecteur propre unitaire de la matrice $[R] \cdot [M]$ orthogonal à V_1 et associé à la plus grande valeur propre λ_2 inférieure ou égale à λ_1 .

c) Recherche des autres axes :

En itérant le procédé, nous déterminons donc les valeurs propres et les vecteurs propres de la matrice $[R] \cdot [M]$ pour obtenir la $l^{\text{ème}}$ composante principale C_l .

Le vecteur propre unitaire de la matrice $[R] \cdot [M]$ définit le $l^{\text{ème}}$ axe orthogonal à $(V_1, V_2, \dots, V_{l-1})$ et associé à la $l^{\text{ème}}$ plus grande valeur propre λ_l . Nous constatons que la mise en équation de ces règles aboutit aux résultats suivants :

- Nous appelons $l^{\text{ème}}$ vecteur principal : le vecteur propre unitaire V_l de la matrice $[R] \cdot [M]$ associée, qui fournit les coefficients qui pondèrent les variables initiales pour le calcul des composantes principales.
- Nous appelons $l^{\text{ème}}$ axe principal, la droite engendrée par le $l^{\text{ème}}$ vecteur principal.
- Chaque composante C_k est portée par le $k^{\text{ème}}$ axe principal.
- La dispersion des projections des variables sur la composante C_k est mesurée par la valeur propre λ_k .
- Les valeurs sont rangées par ordre décroissant : $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_Q$.
- Les CP sont rangées de 1 à Q dans l'ordre des valeurs propres.
- La moyenne de chaque CP est nulle.

Si nous voulons normer les CP, c'est à dire, imposer à chacune d'entre elles d'avoir un écart type unité il suffit de diviser chacune d'elles par la racine de la variance expliquée $(\lambda_k)^{\frac{1}{2}}$ correspondante.

III-4-2 Calcul des composantes principales :

On désigne par CP, la projection du nuage de points initiale sur le nouveau système d'axes fournit par les vecteurs propres. Notons $[C]$ la matrice des CP.

$$[C] = [A]' [X] \quad (\text{III-27})$$

$[X]$: Matrice des données initiales ;

$[A]'$: Matrice transposée de la matrice composée par les vecteurs propres.

III-5 Principaux résultats :

En général une analyse en composantes principales fournit trois sources de renseignements, toutes nécessaires à l'interprétation :

- Un tableau de vecteurs et valeurs propres.
- Un tableau des corrélations des individus avec les axes factoriels.
- Un tableau des corrélations des variables aux axes principaux.

III-6 Représentation graphique :

Une fois les résultats numériques obtenus nous passons à la représentation graphique :

- Nous pouvons représenter chaque variable par un point dont les coordonnées sont les covariances avec les CP réduites.

$$\text{Cov}(X_j, C_k) = (\lambda)^{1/2} * V_k(j) \quad (\text{III-28})$$

Si nous considérons une variable centrée réduite Y_j au lieu de la variable initiale X_j , les coordonnées de cette nouvelle variable dans le plan factoriel sont calculées par :

$$\text{COV}(Y_j, C_i) = \text{COV}\left(\frac{X_j}{S_j}, C_i\right) = \text{Cor}(X_j, C_i) \quad (\text{III-29})$$

C_1 : Composante principale normée ;
 S_j : Ecart type de la $j^{\text{ème}}$ variable.

Comme la variance de Y_j est égale à 1, cela signifie que son point représentatif se trouve sur une sphère de rayon égal à 1. C'est pourquoi le cercle unité tracé sur chacun des plans étudiés s'appelle cercle de corrélation. La variable sera d'autant mieux expliquée, que son point représentatif se rapproche du cercle et inversement.

- Nous pouvons aussi observer la répartition des individus selon les principaux axes choisis. Ainsi deux individus seront proches dans l'espace R^p , s'ils sont proches dans le plan factoriel. Dans le graphique des individus on s'intéresse aux distances inter-individus qu'on peut interpréter comme étant des ressemblances.

En conclusion, nous pouvons dire que l'analyse en composantes principales, dans son aspect qualitative, permet d'établir deux sortes de bilans :

- Un bilan de liaisons entre les variables pouvant nous renseigner sur les points suivants :
 - Quelles sont les variables qui sont liées positivement, et celles qui s'opposent (liées négativement)?
 - Existe-t-il une répartition en groupes des variables inter corrélées ?
- Un bilan de ressemblance entre individus répondant aux questions suivantes :
 - Quels sont les individus qui se ressemblent et ceux qui diffèrent ?
 - Existe-t-il des groupes homogènes d'individus ?

III-7 Application de l'ACP :

En général, le domaine d'utilisation de L'ACP est très large, sans inconvénients, sans dangers et ne nécessite pas au préalable d'hypothèses restrictives.

Si les conditions de normalité ne sont ni impératives ni nécessaires pour pratiquer l'ACP, le type de variables utilisé influe par contre sur les résultats obtenus.

Nous distinguons deux aspects principaux de l'ACP :

- **L'analyse descriptive** : s'intéresse à la structure de la matrice d'observations, basée principalement sur la représentation graphique des résultats obtenus par l'ACP et ce, afin de visualiser certains problèmes non décelables à l'état brut, vu l'indépendance des variables.
- **L'analyse opérationnelle** : l'ACP fournit un outil mathématique en apparence, mais très bien adapté aux problèmes rencontrés en hydrologie à savoir :
 - Optimisation d'un réseau de mesures.
 - Reconstitution des données.
 - Prévision.
 - Simulation.

Dans le domaine descriptif, deux types de phénomènes peuvent être étudiés : temporel et spatial.

Le premier type met en évidence des relations entre des observations à deux époques distinctes de l'année alors que le second fait ressortir la notion de proximité inter stations.

III-8-Influence de la transformation des variables :

La matrice de corrélation (variable centrée réduite) ou de covariance (variable brute) constitue le point de départ de toute analyse en composantes principales, c'est l'une ou l'autre en effet qui permet après diagonalisation de calculer les valeurs et les vecteurs propres, et par conséquent les composantes principales (CP).

Or la valeur du coefficient de corrélation (ou de covariance) entre les variables X_j et X_k dépend essentiellement des couples d'observations X_{ij} et X_{ik} et par conséquent toute transformation appliquée aux p variables X_j modifiera les coefficients de corrélation (ou de covariance), et de ce fait la structure même des composantes principales.

Il existe des transformations usuelles pour l'analyse des données à savoir :

1-Variable centrée :

$$Y_j = X_j - \overline{X_j} \quad (\text{III-30})$$

Avec : X_j : Variable initiale ;

$\overline{X_j}$: Moyenne de la variable initiale ;

Y_j : Variable transformée (centrée).

2-Variable réduite :

$$Y_j = \frac{X_j}{\sigma_{X_j}} \quad (\text{III-31})$$

Avec : X_j : Variable initiale ;
 σ_{X_j} : Ecart type de la variable initiale ;
 Y_j : Variable transformée (réduite).

3-Variable centrée réduite :

$$Y_j = \frac{X_j - \overline{X_j}}{\sigma_{X_j}} \quad (\text{III-32})$$

Avec : X_j : Variable initiale ;
 $\overline{X_j}$: Moyenne de la variable initiale ;
 σ_{X_j} : Ecart type de la variable initiale ;
 Y_j : Variable transformée (centrée réduite).

L'analyse en composantes principales peut donc être effectuée sur quatre types de variables :

- ACP sur variables brutes (sans transformation) ;
- ACP sur variables centrées ;
- ACP sur variables réduites ;
- ACP sur variables centrées réduites.

Le traitement des variables brutes et des variables centrées donne lieu à la même matrice de covariance, et par conséquent, les résultats de la diagonalisation sont identiques.

En effet, considérons la variable brute X et la variable centrée Y résultant de la transformation (III-30), et montrons que la covariance des deux variables brutes (X_j, X_k) est équivalente à celle des deux variables centrées respectives (Y_j, Y_k).

$$Y_j = X_j - \overline{X_j} \quad \Rightarrow \quad \overline{Y_j} = 0 \quad (\text{III-33})$$

$$Y_k = X_k - \overline{X_k} \quad \Rightarrow \quad \overline{Y_k} = 0 \quad (\text{III-34})$$

Or :

$$\text{Cov}(Y_j, Y_k) = \frac{1}{N} \sum_{i=1}^N (Y_{ij} - \overline{Y_j}) * (Y_{ik} - \overline{Y_k}) = \frac{1}{N} \sum_{i=1}^N Y_{ij} * Y_{ik} \quad (\text{III-35})$$

En remplaçant chaque variable par son expression nous obtenons :

$$\text{Cov}(Y_j, Y_k) = \frac{1}{N} \sum_{i=1}^N Y_{ij} * Y_{ik} = \frac{1}{N} \sum_{i=1}^N (X_{ij} - \overline{X_j}) * (X_{ik} - \overline{X_k}) \quad (\text{III-36})$$

$$= \text{Cov}(X_j, X_k) \quad (\text{III-37})$$

d'où :

$$\text{Cov}(Y_j, Y_k) = \text{Cov}(X_j, X_k) \quad (\text{III-38})$$

Le traitement des variables réduites et centrées réduites se fait de la manière suivante :

Soient Y_j et Z_j les variables obtenues par les transformations (III-31) et (III-32), montrons que la covariance des deux variables réduites (Y_j, Y_k) est égale à celle des deux variables centrées réduites (Z_j, Z_k), qui n'est autre que la corrélation entre les deux variables brutes correspondantes.

$$Y_j = \frac{X_j}{\sigma_{X_j}} \quad \Rightarrow \quad \bar{Y}_j = \frac{\bar{X}_j}{\sigma_{X_j}} \quad (\text{III-39})$$

$$Y_k = \frac{X_k}{\sigma_{X_k}} \quad \Rightarrow \quad \bar{Y}_k = \frac{\bar{X}_k}{\sigma_{X_k}} \quad (\text{III-40})$$

$$Z_j = \frac{X_j - \bar{X}_j}{\sigma_{X_j}} \quad \Rightarrow \quad \bar{Z}_j = 0 \quad (\text{III-41})$$

$$Z_k = \frac{X_k - \bar{X}_k}{\sigma_{X_k}} \quad \Rightarrow \quad \bar{Z}_k = 0 \quad (\text{III-42})$$

Evaluons maintenant les covariances des variables (Y_j, Y_k) et (Z_j, Z_k) :

$$\begin{aligned} \text{Cov}(Y_j, Y_k) &= \frac{1}{N} \sum_{i=1}^N (Y_{ij} - \bar{Y}_j) * (Y_{ik} - \bar{Y}_k) \\ &= \frac{1}{N} \sum_{i=1}^N \left(\frac{X_{ij}}{\sigma_{X_j}} - \frac{\bar{X}_j}{\sigma_{X_j}} \right) * \left(\frac{X_{ik}}{\sigma_{X_k}} - \frac{\bar{X}_k}{\sigma_{X_k}} \right) \\ &= \frac{1}{N} \sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k) \\ &= \frac{\sigma_{X_j} * \sigma_{X_k}}{\sigma_{X_j} * \sigma_{X_k}} \\ &= \frac{\text{Cov}(X_j, X_k)}{\sigma_{X_j} * \sigma_{X_k}} \\ &= \text{Cor}(X_j, X_k) \end{aligned} \quad (\text{III-43})$$

Donc :

$$\text{Cov}(Y_j, Y_k) = \text{Cor}(X_j, X_k) \quad (\text{III-44})$$

$$\text{Cov}(Z_j, Z_k) = \frac{1}{N} \sum_{i=1}^N (Z_{ij} - \bar{Z}_j) * (Z_{ik} - \bar{Z}_k) \quad (\text{III-45})$$

$$= \frac{1}{N} \sum_{i=1}^N Z_{ij} * Z_{ik}$$

$$= \frac{1}{N} \sum_{i=1}^N \left(\frac{X_{ij} - \bar{X}_j}{\sigma_{X_j}} \right) * \left(\frac{X_{ik} - \bar{X}_k}{\sigma_{X_k}} \right)$$

$$= \frac{\frac{1}{N} \sum_{i=1}^N (X_{ij} - \bar{X}_j) * (X_{ik} - \bar{X}_k)}{\sigma_{X_j} * \sigma_{X_k}}$$

$$= \frac{\text{Cov}(X_j, X_k)}{\sigma_{X_j} * \sigma_{X_k}}$$

$$= \text{Cor}(X_j, X_k) \quad (\text{III-46})$$

Donc :

$$\text{Cov}(Z_j, Z_k) = \text{Cor}(X_j, X_k) \quad (\text{III-47})$$

D'où :

$$\text{Cov}(Y_j, Y_k) = \text{Cov}(Z_j, Z_k) = \text{Cor}(X_j, X_k) \quad (\text{III-48})$$

En conclusion nous aurons deux types de variantes à appliquer :

- **ACP brute = ACP centrée.**
- **ACP centrée réduite (ou normée) = ACP réduite.**

L'ACP brute peut être utilisée pour tous les types de variables homogènes c'est à dire possédant les mêmes unités de mesure alors que l'ACP normée peut être utilisée pour des variables homogènes et hétérogènes. En effet en centrant et en réduisant la variable, on élimine le problème de l'hétérogénéité de celle-ci (Bermad, 2004).

Chapitre IV : Collecte et critique des données

Introduction

La performance d'une étude hydrologique, est dépendante en premier lieu de la disponibilité et de la qualité des données nécessaires. Dans le cas de la prévision hydrologique, les modèles proposés exigent un historique de données complet et qui contient le maximum d'informations significatives. En effet, des tests statistiques ont été élaborés afin de traiter et d'analyser les séries temporelles pour en tirer le maximum d'informations. Ainsi, des méthodes peuvent être utilisées pour combler les lacunes des séries tronquées. Dans ce chapitre, on présente la théorie des différents tests et méthodes traitant les séries chronologiques. Et on montre les résultats obtenus pour nos séries après l'application des tests et méthodes en question.

IV- 1- Définition et intérêt d'une série temporelle

C'est la suite d'observations d'une seule variable x durant une période de temps t donné.

En pratique, c'est tout ce qui est chiffrable et varie en fonction du temps. La dimension temporelle est ici importante car il s'agit de l'analyse d'une chronique historique : des variations d'une même variable au cours du temps, afin de pouvoir comprendre la dynamique. La périodicité de la série n'importe en revanche pas : il peut s'agir de mesures quotidiennes, mensuelles, trimestrielles, annuelles, voire même sans périodicité.

De façon générale, il est d'usage de considérer l'intérêt des séries temporelles selon trois perspectives : descriptive, explicative et prévisionnelle.

Description

- L'analyse temporelle permet de connaître la structure de la série de données étudiée ;
- Elle peut être utilisée pour comparer une série à d'autres séries (varicelle et oreillons, par exemple) ;

Explication

- Les variations d'une série peuvent être expliquées par une autre série (exposition météorologique, pollution atmosphérique, etc.) ;
- Il est possible de modéliser une intervention externe grâce à l'analyse de séries temporelles ;
- Ces analyses permettent de réaliser des scénarios pour la période contemporaine : en agissant sur une variable explicative, il est possible d'observer le comportement de la variable expliquée ;

Prévision

La prévision *a priori* permet la planification ;

La prévision *a posteriori* permet d'estimer l'impact d'une perturbation sur la variable expliquée ;

Des scénarios pour le futur, enfin, peuvent être réalisés (Emmanuel, 2006).

Dans notre étude, les données de base sont la pluviométrie, évapotranspiration potentielle, et de la station d'Alger.

IV-2- Présentation des données

Les séries de données ont été gracieusement offertes par l'**Office Nationale de Météorologie** (l'ONM). Il s'agit de la pluie mensuelle et décadaire, les températures journalières maximales, et les paramètres de calcul de l'évapotranspiration potentielle de la formule de Penman, de la station de **Dar El Beida** dont les caractéristiques sont présentés dans le tableau suivant :

Tableau IV-1 **Caractéristiques de la station de Dar El Beida**

Station	Numéro de la station	Code	Latitude (Nord)	Longitude (E/W)	Altitude (Mètre)
D.E.B	6	390	36°43	03°15E	24

Les séries de données qu'on a traitées sont complètes et sont présentées de la manière suivante :

- La pluie en mm (1936-2000) au pas de temps mensuel et décadaire.

L'évapotranspiration potentielle a été calculée au pas de temps mensuel et décadaire, par la formule de Penman et en exploitant les données suivantes :

- La température moyenne (°C) :
(1958 – 1992) au pas de temps mensuel et décadaire.
- L'humidité relative (%) :
(1958 – 1992) au pas de temps mensuel et décadaire.
- L'insolation (h et 1/10h) :
(1958 – 1992) au pas de temps mensuel et décadaire.
- La vitesse du vent (m/s) :
(1958 – 1992) au pas de temps mensuel et décadaire.

IV -2-1 Calcul de l'évapotranspiration potentielle

L'évapotranspiration potentielle (ETP) est théoriquement la quantité d'eau que cède une culture en pleine croissance couvrant totalement un sol largement pourvu d'eau par transpiration des plantes et évaporation du sol. Elle dépend donc du stade végétatif, de la plante, et des conditions climatiques.

L'ETP est aussi une mesure climatique, dont la valeur est établie à partir de relevés météorologiques selon des formules plus ou moins complexes impliquant des variables météorologiques mesurables. Il est à noter que l'ETP correspond à une valeur moyenne tous couverts végétaux confondus (Bermad, 2004).

En effet, l'évapotranspiration en Algérie n'est pas mesurée, elle est calculée à l'aide de la formule de Penman. L'utilisation de cette formule est justifiée du fait qu'elle est la plus représentative de l'ETP, en tenant compte de quatre paramètres essentiels qui l'influencent : la température, l'humidité relative, l'insolation et la vitesse du vent.

IV-2-2 La formule de Penman

La formule développée prend la forme suivante :

$$ETP = \frac{1}{1 + \frac{F_T}{\gamma}} \left[\frac{Iga \cdot F_T}{59 \cdot \gamma} (1 - a) \left(0.18 + 0.62 \frac{h}{H} \right) - \frac{\sigma \cdot F_T}{59 \cdot \gamma} \cdot T^4 \cdot (0.56 - 0.08 \sqrt{e}) \left(0.10 + 0.9 \frac{h}{H} \right) \right]$$

(IV-1)

- ETP : Evapotranspiration potentielle en mm/j ;
 Iga : Radiation directe solaire en l'absence d'atmosphère (cal/cm²/j) ;
 a : Albédo de la surface évaporante (0.2 pour la végétation) ;
 h : Durée réelle d'insolation en heures du jour considéré ;
 H : Durée maximale possible d'insolation pour ce jour en heures ;
 σ : Constante de Boltzman (1.18 10⁷ cal/cm²/j/K) ;
 T : Température moyenne sous abri en degré Kelvin ;
 e : Tension moyenne de vapeur d'eau de l'air en millibars ;
 F_T : Pente de la courbe de tension de vapeur saturante à la température T ;
 γ : Constante psychométrique ;
 e_w : Tension de vapeur saturante à la température T ;
 V : Vitesse moyenne journalière du vent à deux mètres du sol en m/s.

IV-3 Critique des données

Les types de problèmes le plus souvent rencontrés lors du traitement des données en hydrologie sont (Llamas, 1985) :

- Le manque de données : station non existante,
- Non continuité des données : manque d'enregistrements à une station,
- La mauvaise qualité des données, qui peut dépendre de deux causes :

- Données inconsistantes : lorsqu'elles sont entachées d'une erreur systématique due à des instruments de mauvaise qualité ou non calibrés.
- Données non homogènes : lorsqu'il se produit à un moment donné, un changement de l'emplacement de l'appareil qui affecte la mesure (déplacement de la station de mesure ou changement d'appareil).

Afin de garantir la qualité des données, nous devons suivre les étapes suivantes :

- Comblent les lacunes des séries observées si elles existent.
- Détecter et corriger les valeurs douteuses.
- Vérifier l'homogénéité et l'indépendance des données.

IV-3-1 Méthodes de comblement des lacunes

Dans la pratique courante, de nombreuses causes concourent pour entraîner l'absence de certaines informations (absence de mesures, défaut de transmission, erreur de stockage, manipulation erronée, ...), ce qui fait que dans les fichiers disponibles certaines données sont manquantes.

Nous pouvons estimer les valeurs manquantes ou erronées à une station à partir des valeurs provenant des stations voisines soumises aux mêmes conditions climatiques et situées dans la même zone géographique. Plusieurs méthodes sont proposées pour cela, (El jabi et al., 1987) :

- Remplacer la valeur manquante par celle de la station la plus proche.
- Remplacer la valeur manquante par la moyenne des stations voisines. Cette méthode est utilisée lorsque les valeurs moyennes annuelles de la station à compléter ne diffèrent pas plus de 10 % des valeurs moyennes annuelles des stations de référence.
- Remplacer la valeur manquante par une moyenne pondérée par la tendance annuelle des stations pluviométriques par exemple, soit :

$$P_x = \frac{1}{n} \sum_{i=1}^n \left[\frac{\overline{P}_x}{\overline{P}} P_i \right] \quad (\text{IV-2})$$

P_x : Donnée manquante de précipitation (par exemple) à estimer ;

n : Nombre de stations de référence ;

P_i : Précipitation à la station de référence ;

\overline{P}_x : Précipitation moyenne de la station ;

\overline{P} : Précipitation moyenne de la station de référence.

- Remplacer la valeur manquante par la moyenne inter-annuelle de la série considérée :

Une donnée journalière manquante, par exemple, peut être remplacée soit par la moyenne entre la valeur des jours précédent et suivant, soit par la moyenne du mois ou bien par la moyenne journalière inter-annuelle du jour considéré. S'il s'agit d'une valeur mensuelle, nous estimons la valeur manquante soit par la moyenne mensuelle des mois disponibles ou bien par la moyenne mensuelle inter-annuelle du mois considéré.

- Les méthodes basées sur l'analyse de régressions et corrélations :

Le calcul de régression se fait en déterminant les valeurs des paramètres de la relation entre la variable expliquée et la (ou les) variable(s) explicative(s). Cette régression peut être simple ou multiple.

Exemple : Soient Y une variable climatologique dont nous connaissons n valeurs : y_1, y_2, \dots, y_n , et X une autre variable observée de façon concomitante, avec les observations x_1, x_2, \dots, x_n , mais nous possédons par ailleurs sur une période d'observation antérieure par exemple N-n réalisations complémentaires $x_{n+1}, x_{n+2}, \dots, x_N$.

Il s'agit de reconstituer les valeurs manquantes $y_{n+1}, y_{n+2}, \dots, y_N$, concomitantes aux valeurs de la série complémentaire $x_{n+1}, x_{n+2}, \dots, x_N$ comme illustré par la figure ci-dessous :

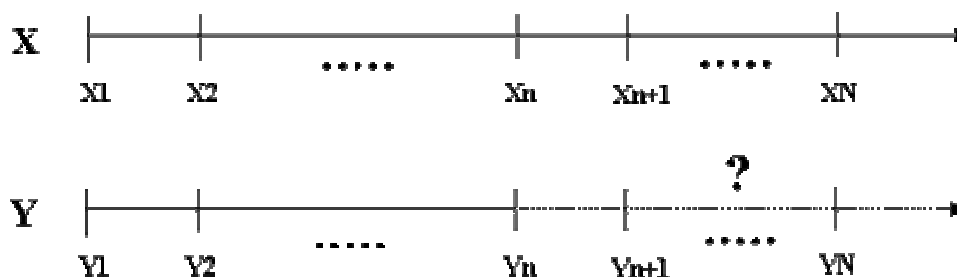


Figure IV-1 reconstitution des séries manquantes.

Une méthode générale consiste à utiliser un modèle linéaire de régression du type :

$$y = a + bx + \varepsilon \quad (\text{IV-3})$$

Nous estimons les paramètres a et b sur la série de couples connus.

Nous pouvons alors estimer les valeurs complémentaires y_i pour chaque x_i connu ($i = n+1, \dots, N$), à partir de la relation estimée :

$$y_i = \hat{a} + \hat{b}x_i \quad (\text{IV-4})$$

- Méthode basée sur la simulation

Pour étudier d'éventuelles anomalies au niveau des séries climatologiques prises en considération, nous avons élaboré un logiciel en fortran 90 (CRIDON) permettant de vérifier les valeurs douteuses et d'étudier l'homogénéité et l'indépendance des données collectées.

L'examen des fichiers fournis par l'O.N.M. pour les différents paramètres ne présente pas de valeurs manquantes. Ceux-ci peuvent être considérées comme étant fiable car ils font l'objet de contrôles réguliers de la part des services gestionnaires de l'O.N.M. vue leur importance (communication de bulletins météorologiques pour le ministère des transports).

IV-3-2 Détection des valeurs singulières

Il arrive fréquemment que des erreurs se glissent dans les données de base et en dénaturent complètement le contenu. Cela peut se produire au cours de la mesure, de la transmission, du codage, du stockage ou du décodage des observations, ou à cause des valeurs intégrées après le comblement des lacunes.

Il est important donc avant tout traitement de détecter les erreurs grossières de ce genre, afin que l'on puisse effectuer une analyse plus fine pour décider si cette donnée douteuse est aberrante et doit donc être éliminée, ou si c'est une donnée exceptionnelle, qui par conséquent doit être au contraire précieusement conservée.

- **Test de Grubbs et Beck :**

Ce test est très utilisé pour sa simplicité. Il peut détecter les points singuliers de la série en calculant les statistiques suivantes :

$$X_h = \exp(\bar{x} + K_n * s_x) \quad (IV-5)$$

$$X_l = \exp(\bar{x} - K_n * s_x) \quad (IV-6)$$

\bar{x} et s_x : Sont respectivement, la moyenne et l'écart type des logarithmes naturels des éléments constituant l'échantillon.

K_n : Valeur statistique du test de Grubbs et Beck, tabulée pour différentes tailles d'échantillons et de niveaux de signification.

Au niveau de risque de 10%, l'approximation polynomiale suivante a été proposée par Bobée (1991) :

$$K_n = 3.62201 + 6.28446 * n^{1/4} - 2.49835 * n^{1/2} + 0.491436 * n^{3/4} - 0.037911 * n \quad (IV-7)$$

n : Taille de l'échantillon.

En appliquant ce test, tous les points de l'échantillon supérieurs à X_h ou inférieurs à X_l seront considérés comme points singuliers.

Le tableau IV-2 récapitule les résultats obtenus après l'application du test de Grubbs et Beck.

Tableau IV-2 Résultats du test de Grubbs et Beck

	n	\bar{x}	s_x	K_n	X_l	X_h
Plj (mm)	21915	0.26	0.90	-235.42	0	1.4E+92
Temp.(°C)	12784	2.80	0.34	-105.85	0	1.2E+17
Hum.(%)	12784	4.28	0.14	-113.09	0	8.9E+08
Inso.(h+0.1h)	12784	1.79	0.90	-113.09	0	1.0E+50
Vv (m/s)	12784	0.98	0.58	-113.09	0	1.7E+29
Pl.men(mm)	792	3.08	1.77	2.75	0.17	2800.8

Pl_j : Pluie journalière à la station d'Alger ;
 Temp : Température journalière à la station d'Alger ;
 Hum : Humidité relative journalière à la station d'Alger ;
 Inso : Insolation journalière à la station d'Alger ;
 Vv : Vitesse du vent journalière à la station d'Alger ;
 Pl.men : Pluie mensuelle à la station d'Alger ;

L'application de ce test au pas du temps journalier et mensuel a détecté les valeurs nulles comme étant singulières, mais nous ne pouvons rejeter ces valeurs car elles correspondent aux jours ou aux mois secs.

IV-3-3 Traitement statistique des données

Après avoir analysée et critiquée les données recueillies, il est indispensable de vérifier leur homogénéité et leur indépendance.

IV-3-4 Test d'homogénéité

Les échantillons sont dits homogènes, s'ils sont caractérisés par une même loi de probabilité et par les mêmes paramètres, les données recueillies sont alors considérées comme extraites d'une même population.

Il existe plusieurs tests pour vérifier l'homogénéité des séries de données, nous avons utilisé dans ce cas celui de Wilcoxon-Mann-Withney.

Il permet de tester si une série statistique est homogène, c'est à dire si les éléments qui la constituent proviennent de la même population. En hydrologie, cela veut dire que les conditions qui ont prévalu lors de la collecte des données, ou de l'avènement du phénomène considéré (pluie, évaporation, etc.) n'ont pas changé pendant toute la durée de la collecte, ou en d'autres termes qu'il n'y a pas eu un phénomène extraordinaire qui aurait pu modifier les données hydrologiques considérées (changement de site de la station de mesure, construction d'un barrage à l'amont, urbanisation etc...).

Pour appliquer ce test nous avons procédé comme suit : l'échantillon est subdivisé en deux sous-ensembles de tailles respectives p et q avec $p \geq q$.

$$\begin{array}{c}
 x_1, x_2, \dots, x_i, \dots, x_q \\
 y_1, y_2, \dots, y_i, \dots, y_p
 \end{array}$$

La taille de l'échantillon original est $n=p+q$.

Ensuite nous classons les valeurs par ordre croissant de 1 à n et nous notons les rangs $R(x_i)$ des éléments du premier sous-ensemble et $R(y_i)$ des éléments du second sous-ensemble dans l'échantillon original.

Nous définissons deux statistiques K et F comme suit :

$$K = L - \frac{q(q+1)}{2} \quad \text{et} \quad F = pq - K \quad (\text{IV-8})$$

$$L = \sum_{i=1}^q R(x_i) \quad (\text{IV-9})$$

L : Somme des rangs des éléments de l'échantillon 1 dans l'échantillon original.

K : Somme des nombres de dépassements de chaque élément du second échantillon par ceux du premier.

F : Somme des nombres de dépassements des éléments du premier échantillon par ceux du second.

Lorsque n, p et q sont suffisamment grands, K et F suivent une distribution normale :

- de moyenne $\bar{K} = \bar{F} = \frac{pq}{2}$,
- et d'écart type $s_K = s_F = \left(\frac{pq}{12}(p+q+1) \right)^{1/2}$

Et nous pouvons alors tester si les deux échantillons proviennent de la même population au niveau de signification α en comparant la grandeur statistique :

$$T = \left| \frac{K - \bar{K}}{s_K} \right| \quad (\text{IV-10})$$

Avec la variable normale centrée réduite U de probabilité de dépassement $\frac{\alpha}{2}$.

Si $T \leq U_{\alpha/2}$ alors les deux échantillons proviennent de la même population, dans le cas contraire l'échantillon étudié est considéré comme non homogène ($T \geq U_{\alpha/2}$).

Le tableau IV-3 résume les résultats de ce test.

Tableau IV-3 Résultats du test de Wilcoxon-Mann-Withney.

	Plj.(mm)	Temp.(°C)	Hum.(%)	Inso.(h+0.1h)	Vv.(m/s)	Pl.men(mm)
T	.3153	.0071	.0943	.0409	.0251	.0003
$U_{\alpha/2}$ ($\alpha=5\%$)	1.96					

Les résultats de ce test, montrent que les observations de tous les paramètres étudiés de la station d'Alger à différents pas de temps sont homogènes.

IV-3-5 Test d'indépendance

L'application des méthodes statistiques repose sur la validité de certaines hypothèses de base relatives à la constitution de l'échantillon observé. En effet le fichier recueilli doit être constitué par une série d'observations indépendantes formant ce que l'on appelle un échantillon extrait au hasard de la population statistique. En d'autres termes, l'échantillon

$X=(x_1, \dots, x_j, \dots, x_n)$ est dit « au hasard », si le test permet de vérifier l'hypothèse H_0 d'indépendance des observations.

Dans cette étude, nous avons appliqué le test de Wald-Wolfowitz à toutes les séries recueillies sur des pas de temps journaliers et mensuels.

A la suite des travaux d'Anderson (Dagnelie, 1998), Wald et Wolfowitz ont développé un test non paramétrique du coefficient d'autocorrélation, celui-ci est basé sur le critère qu'aucune observation appartenant à l'échantillon ne devrait influencer l'observation suivante :

La statistique du test est calculée comme suit :

$$R = \sum_{i=1}^{n-1} x_i x_{i+1} + x_n x_1 \tag{IV-11}$$

Pour n suffisamment grand, cette statistique suit une distribution normale de moyenne et de variance :

$$E[R] = \frac{(S_1^2 - S_2)}{n-1} \tag{IV-12}$$

$$Var[R] = \frac{(S_2^2 - S_4)}{n-1} + \frac{(S_1^4 - 4S_1^2 S_2 + 4S_1 S_3 + S_2^2 - 2S_4)}{(n-1)(n-2)} - \{E[R]\}^2 \tag{IV-13}$$

$$S_k = \sum_{i=1}^n x_i^k \tag{IV-14}$$

S_k : Moment d'ordre k de l'échantillon.

La valeur statistique $T = \frac{R - \bar{R}}{\sqrt{Var(R)}}$ suit une distribution normale centrée réduite $N(0,1)$, et est utilisée pour tester à un niveau de signification, l'hypothèse d'indépendance en comparant la valeur absolue de T avec la valeur normale centrée réduite $U_{\alpha/2}$ correspondant à la probabilité de dépassement de $\alpha/2$.

Si $|T|$ est inférieure à $U_{\alpha/2}$ l'hypothèse d'indépendance est vérifiée, celle-ci est rejetée dans le cas contraire.

Le tableau IV-4 présente les résultats de ce test.

Tableau IV-4 Résultats du test de Wald-Wolfowitz

	Plj.(mm)	Temp.(°C)	Hum.(%)	Inso.(h+0.1h)	Vv.(m/s)	Pl.men(mm)
T	49.05	156.45	155.75	133.72	125.48	97.40
$U_{\alpha/2}$ ($\alpha=5\%$)	1.96					

Les résultats illustrés dans le tableau IV-4 montrent que les observations de tous les paramètres de la station d'Alger à différents pas de temps ne sont pas indépendantes (Bermad, 2004).

Chapitre V : Elaboration des modèles de prévision

L'objectif de cette étude est de tester trois modèles différents, pour la prévision de deux variables hydrologiques distinctes et qui sont la pluie et l'ETP. Pour pouvoir décider quel est le meilleur parmi ces modèles, une comparaison entre les résultats fournis par chacun de ces derniers est nécessaire. En effet, on a opté pour les réseaux de neurones, et l'Analyse en Composantes Principales comme modèles de prévision. Ainsi, une combinaison entre ces deux derniers constitue un troisième modèle testé dans ce travail.

Le choix de ces deux variables se justifie, par l'intérêt apporté par leur prévision dans le domaine de gestion des ressources en eau. Par ailleurs, la distinction de ces deux variables par leur type, nous ramène à confronter deux problèmes de prévision différents. Etant cyclique, l'ETP est facilement prévue. A l'opposé, la pluie présente un phénomène très complexe, difficile à prévoir à causes de son caractère aléatoire.

Il est prouvé que le problème de prévision devient plus délicat au fur et à mesure que le pas de temps pris en considération est grand. Pour cette raison, on a choisi dans cette étude les pas de temps mensuel et décadaire.

V-1-Le perceptron multicouches

Dans le cadre de cette étude, un perceptron multicouches ayant 10 neurones dans sa couche cachée est adopté comme modèle de prévision de la pluie et de l'ETP. Ce modèle est détaillé par la suite.

V-1-1 L'architecture

L'unité de base de chaque réseau de neurone est le neurone formel. Il est constitué d'un intégrateur qui effectue une somme pondérée de ses entrées.

Le perceptron multicouches est un réseau constitué de plusieurs couches disposées l'une après l'autre. Chaque couche contient plusieurs neurones.

Dans ce cas là, le fonctionnement de chaque neurone est le même. Une somme pondérée doit être effectuée, avant de passer au calcul de la sortie a_j du neurone j par la fonction de transfert f . On peut résumer le calcul de la sortie du a_j par l'expression ci dessous :

$$a_j = f(\sum w_{ij} \times p_i + b_j) \quad (V-1)$$

Où :

a_j est la sortie du neurone j ;

f est sa fonction d'activation ;

p_i est un élément du vecteur d'entrée ;

w_{ij} est la connexion (poids) reliant le neurone j à l'entrée p_i ;

b_j est le biais.

Ainsi, pour chaque couche k constituée de s neurones, les vecteurs de sortie (a^k) et de biais (b^k) du même taille s . Les poids w_{ij} forment à leur tour une matrice de dimension $s.r$.

Le vecteur d'entrée peut être le stimuli si c'est le cas de la première couche, comme il peut être le vecteur de sortie de la couche précédente s'il s'agit d'une couche intermédiaire.

La fonction d'activation adoptée dans ce type de réseaux de neurones, est souvent la fonction sigmoïde.

V-1-2 L'apprentissage [Lehsen 2005]

L'apprentissage du perceptron multicouches est un apprentissage supervisé par l'algorithme de rétro propagation de l'erreur (backpropagation).

Il consiste en la modification des poids des connexions entre les différentes couches, par rapport à une erreur calculée en sortie de réseau: on dit alors que l'erreur est rétro propagée (Davallo, 1991).

Cette erreur notée $e(t)$ au temps t est déterminée par la présentation d'une série $\{(p_1, d_1), (p_2, d_2), \dots, (p_Q, d_Q)\}$ de couples d'entrée p et de sortie désirée d .

On confronte alors à chaque fois la sorti effective $a(t)$ à $d(t)$, la sortie que l'on désire obtenir, l'erreur est alors calculée par l'expression qui suit (Parizeau, 2004) :

$$e(t) = d(t) - a(t) \quad (V-2)$$

Où :

$a(t)$: désigne la sortie calculée par le réseau au temps t ;

Un indice de performance est alors calculé : il s'agit de l'espérance de l'erreur instantanée :

$$F(x) = E[e^T(t) \cdot e(t)] \quad (V-3)$$

Où :

$[e^T(t) \cdot e(t)]$: désigne le produit de $e(t)$ et de $e^T(t)$, deux erreurs calculées à un temps T d'intervalle.

L'indice de performance précédemment défini peut être approximé par l'erreur instantanée $F1(x)$:

$$F1(x) = e^T(x) \cdot e(t) \quad (V-4)$$

C'est finalement cette erreur instantanée qui devra être optimisée, et qui permettra de modifier les poids des différentes connexions ainsi que les biais des différents neurones, par les expressions suivantes :

$$\Delta W^k = -\eta \cdot S^k(t) \cdot (a^{k-1})^T(t) \quad (V-5)$$

$$\Delta b^k(t) = -\eta \cdot S^k(t) \quad (V-6)$$

Où :

η est le taux d'apprentissage (un constant aléatoire compris entre 0 et 1, et qui permet, selon les cas, d'accélérer la convergence de l'algorithme d'apprentissage).

S^k définit comme la *sensibilité* aux changements dans le niveau d'activation des neurones, elle s'exprime par :

$$S^k = \frac{\partial F1}{\partial n^k} \quad (\text{V-7})$$

Où :

$n^k(t)$: désigne l'activation de la couche k .

V-1-3 Programmation du modèle

Le modèle des réseaux de neurones opté dans ce travail est le perceptron multicouches.

Après plusieurs essais et tests, il s'est avéré que le modèle à une seule couche cachée, constituée de *dix neurones*, est celui qui a donné les meilleurs résultats et a nécessité un temps de calcul moyen. Nos essais et tests consistaient à élaborer plusieurs modèles de réseaux de neurones, ceci en modifiant le nombre de couches, le nombre de neurones dans chaque couche et la fonction de transfert. Notre guide qui nous a aidé à trouver le meilleur modèle était *les critères de validation*. Cette démarche a été adoptée pour la pluie comme pour l'ETP.

Ainsi, notre choix s'est limité pour le modèle décrit précédemment ayant une fonction transfert de type sigmoïde dans la couche cachée, et de type linéaire dans la couche de sortie qui ne contient qu'un seul neurone.

La programmation de ce modèle a été faite à l'aide du logiciel MATLAB version 7.0.

L'étude a nécessité d'exécuter chaque programme plusieurs fois (20 à 30 fois).

V-2 L'Analyse en Composantes Principales

Le modèle de prévision adopté est le modèle normé dont l'expression est la suivante :

$$\hat{x}_j = B_{j0} + \sum_{i=1}^Q B_{ji} * C_i + E_j \quad (\text{V-8})$$

Avec :

$B_{j0} = \overline{X_j}$ (est la moyenne de la variable d'ordre j)

$$E_j = \sigma_{x_j} \varepsilon_j$$

En se basant sur la structure de ce modèle de prévision, on détermine

- Les éléments de la matrice $[\beta]$: les vecteurs de la matrice ont pour composantes les coefficients de régression entre les variables et les CP.
- Les éléments de la matrice des CP : $[C]$.
- Les éléments de la matrice des résidus $[E]$.

Le vecteur β_0 , est constitué des moyennes mensuelles inter-annuelles de la série historique. Ce modèle a été élaboré à l'équipe de recherche **GESTOP** de l'ENP par **Bermad** en 2004.

V-3- Modèle combiné (Réseaux de neurones +ACP)

Dans cette étape de ce travail, la variable à prévoir (pluie ou ETP) au temps $t+\Delta t$ est écrite selon l'expression (V-8) :

$$\hat{x}_j = B_{j0} + \sum_{j=1}^Q B_{j1} * C_i + E_j$$

On sait que B_0 est le vecteur des moyennes mensuelles inter annuelles, et B_{ij} est constant après un certain moment. Il nous reste à prévoir la composante principale de l'année de prévision est les 12 résidus correspondant à chaque mois de cette année.

En effet, on a exploité le modèle du perceptron multicouches cité au par avant pour prévoir la composante principale et les résidus de l'année en question, en se basant sur les composantes principales et les résidus des années précédentes.

V-4 -Les critères de validation

Pour juger de la qualité d'une prévision, nous faisons appel à des fonctions critères qui permettent d'apprécier quantitativement et qualitativement les résultats mensuels, estimant globalement l'écart entre les sorties calculées (pluies ou ETP mensuelles prévues), et les variables historiques, il existe deux types de critères :

- Les critères graphiques : l'analyse graphique est indispensable et primordial, cela est obtenu en portant sur un graphique les valeurs observées et les valeurs calculées sur une période test.

- Les critères numériques (Rognon, 1996) :

1- le critère de Nash (NS) est défini par :

$$NS = \left[1 - \frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2} \right] * 100 \quad (V-9)$$

Avec :

Q_i : Valeurs des données observées ;

\hat{Q}_i : Valeurs prévues par le modèle ;

\bar{Q} : Moyenne des données observées.

Ce critère a beaucoup séduit les hydrologues, il a été utilisé au début pour juger un modèle fonctionnant par épisodes (Nash et Sutcliffe, 1970 ; O'Connell et all, 1970). Il exprime le pourcentage de la variance de la série mesurée expliqué par le modèle, ou en d'autres termes, il

explique ce que l'on gagne par rapport à un modèle médiocre qui donnerait comme valeurs calculées la moyenne des valeurs observées (Nascimento, 1995).

Le critère de Nash peut prendre des valeurs de $-\infty$ à 100 % (Bermad, 2004).

2-Coefficient de détermination donné par la formule suivante :

$$R^2 = \frac{\left[\sum_{i=1}^n \left((Q_i - \bar{Q}) \times (\hat{Q}_i - \bar{\hat{Q}}) \right) \right]^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2 \times \sum_{i=1}^n (\hat{Q}_i - \bar{\hat{Q}})^2} \quad (\text{V-10})$$

Ce critère exprime la corrélation entre les valeurs observées et celles prévues.

Avec :

Q_i : Valeurs des données observées ;

\hat{Q}_i : Valeurs prévues par le modèle ;

\bar{Q} : Moyenne des données observées ;

$\bar{\hat{Q}}$: Moyenne des valeurs prévues par le modèle.

En plus des critères numériques et graphique cités ci-dessus, nous avons validé nos modèles à l'aide du critère graphique $x=y$ qui exprime le degré de corrélation entre les valeurs prévues et observées ainsi que le critère de l'ONM.

3- Critère de l'ONM :

Ce critère considère qu'on a une bonne prévision si l'écart entre les valeurs prévues et observées ne dépasse pas les 30mm au pas de temps mensuel et 10mm au pas de temps décadaire.

Chapitre VI : Résultats et interprétations

La prévision est l'estimation préalable de l'état futur d'une grandeur bien déterminée à différentes échéances.

Ce chapitre est consacré à la présentation des résultats et leurs interprétations. En fait, plusieurs modèles ont été testés.

VI-1- Modèle des réseaux de neurones

Nous avons exploité le modèle du perceptron multicouches décrit dans le chapitre précédent. Nous avons adopté une démarche qui consiste à prévoir une variable x donnée (pluie ou ETP) au temps $t+1$, en se basant sur l'historique de cette même variable aux temps t , $t-1$, $t-2$, etc. Donc, l'entrée du modèle est une matrice à n entrées qui représentent la variable x aux temps t , $t-1$, ..., $t-n$, la sortie est le vecteur ayant comme éléments x au temps $t+1$. Dans une première étape, nous avons testé plusieurs modèles à différent nombre d'entrées, afin de trouver le meilleur. On a préféré de choisir l'ETP comme variable à prévoir comme une première étape. La validation est effectuée pour des années 1980 et 1992. Ensuite la même démarche est suivie pour la prévision de la pluie, néanmoins la validation est effectuée sur l'année 2000. Notons que ce choix des années est arbitraire.

➤ Modèle de l'ETP

- En décadaire

Les résultats de la prévision des années 1980 et 1992 sont donnés dans le tableau suivant :

Tableau VI-1 : Variation de nombre d'entrées pour la prévision d'ETP décadaire des années 1980 et 1992

Nombre d'entrée	1980		1992	
	Nash (%)	R ²	Nash (%)	R ²
ETP (t)	84,65	0,85	86,67	0,87
ETP (t) +ETP (t-1)	85,20	0,85	87,52	0,88
ETP (t) +ETP (t-1) +ETP (t-2)	85,77	0,86	87,83	0,88
ETP (t) +ETP (t-1) +ETP (t-2) +ETP (t-3)	89,71	0,89	89,71	0,90
ETP (t) +ETP (t-1) +ETP (t-2) +ETP (t-3) +ETP (t-4)	88,28	0,88	89,14	0,90

Le tableau VI-1 montre que le meilleur modèle de prévision des ETP décadaires pour les années 1980 et 1992, est celui à quatre vecteurs dans la matrice d'entrées.

- En mensuel

Les résultats de la prévision des années 1980 et 1992 sont donnés dans le tableau suivant :

Tableau VI-2 : Variation de nombre d'entrées pour la prévision d'ETP mensuelle des années 1980 et 1992

Les entrées	1980		1992	
	Nash (%)	R ²	Nash (%)	R ²
ETP (t)	75,46	0,76	74,12	0,75
ETP (t) +ETP (t-1)	95,57	0,96	91,49	0,92
ETP (t) +ETP (t-1) +ETP (t-2)	96,23	0,96	92,95	0,95
ETP (t) +ETP (t-1) +ETP (t-2) +ETP (t-3)	97,40	0,98	94,65	0,98
ETP (t) +ETP (t-1) +ETP (t-2) +ETP (t-3) +ETP (t-4)	93,90	0,94	93,51	0,96

Le tableau VI-2 montre très bien que le modèle à quatre entrées est le plus adéquat à la prévision des ETP mensuelles.

➤ **Modèle de la pluie**

• **En décadaire**

Les résultats obtenus sont montrés dans le tableau VI-3.

TableauVI-3 : Résultats de la prévision des pluies au pas de temps décadaire

Les entrées	Nash (%)	R ²
P (t)	-50,98	0,0074
P(t)+P(t-1)	-40,39	0,0232
P(t)+P(t-1)+P(t-2)	-40,4725	0,0328
P(t)+P(t-1)+P(t-2)+P(t-3)	12,34	0,04
P(t)+P(t-1)+P(t-2)+P(t-3)+P(t-4)	7,6782	0,0275

On remarque très bien que le modèle à quatre entrées est celui qui a donné les meilleurs résultats.

• **En mensuel**

Les résultats sont récapitulés aux niveaux du tableau VI-4.

Tableau VI- 4 : Résultats de la prévision des pluies mensuelles

Les entrées	Nash (%)	R ²
P (t)	-102,6690	0,1295
P(t)+P(t-1)	-103,3637	0,0877
P(t)+P(t-1)+P(t-2)	-80,7748	0,1818
P(t)+P(t-1)+P(t-2)+P(t-3)	10,06	0,33
P(t)+P(t-1)+P(t-2)+P(t-3)+P(t-4)	4,9820	0,1732

Le modèle à quatre entrées pour la prévision de la pluie mensuelle par les réseaux de neurones est celui qui a donné les meilleurs résultats.

Conclusion : On opte pour le modèle à quatre entrées pour la prévision des ETP et des pluies en utilisant les réseaux de neurones.

VI-2-Prévision de la pluie et de l'ETP

VI-2-1-Prévision de l'ETP

L'historique de l'évapotranspiration étudié dans ce travail, s'étale de 1958 à 1992 au pas de temps décadaire et mensuel. Ce qui fait 35 ans d'observation. Dans ce qui suit, les modèles utilisés sont validés sur les années 1980, 1985 et 1992.

VI-2-1-1- Prévision de l'ETP au pas de temps mensuel

➤ Prévision par les réseaux de neurones

La démarche poursuivie consiste à exploiter le modèle du perceptron multicouches à quatre entrées, pour la prévision des ETP mensuelles des années 1980, 1985, et 1992. Ce choix des années pour la prévision est arbitraire, seulement là on veut connaître l'apport de la variation des années d'apprentissage. Les résultats obtenus sont donnés dans le tableau VI-5.

Tableau VI-5 : Résultats de la prévision de l'ETP mensuelle par les réseaux de neurones

Année	Nash (%)	R ²
1980	97,40	0,98
1985	95,66	0,96
1992	93,97	0,96

Les valeurs des critères de Nash et de R² donnés par le tableau précédent, prouvent que le modèle des réseaux de neurones utilisé permet de prévoir avec satisfaction l'ETP mensuelle.

Les critères de validation obtenus pour la prévision de l'année 1980 sont les meilleurs par rapport aux autres.

La figure VI-1 représente la prévision de l'ETP mensuelle de l'année 1980, et la figure VI-2 montre la corrélation entre les valeurs prévues et celles observées.

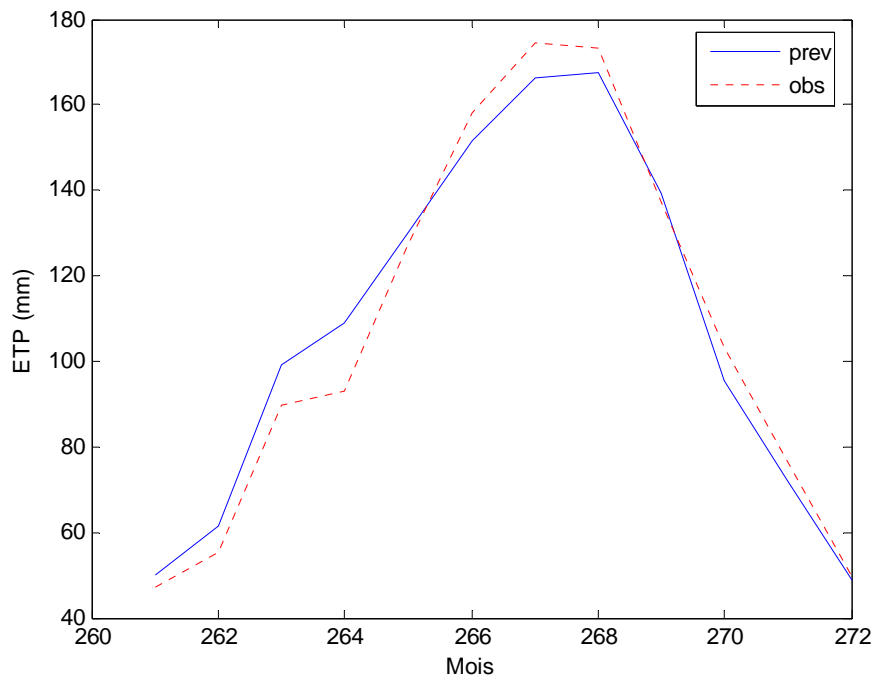


Figure VI-1 : Prédiction de l'ETP de l'année 1980

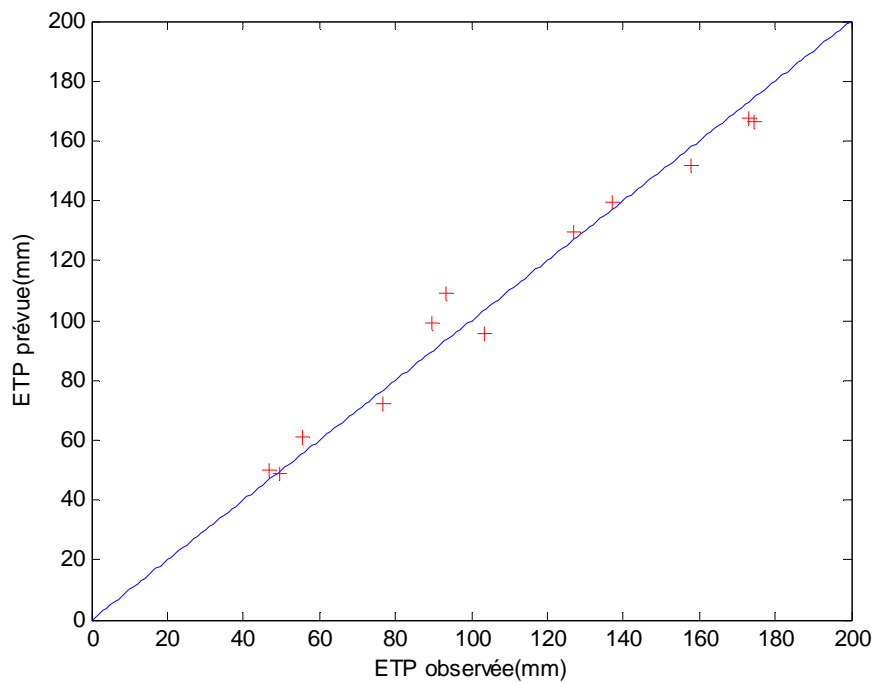


Figure VI-2 : Corrélation entre les ETP prévues et observées de l'année 1980

La figure VI-1 montre que les écarts entre les valeurs prévues et observées ne dépassent pas les 15 mm. Ce qui est négligeable pour une prévision mensuelle.

Ainsi, la figure VI-2 montre une très bonne corrélation entre les valeurs prévues et observées, du fait que le nuage des points est aligné sur la première bissectrice.

Les courbes de tendance des ETP prévues et observées, des deux autres années, ainsi que la corrélation ces deux séries, sont représentées dans l'annexe VI-1. Les courbes de tendance et les corrélations montrent que le modèle du perceptron multicouches permet une bonne prévision des ETP mensuelles.

➤ Prévision de l'ETP par l'Analyse en Composantes Principales

Les critères de validation pour la prévision des trois années sont regroupés dans le tableau VI-6

Tableau VI-6 : Résultats de la prévision de l'ETP mensuelle par l'ACP

Année	Nash (%)	R ²
1980	97,40	0,98
1985	98,35	0,98
1992	93,36	0,97

On remarque que les critères de validation obtenus par ce modèle sont très bons. Ils sont proches de ceux obtenus par les réseaux de neurones, néanmoins on remarque une amélioration considérable des critères de validation de l'année 1985.

Les variations des valeurs observées et prévues de l'année 1985 sont représentées par la figure VI-3. Les évolutions des ETP prévues et observées sont données dans l'annexe VI-5.

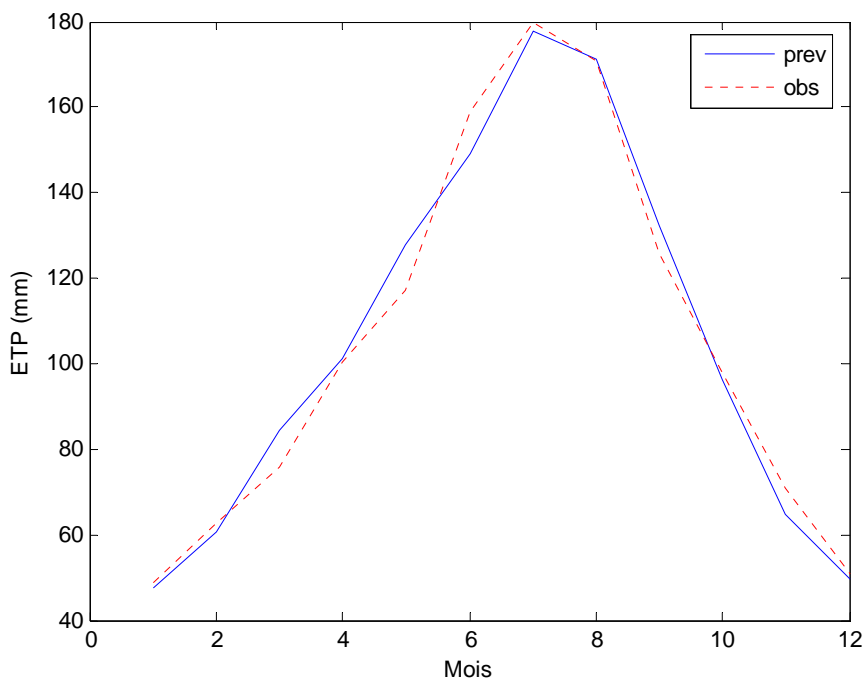


Figure VI-3 : Prévision de l'ETP de l'année 1985

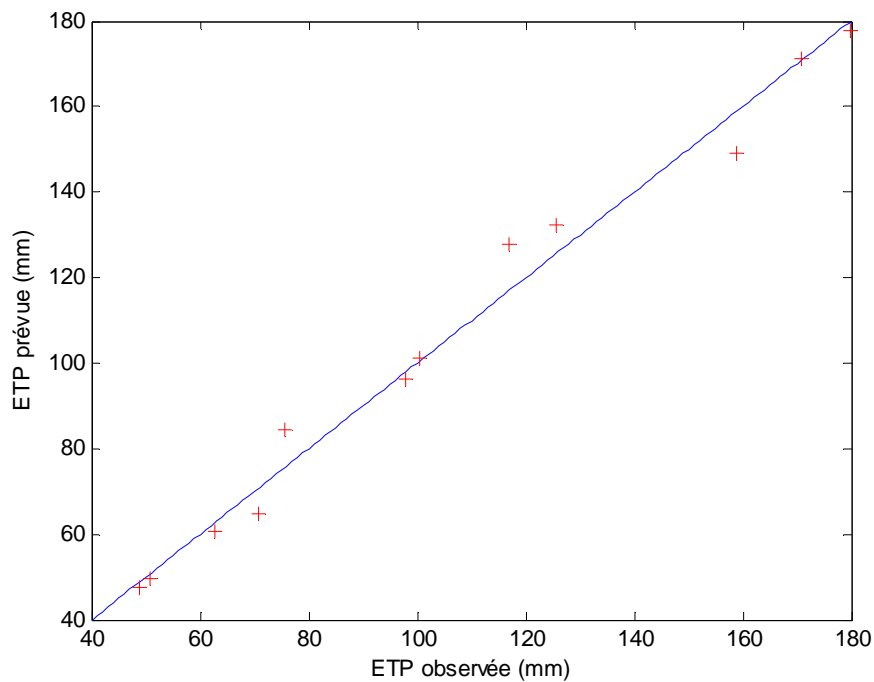


Figure VI-4 : Corrélation entre les ETP prévues et observées de l'année 1985

Il apparaît nettement dans les figures des évolutions de l'ETP, que les écarts entre les valeurs observées et prévues sont très faibles. Aussi, la corrélation entre l'observé et le prévu est très grande, d'après les courbes $x=y$ les nuages des points sont presque alignés sur la première bissectrice.

On constate alors, que l'Analyse en Composantes Principales est un très bon modèle pour la prévision des ETP mensuelles, qui a apporté un plus par rapport aux réseaux de neurones.

➤ Prévision de l'ETP par le modèle combiné (ACP+RN)

Les résultats obtenus sont collectés dans le tableau VI-7.

Tableau VI-7 : Résultats de la prévision de l'ETP mensuelle par le modèle combiné

Année	Nash (%)	R ²
1980	95,90	0,97
1985	96,40	0,97
1992	93,41	0,96

On remarque du tableau VI-7 que les critères numériques de validation obtenus par la prévision avec le modèle combiné, sont proches de ceux obtenus par les modèles précédents.

La courbe de tendance ainsi que la corrélation entre les ETP prévues et observées de l'année 1985 sont données par la figure VI-5 et VI-6 successivement. Pour les autres années, l'annexe VI-9 donne les différentes figures concernant les deux autres années.

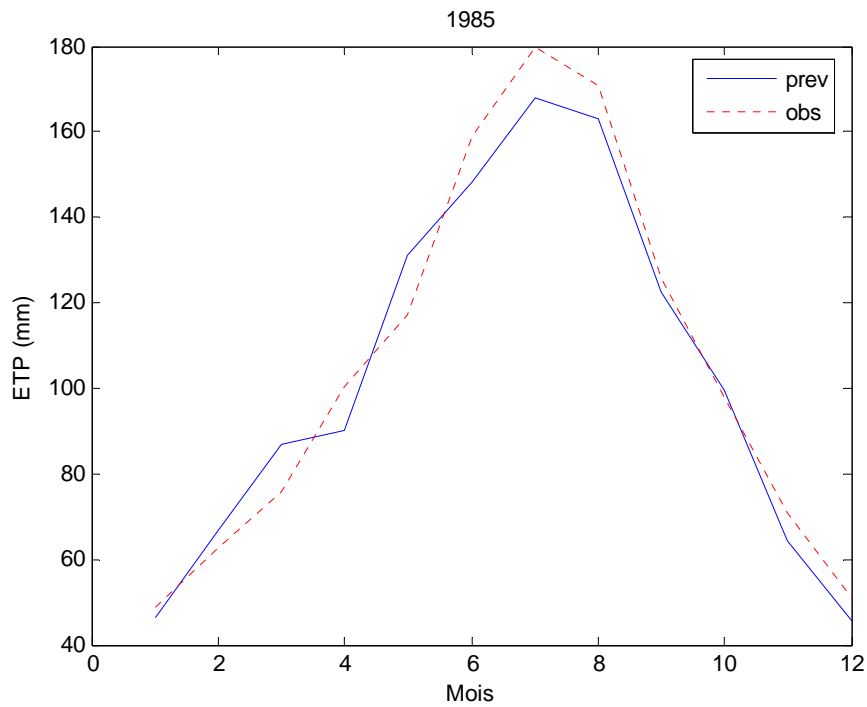


Figure VI-5 : Evolution de l'ETP prévue et observée DE l'année 1985

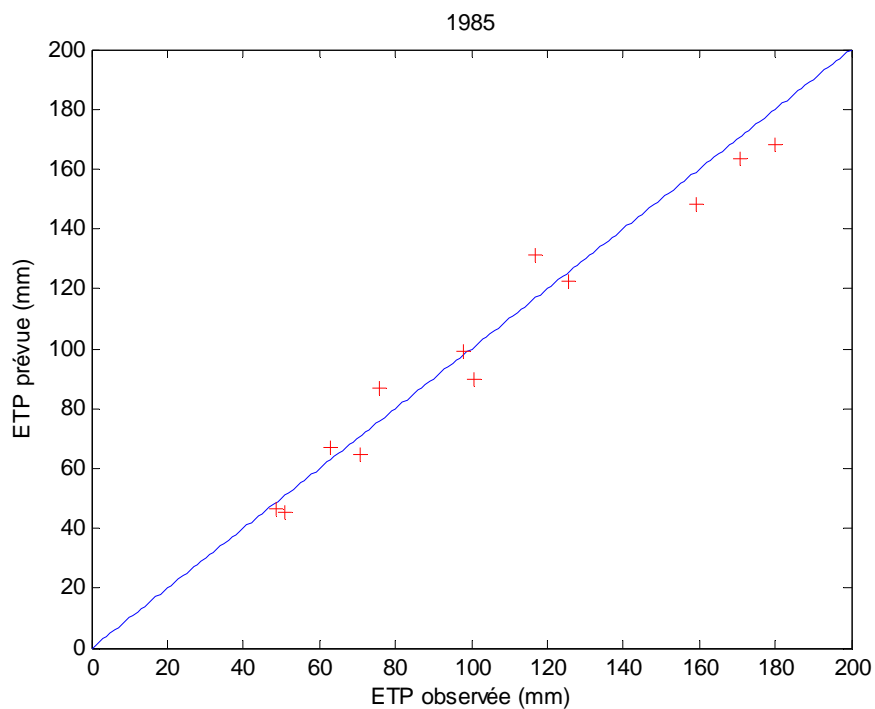


Figure VI-6 : Corrélation entre l'ETP prévue et observée de l'année 1985

D'après les figures VI-5 et VI-6, on remarque que les écarts de prévision sont très faibles et la corrélation est très bonne entre le prévu et l'observé. DE là, on constate que le modèle combiné

permet une bonne prévision des ETP mensuelle et que les résultats qu'il donne sont semblables à ceux des deux autres modèles.

VI- 2-1-2 Prévision de l'ETP décadaire

Dans ce cas là, on ne teste que deux modèles : les réseaux de neurones et l'Analyse en Composantes Principales.

➤ Prévision de l'ETP décadaire par les réseaux de neurones

Le même modèle utilisé pour la prévision des ETP mensuelles est testé pour la prévision des ETP décadaires. Les critères de validations statistiques obtenus sont donnés par le tableau VI-8 :

Tableau VI-8 : Résultats de prévision de l'ETP décadaires par les réseaux de neurones.

Année	Nash (%)	R ²
1980	89,18	0,89
1985	85,17	0,85
1992	89,71	0,90

L'observation du tableau VI-8 nous permet de remarquer des bons critères de validation. Néanmoins ces critères ont des valeurs inférieures à ceux obtenus en mensuel.

La figure VI-7 représente l'évolution des ETP décadaires, et la figure VI-8 montre la corrélation entre les prévues et les observées, pour l'année 1992. Les autres graphes de tendance et de corrélation sont donnés dans l'annexe VI-2.

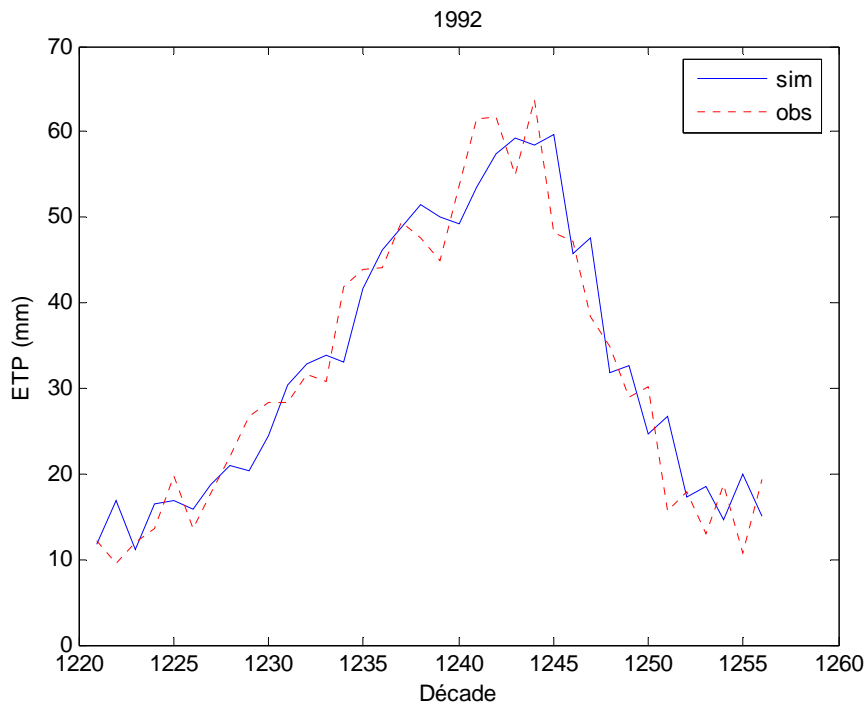


Figure VI-7 : Evolution des ETP décadaires prévues et observées de l'année 1992

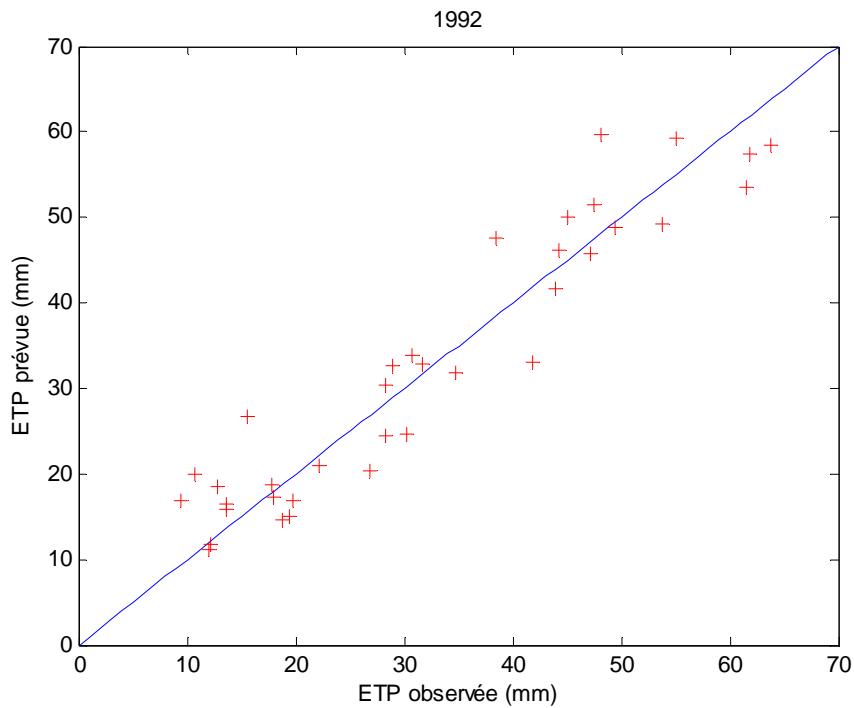


Figure VI-8 : Corrélation entre les ETP décadaires prévues et observées de l'année 1992

De la figure VI-7, on remarque que les écarts entre les valeurs prévues et observées sont très faibles et ne dépassent pas les 10mm. De même, la figure VI-8 fait apparaître un nuage de points

regroupé autour de la première bissectrice. Ceci est généralisé pour les trois années. De là, on constate que le modèle des réseaux de neurones permet une bonne prévision des ETP décadaires.

➤ Prévion des ETP décadaires par l'ACP

Les critères de validation sont donnés par le tableau VI-9

Tableau VI-9 : Résultats de prévision de l'ETP décadaire par l'ACP

Année	Nash (%)	R ²
1980	94,75	0,95
1985	93,84	0,94
1992	91,27	0,95

Les critères de validation donnés par ce dernier tableau sont des valeurs très bonnes, et qui dépassent celles trouvées par les réseaux de neurones. Ce qui explique la puissance de l'ACP à prévoir l'ETP décadaire. Ceci va être confirmé après l'observation des tendances et les corrélations. Ces dernières sont représentées dans l'Annexe VI-6 pour les années 1985 et 1992. Les critères de validation de l'année 1980 sont meilleurs par rapport aux autres, donc on va représenter les courbes de tendance et de corrélation par les figures suivantes :

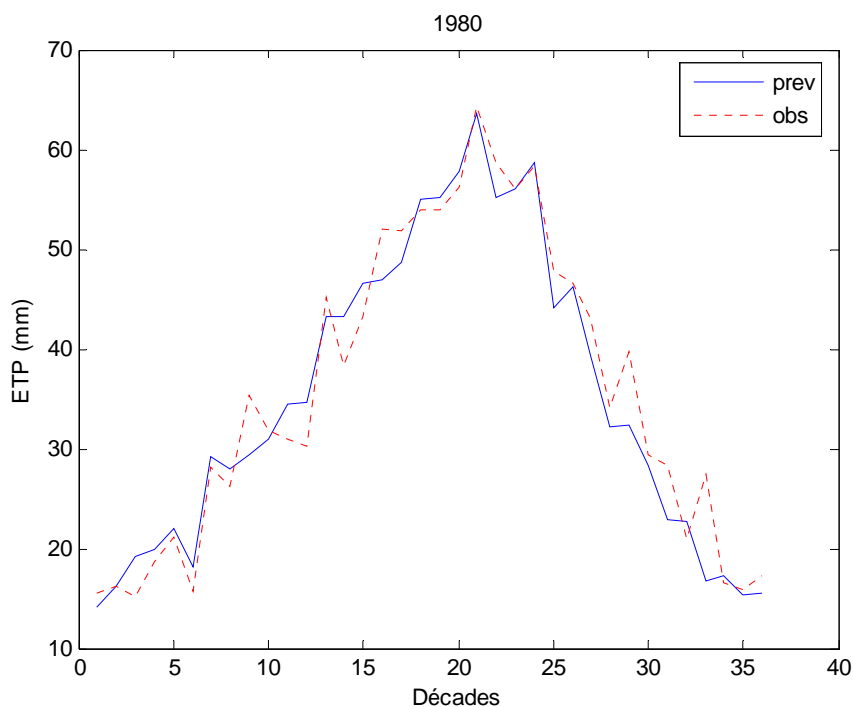


Figure VI-9 : Evolution des ETP décadaires prévues et observées de l'année 1980

De cette figure, on constate que le modèle d'ACP a pu reproduire les valeurs de toutes les décades avec des écarts négligeable qui ne dépassent pas les 10 mm.

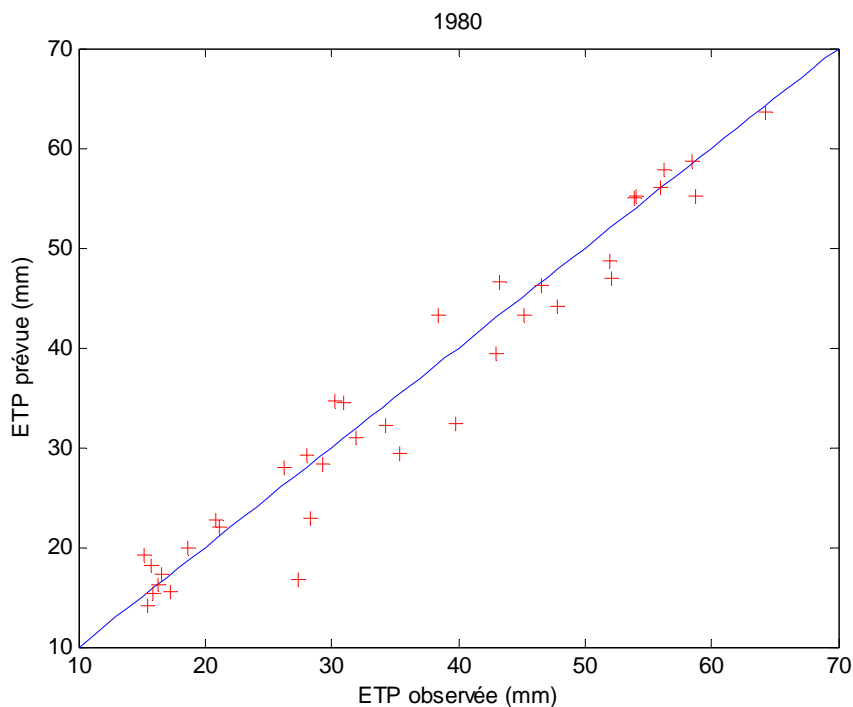


Figure VI-10 : Corrélation entre les ETP prévues et observées de l'année 1980

La figure VI-10 montre une forte corrélation entre les ETP prévues et observées car les points sont très proches de la première bissectrice.

Après observation de l'annexe VI-6 ces constats peuvent être généralisés. De là, on peut dire que l'ACP est un très bon modèle de prévision des ETP décennaires.

VI-2-1-3-Conclusion

L'ETP est un phénomène cyclique, très facile à prévoir. Des remarques précédentes, on peut conclure que tous les modèles utilisés pour la prévision de cette variable en mensuel ou en décennaire, ont donné des excellents résultats. En comparant les modèles, on trouve que le modèle de l'Analyse en Composantes Principales est celui qui a donné les meilleurs résultats, aussi bien en mensuel qu'en décennaire.

Appliqué pour la prévision des ETP mensuelles seulement, le modèle combiné s'est montré comme un bon modèle de prévision de celles-ci, en ramenant quelque amélioration par rapport aux deux autres.

La variation du nombre d'année de calage n'a pas d'influence sur les résultats.

VI-2-2 Prévision de la pluie

La pluie est un phénomène très aléatoire, complexe et très difficile à prévoir. Nous avons essayé d'appliquer les mêmes modèles utilisés pour la prévision des ETP. Notre série de pluie est de 65 ans d'observation, elle s'étale de 1936 à 2000.

VI-2-2-1 Prédiction de la pluie mensuelle

Nous avons testé trois modèles : les réseaux de neurones, l'Analyse en Composantes Principales et le modèle combiné entre ces deux précédents.

➤ Prédiction de la pluie mensuelle par les réseaux de neurones

Le même modèle que précédemment est utilisé : un perceptron multicouches à quatre entrées. Les résultats se résument par le tableau suivant :

Tableau VI-10 : Résultats de prédiction des pluies mensuelles par les réseaux de neurones

Année	Nash (%)	R ²
1980	33,09	0,40
1985	27,60	0,32
1990	38,57	0,36
1995	31,00	0,30
2000	10,06	0,33

Les critères de validation donné par le tableau ci-dessus, apparaissent *insuffisants*. Le critère de Nash tourne autour de 30% alors que le R² est de l'ordre de 0,3. Les graphes représentant la tendance et la corrélation entre les pluies observées et prévues sont donnés par l'annexe VI-3. On montre comme exemple les figures VI-11 et VI-12, qui représentent successivement les courbes de tendance et de corrélation de la pluie mensuelle pour l'année 1990.

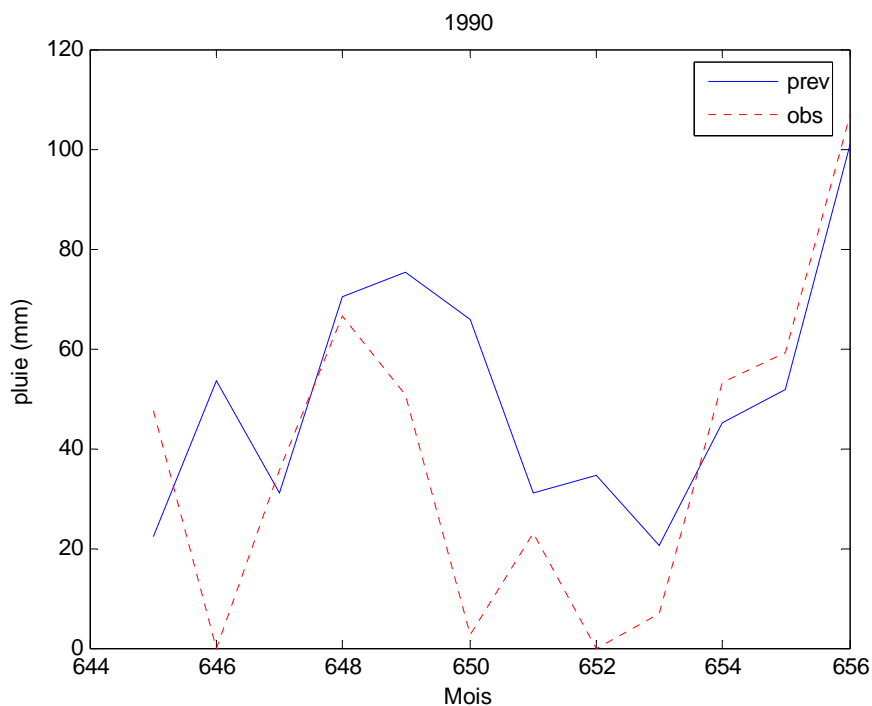
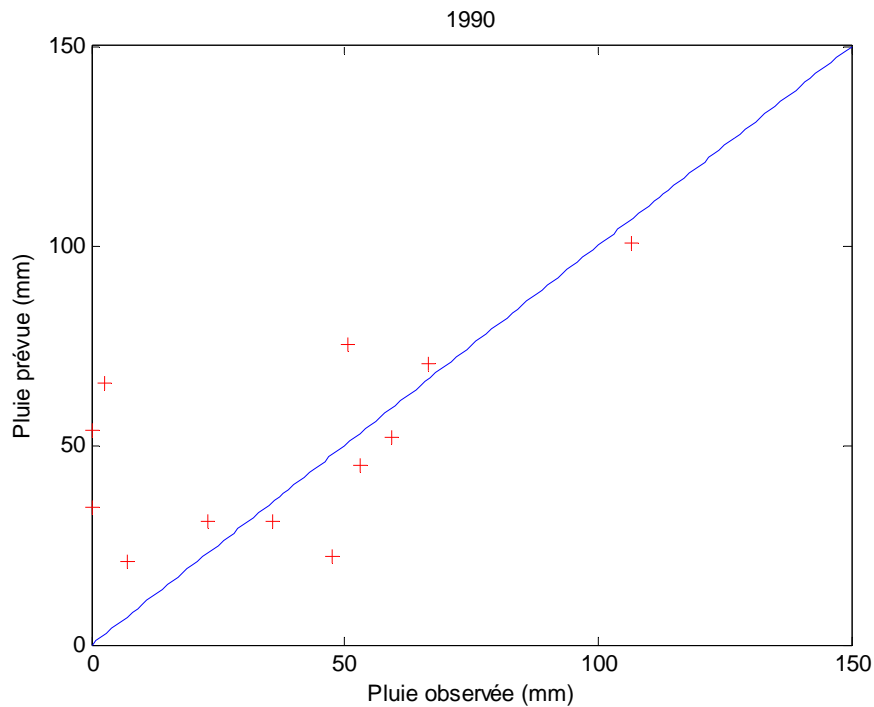


Figure VI-11 : Evolution des pluies mensuelles prévues et observées de l'année 1990**Figure VI-12** : Corrélation entre les pluies mensuelles prévues et observées de l'années 1990

La figure VI-11 montre que le modèle a pu prévoir 9 mois parmi les 12 avec des écarts inférieurs à 30 mm. Il a prévu les mois secs avec des écarts faibles, à l'exception du moi de juin, où il existe un écart de plus de 60mm. Pour les mois humides, on remarque qu'il a donné des valeurs maximales excellentes. Notons que pour le moi de Février qui a une valeur observée presque nulle, le modèle a prévu une valeur de plus de 53mm. On peut expliquer ceci du fait que le modèle, d'après l'apprentissage a considéré le moi de Février comme étant un moi humide, ayant des valeurs fortes de précipitations.

De la figure VI-12, on voit six points proche de la première bissectrice. On constate alors qu'il existe six mois ayant des valeurs prévues et observées bien corrélées, ce qui est satisfaisant pour la prévision de la pluie.

Pour les années 1980, 1985 et 1995, l'observation de l'annexe indique que le modèle a donné des valeurs qui tournent autour de la moyenne pour les mois humides. Les valeurs maximales donc sont mal prévues. Mais, on remarque que pour chaque année 6 à 9 mois sont prévus avec un écart de moins de 30 mm.

Pour l'année 2000, les pluies des mois secs sont prévus avec satisfaction sauf le moi de Juin qui a une valeur prévue dépassant celle observée d'un écart de plus de 80 mm. Pour les mois humides, aussi, les écarts entre le prévu et l'observé ne dépasse pas les 30mm, à l'exception du moi de Janvier dont l'écart entre l'observé et le prévu est de 60 mm. Cela s'explique du fait que le moi de Janvier représente un moi humide, alors qu'il a une valeur de pluie faible dans cette année.

On conclue que le modèle des réseaux de neurones permet une prévision excellente des mois secs, sauf pour quelques exceptions. Pour les mois humides, les écarts de prévision sont de moins de 30 mm, sauf pour les mois à des valeurs observés qui ne reflètent pas la saison. Le modèle arrive à prévoir chaque année de 6 à 9 mois avec des écarts de prévision de moins de 30mm, et des corrélations moyennes entre les prévues et les observées.

➤ Prévision de la pluie mensuelle par l'Analyse en Composantes Principales

Les critères statistiques de validation obtenus par ce modèle sont donnés par le tableau suivant :

Tableau VI-11 : Résultats de prévision de la pluie par l'ACP

Année	Nash (%)	R ²
1980	44,46	0,47
1985	39,30	0,39
1990	-1,22	0,54
1995	-6,50	0,14
2000	-232,14	0,35

Des critères de validation très insuffisants apparaissent dans le tableau ci-dessus. On remarque un Nash et un R² qui se divergent l'un de l'autre pour les années 1990, 1995 et 2000. Les courbes de tendance et de corrélation sont représentées dans l'annexe VI-7. Les courbes de tendance et de corrélation de l'année 1990 sont données comme exemples par les figures VI-13 et VI-14.

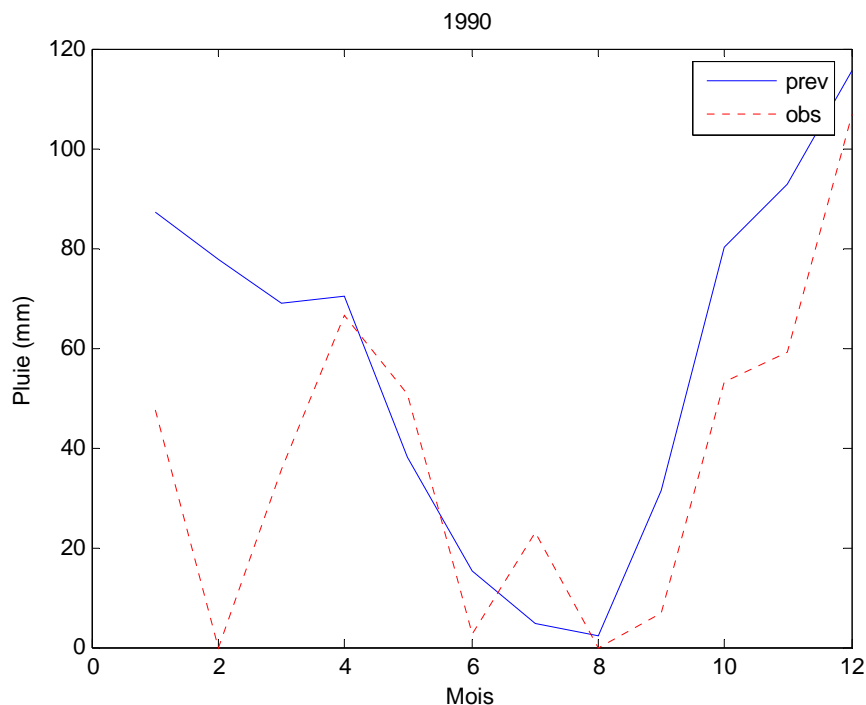


Figure VI-13 : Evolution des pluies mensuelles prévues et observées de l'année 1990

L'observation de la figure VI-13 nous permet de faire plusieurs constatations. Les pluies des mois secs de l'année 1990 sont prévues avec excellence. En fait, on n'a pas de pluies pendant les mois d'été. Les valeurs maximales des mois humides sont prévus avec des écarts négligeables. Le mois de Février a une pluie observée nulle, ce qui est contradictoire avec la réalité apprise par le modèle dans la phase de calage. Sinon, le modèle a réussi de prévoir 9 mois parmi les 12 avec des écarts de moins de 30mm.

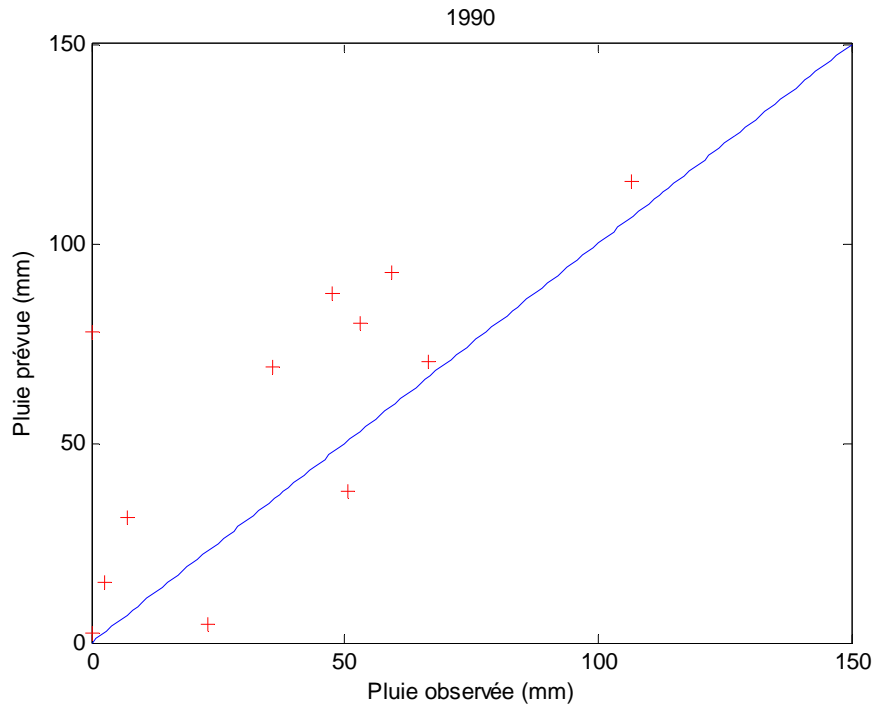


Figure VI -14 : Corrélation entre les valeurs prévues et observées de l'année 1990

Le nuage des points qui se trouve sur la figure VI-14 est dispersé, néanmoins on trouve six points qui sont proches de la première bissectrice. Donc six mois ont des précipitations prévues et observées corrélées.

En ce qui concerne la prévision des années 1980, 1985 et 1995, on remarque que le modèle ne permet pas d'avoir les valeurs maximales mais plusieurs mois sont prévus avec des écarts de moins de 30mm. Mais, il réussit à prévoir les pluies des mois secs. En moyenne, il arrive à prévoir six mois avec satisfaction. On peut avoir des bonnes corrélations entre les pluies prévues et observées pour quelques mois.

Pour l'année 2000, on voit clairement que les pluies des mois secs sont parfaitement prévues. On ne remarque que des écarts légers entre les pluies prévues et observées de la saison humide, à l'exception des mois de Janvier, Février, Mars, et Avril. Ces mois ont des valeurs de précipitation très faibles pendant cette année, ce qui s'oppose à la *connaissance* du modèle requise pendant le calage.

L'Analyse en Composante Principales a permis une prévision des mois secs meilleurs par rapport aux réseaux de neurones. Les grandes valeurs maximales de quelques années, ainsi que les mois ayant des pluies observées non conformes à la saison restent toujours un problème.

➤ **Prévision de la pluie mensuelle par le modèle combiné**

Les deux critères de validation sont représentés dans le tableau suivant :

Tableau VI-12 : Résultats de prévision de la pluie par le modèle combiné

Année	Nash (%)	R ²
1980	48,24	0,53
1985	25,38	0,27
1990	38,87	0,69
1995	-5,06	0,15
2000	-132,97	0,47

Les critères de validation du tableau VI-12 apparaissent moyens pour l'année 1980 et insuffisants pour les autres. On remarque une divergence entre les deux critères pour les années 1990, 1995, et 2000.

Les graphes de l'évolution et de corrélation entre les valeurs prévues et observées sont donnés Dans l'annexe VI-10. La figure VI-15 est l'évolution des pluies prévues et observées pour l'année 1990.

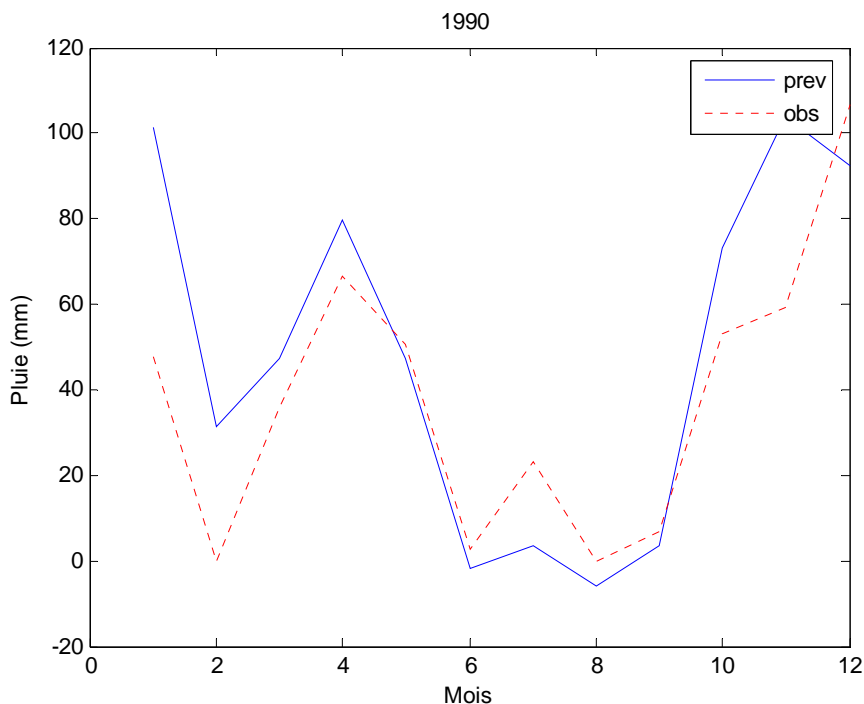


Figure VI-15 : Evolution des pluies prévues et observées de l'année 1990

En admirant la figure VI-15, on remarque que le modèle a permis une prévision très bonne des valeurs maximales avec des écarts inférieurs à 30 mm. L'écart de prévision de la pluie du mois de Février a diminué par rapport à celui obtenu dans les modèles précédents. Ce modèle a permis de prévoir 5 mois humides, mais il a échoué à reproduire les pluies des mois secs en donnant des valeurs négatives.

Pour les années 1980, 1985, 1995, on remarque que les prévisions des mois humides est atteintes pour la majorité. Néanmoins, les pluies des mois secs sont parfaitement reproduites.

L'année 2000 est une année exceptionnelle, elle est caractérisée de la sécheresse pendant les mois Janvier, Février, Mars, Avril. Pour cette raison la prévision de ces mois est délicate du fait que le modèle apprend que c'est des mois fortes précipitations. Sinon, les précipitations des autres mois de la saison hivernale sont prévues avec des écarts inférieurs à 30mm. La pluie de la saison d'été de cette année est reproduite avec excellence par le modèle combiné.

On conclue que le modèle combiné, étant appliqué pour la prévision des pluies mensuelles a fournit des résultats semblables à ceux obtenus par l'ACP, sauf pour l'année 1990. Pour cette année le modèle a résolu le problème du moi de Février mais il a fait des mauvaises prévisions des mois secs.

VI-2-2-2- Prévision de la pluie décadaire

Sauf les modèles des réseaux de neurones et d'Analyse en Composantes Principales sont testés dans cette partie du travail.

➤ Prévision de la pluie décadaire par les réseaux de neurones

Tableau VI-13 : Résultats de prévision de la pluie par les réseaux de neurones

Année	Nash (%)	R ²
1980	31,73	0,36
1985	27,57	0,28
1990	14,01	0,11
1995	18,07	0,20
2000	12,34	0,04

Les résultats du tableau VI-13 se montrent insuffisants du fait qu'ils ne dépassent pas la moyenne. Les évolutions et les corrélations sont représentées dans l'annexe VI-4. La figure VI-16 représente l'évolution des pluies prévues et observées de l'année 1990.

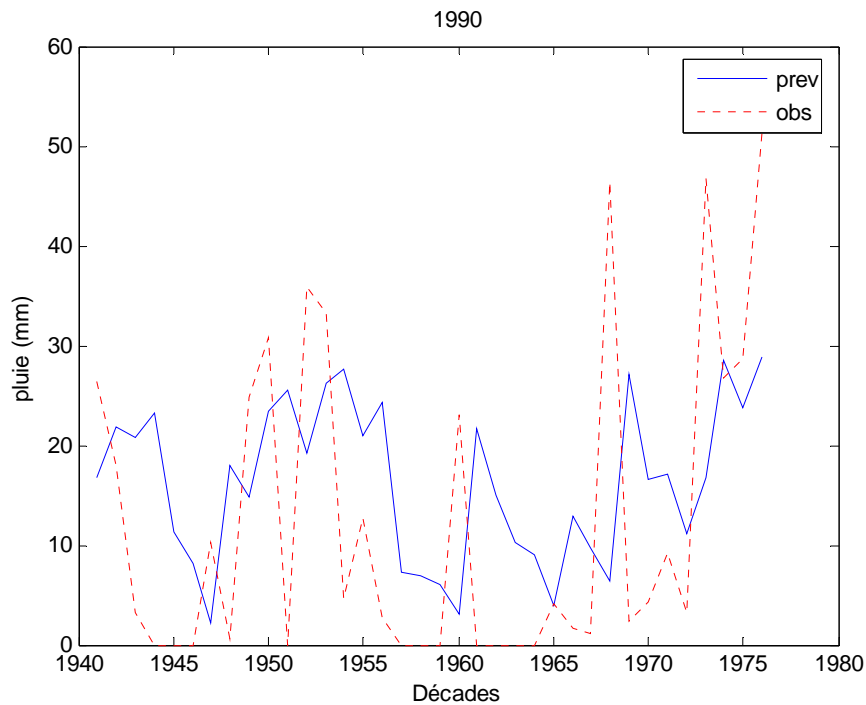


Figure VI-16 : Evolution des pluies décadaires et prévues de l'année 1990

Il est apparent que le modèle des réseaux de neurones n'a pas pu prévoir des valeurs maximales et il s'est confronté avec le problème de prévision des décades se trouvant pendant le moi de Février. Pour la même raison que pour la prévision en mensuel, ces décades sont considérées par le modèle comme humides, donc il les attribue des valeurs de pluie dépassant les 20mm.

On remarque que pour les autres années, le modèle arrive à prévoir les valeurs des décades sèches. Pour les décades humides il n'arrive pas à prévoir les valeurs max. Mais la majorité des décades sont prévues avec des écarts de moins de 10mm.

➤ Prédiction de la pluie par l'ACP

Les critères de validation sont représentés par le tableau suivant :

Tableau VI-14 : Résultats de prévision de la pluie par l'ACP

Année	Nash (%)	R ²
1980	35,95	0,37
1985	19,33	0,20
1990	4,49	0,29
1995	0,83	0,07
2000	-89,86	0,08

Les critères statistiques de validation apparaissent insuffisants. Mais les observations faites sur les graphes de l'annexe VI-6 donnent plus d'information sur la qualité de prévision par le modèle

d'ACP. La figure VI-17 représente les courbes de tendance des pluies prévues et observées de l'année 2000.

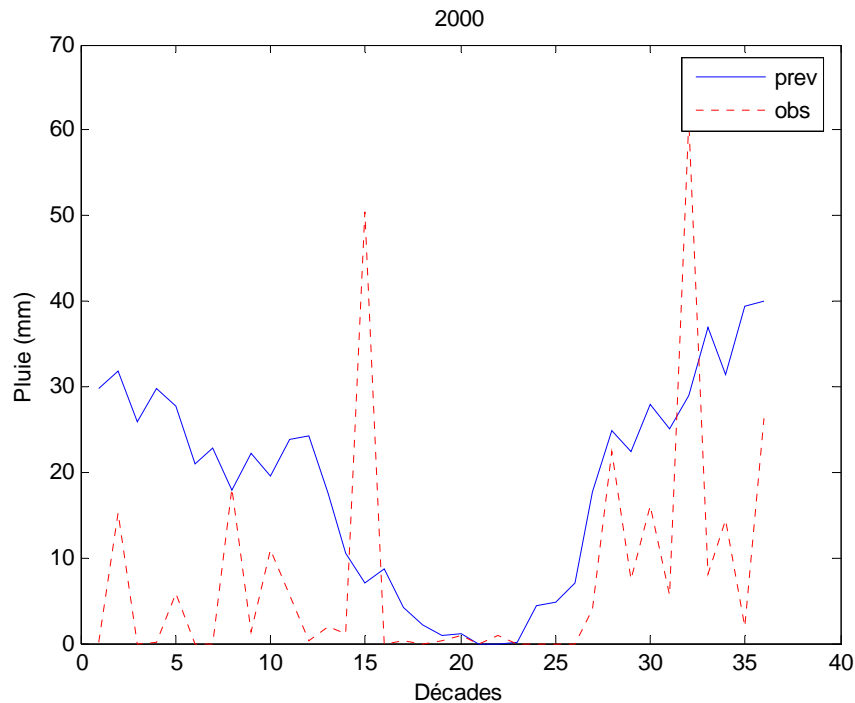


Figure VI-17 : Evolution des pluies observées et prévues de l'année 2000

De la figure VI-17, on constate que le modèle a prévu les décades sèches avec réussite, mais il a montré un échec lors de la prévision des valeurs maximales. Et il a attribué des valeurs prévues, grandes de plus de 10mm à celles observées, aux décades situées en mois de Janvier, Février, Mars, et Avril, du fait qu'il les considère comme humide.

Pour les autres années, le modèle arrive à prévoir les décades sèches, et il n'arrive pas à prévoir les valeurs maximales.

Pour la prévision de l'année 1990, le problème des décades du moi de Février apparaît toujours.

VI-2-2-3 Conclusion

La problématique de prévision de la pluie est très complexe. Les modèles traités dans ce travail ont échoué à la prévoir avec exactitude.

Les réseaux de neurones ont permis une prévision des mois et des décades secs sauf pour quelques exceptions. Les *valeurs maximales* sont prévues en mensuel généralement, et non pas en décadaire, par ce modèle. On arrive à prévoir 6 à 9 mois avec des écarts inférieurs à 30mm. Les décades prévues avec un écart de moins de 10mm peuvent atteindre les 20 décades.

De même l'ACP, a donné une prévision excellente des mois et des décades secs. Mais ce modèle reste incapable à prévoir les valeurs maximales. On trouve, néanmoins des mois prévus avec satisfaction.

Le modèle combiné appliqué à la prévision des pluies mensuelles, a amélioré les critères statistiques de validation de certaines années. Il a prévu les mois secs avec satisfaction. Les résultats de prévision des mois humides sont semblables à ceux obtenus par les deux autres modèles. Mais, il a ramené de plus pour la prévision de Février 1990.

Le problème de Février 1990 et de Janvier, Février, Mars, Avril 2000, S'est imposé toujours aux trois modèles. C'est un problème qui touche la qualité des données fournit aux modèles. Les valeurs des pluies nulles risquent d'engendrer des erreurs sur la prévision.

Les critères statistiques adoptés ont eu des valeurs insuffisantes pour les trois modèles de prévision.

L'effet de variation du nombre d'années de calage n'apparaît pas à ce stade. Les résultats trouvés varient selon l'année de prévision considérée.

Conclusion générale

La pluie et l'évapotranspiration potentielle sont deux paramètres essentiels dans le calcul du bilan hydrique. Leur prévision apporte des intérêts majeurs pour la gestion des eaux destinées à l'irrigation. Par ailleurs, la prévision de la pluie permet de mieux dimensionner les ouvrages de retenue et d'assainissement.

Pour cela, cette étude a pour objectif d'élaborer des modèles de prévision des deux variables : pluie et ETP aux pas de temps mensuel et décadaire. Ce choix de pas de temps est bénéfique pour la gestion et la planification en domaine des ressources hydriques.

En fait, trois modèles ont été testés dans ce travail : les réseaux de neurones, l'ACP et un modèle combiné des deux précédents. Les résultats des trois modèles sont comparés afin de trouver le meilleur modèle.

Pour l'approche des réseaux de neurones, plusieurs modèles ont été testés en modifiant le nombre d'entrées du modèle. Il a été prouvé que le modèle à quatre entrées est celui qui a donné les meilleurs résultats pour la pluie et l'ETP. Donc, on a opté pour ce modèle pour la prévision de ces deux variables par les réseaux de neurones.

Nous avons choisi arbitrairement 3 années à prévoir, pour l'ETP décadaire et mensuelle, à savoir : 1980, 1985, 1992. En ce qui concerne la pluie mensuelle et décadaire, nous avons choisi les années 1980, 1985, 1990, 1995, et 2000 pour la prévision. Ce choix est purement aléatoire. Il a comme but de trouver l'influence de la variation du nombre des années de calage.

Par conséquent, les résultats trouvés par les trois modèles prouvent qu'il n'y a pas d'influence du nombre d'année de calage pour toutes les années et les deux variables. Les résultats de prévision varient selon l'année considérée.

L'évapotranspiration potentielle, étant une variable cyclique, a été prévue par les trois modèles avec excellence pour les deux pas de temps.

Le modèle combiné a apporté un plus dans le domaine de prévision de cette variable, du fait que les résidus sont faibles et ont une tendance cyclique.

Le modèle d'ACP reste celui qui a donné les meilleurs résultats en domaine de la prévision d'ETP mensuelle et décadaire.

Le régime pluviométrique algérien est très aléatoire. Le volume de précipitation varie d'une année à une autre d'une manière très significative. Problématique qui fait de la prévision de la pluie un domaine d'une complexité extrême.

D'après les résultats trouvés dans ce travail, aucun des trois modèles n'a pu prévoir la pluie de la station d'Alger de la bonne manière. Ceci n'est pas dû aux modèles mais à la qualité des données fournies. Cette *mauvaise qualité* ne s'exprime pas par la troncature de la série ou par l'existence

des valeurs erronées, mais plutôt par l'existence des valeurs nulles qui risquent d'induire les modèles en erreur. Le mois de Février de l'année 1990, peut être vu comme exemple dans ce cas.

Les graphiques montrent que tous les modèles ont réussi à prévoir les saisons sèches, ce qui ne pose jamais de problème. Par contre, ils ont tous échoué à prévoir les pluies maximales à cause toujours de la mauvaise façon d'introduire les données.

Nous avons confronté le problème des grands résidus variants d'une manière très aléatoire en testant le modèle combiné pour prévoir la pluie mensuelle. Ceci peut être expliqué par l'existence d'un effet saisonnier qui n'est pas tenu en compte.

Les critères statistiques apparaissent très insuffisants en prévision mensuelle et décadaire de la pluie. Ceci n'est pas du aux modèles choisis, mais plutôt aux données introduites. En effet, dans le cadre de cette étude, nous pouvons affirmer que les critères statistiques optés, à savoir le critère de Nash et le coefficient de détermination, n'ont pas une valeur significative dans le domaine de prévision de la pluie. La recherche d'autres critères de validation des modèles de prévision, apparaît une démarche très sage qui peut apporter des améliorations dans ce domaine.

Perspectives

- Tenir en compte de l'effet saisonnier en introduisant les données aux modèles
- Tester d'autres modèles de réseaux de neurones, autres que le perceptron multicouches, dans le domaine de prévision d'ETP et de pluie.
- Généraliser les trois modèles testés pour la station d'Alger, à d'autres stations algériennes.
- Prévoir les canicules.

BIBLIOGRAPHIE

ANDERSON, T.W., (1958) : An Introduction to Multivariate Statistical Analysis (Second edition : 1984). J. WILEY, New York.

BENZECRI, J.P. (1982). Histoire et préhistoire de l'analyse des données. DUNOD.

BERMAD, A., (2004) : Elaboration d'une méthodologie de simulation et de prévision des variables climatologiques aléatoires et cycliques par l'Analyse en Composantes Principales. Thèse de doctorat. Ecole Nationale Polytechnique, Alger.

BOX, G.E.P., JENKINS, G.M., (1970) : Time series analysis, forecasting and control, Holden Day, SAN FRANCISCO.

BOX, G.E.P., JENKINS, G.M., (1976) : Time series analysis, forecasting and control, Revised Edition. Holden Day, SAN FRANCISCO.

BURT, C., (1950) : The factoriel analysis of qualitative data. British J. of Statist. Psychol. 3, pp. 166-185.

CAMPOLO, M., ANDREUSSI, P., SOLDATI, A., (1999) : River flood forecasting with a neural network model. Water Ressources Reasarch, 35(4), pp. 1191-1197.

COULIBALY, P., ANCTIL, F., BOBEE, B., (1999) : Prévision hydrologique par réseaux de neurones artificiels. Etat de l'art. Journal of Civil Engineering Canada n° 26, pp. 293-301

DAGNELIE, P. (1998). Statistique théorique et appliquée. Tome 1, De Boeck Université, Bruxelles.

DAVALO, E., NAÏM, P., (1991) : Des réseaux de Neurones, Edition Eyrolles, Paris.

DELAGARDE, J. (1983). Initiation à l'analyse des données. DUNOD.

DREYFUS, G., (2002) : Réseaux de neurones : Méthodologies et applications. Edition Eyrolle, Paris.

DUBAND, D. (1989). Hydrologie statistique approfondie. I.N.P.G.

EMMANUEL, C., Bruno, R., (2006) : Les Séries Temporelles. Cours Module XML et Data Mining, Université de Versailles Saint-Quentin-en-Yvelines.

EL JABI, N., ROUSSELLE, J., (1987) : Hydrologie fondamentale. Ecole Polytechnique de Montréal.

FISHER, R.A., (1915) : Frequency distribution of the value of the correlation coefficient in samples from an indefinitely large population. Biometrika, 10, pp. 507-521.

- FIORDALISO, A., (1999) : Systèmes flous et prévision de séries temporelles. Edition HERMES.
- FRAMLING, K., (1992) : Les réseaux de neurones comme outils d'aide à la décision floue. Rapport de D.E.A. Ecole Nationale Supérieure des Mines, Saint-Etienne
- FRENCH, M.N., KRAJEWSKI, W.F., CUYKENDALL, R.R., (1992) : Rainfall Forecasting in Space and Time using a Neural Network, Journal of Hydrology, Vol. 137, p. 1-31.
- HOTELLING, H., (1933) : Analysis of a complex statistical variables into principal component. J. Educ. Psy., 24, pp. 417-441 & pp. 498-520.
- HOTELLING, H., (1936) : Relation between two sets of variables. Biometrika, 28, pp. 129-149.
- HSU, K., GUPTA, H., SOROOSHIAN, S., (1995) : Artificial neural network modeling of rainfall-runoff process. Water Ressources Reasarch, 31(10), pp. 2517-2530.
- JAMBU, M., (2000) : Méthode de base de l'analyse des données, Edition Eyrolles, Paris p. 155-185.
- JODOUIN, J. F., (1994) : Les réseaux de neurones, Edition Hermès, Paris .
- LAHSEN, R., (2005) : Modélisation de l'évapotranspiration potentielle par les réseaux de neurones. Projet de fin d'étude. Ecole Nationale Polytechnique, Alger.
- LE FABLEC, Y. (1999) : Prévision de trajectoires d'avions par réseaux de neurones. Thèse Informatique Fondamentale et Parallélisme. Ecole Nationale de l'Aviation Civile, Toulouse.
- LEK, S., DIMOPOULOS, I., DERRAZ, M., ELGHACHTOUL, Y., (1996) : Modélisation de la Relation Pluie - Débit à l'aide des Réseaux de Neurones Artificiels, Revue des Sciences de l'eau, Vol. 3, p. 319-331.
- LIAMAS, J. (1985). Hydrologie générale : principes et applications. GAETAN MORIN, Québec.
- MAHALANOBIS, P.C., (1936) : On the generalized distance in statistics. Proc. Nat. Inst. Sci. India, 12, pp. 49-55.
- MINNS, A., HALL, J., (1996) : Artificial neural network as rainfall-runoff models. Journal des sciences hydrologiques, 41(3), pp. 399-417
- MIQUEL, J., (2006) : Hydrologie statistique : Introduction à l'Etude des Processus Hydrométéorologiques, Application à la prédétermination des Débits de Crues. Cours. Ecole Nationale Des Ponts et Chaussée, Pari.
- MOKHNACHE, L., (2004) : Application des réseaux de neurones dans le diagnostic et la prédiction des isolations de haute tension. Thèse de doctorat. Ecole Nationale Polytechnique, Alger.

NASCIMENTO, N.O., (1995) : Appréciation à l'aide d'un modèle empirique des effets d'actions anthropiques sur la relation pluie-débit à l'échelle d'un bassin versant. Thèse de doctorat, ENPC. CERGRENE.

NASH, J.E., SUTCLIFFE, J.V., (1970) : River flow forecasting through conceptual models : Part I – a discussion of principles. *Journal of hydrology*, 10, pp. 282-290.

O'CONNELL, P.E., NASH, J.E., FARREL, J.P., (1970) : River flow forecasting through conceptual models : Part II – The brasna catchment at ferbzne. *Journal of hydrology*, 10, pp. 317-329.

PEARSON, K., (1901) : On lines and planes of closest fit to systems of points in space. *Phil. Mag.* 2, n° 11, pp. 559-572.

ROGNON, P., (1996) : Sécheresse et aridité, leur impact sur la désertification au Maghreb. *Cahier de sécheresse* n°7, pp. 287-297.

SAJIKUMAR, N., THANDAVESWARA, B.S., (1999) : A Non-Linear Rainfall-Runoff model using an Artificial Neural Network, *Journal of Hydrology*, Vol. 216, p. 32-55.

SEIDOU, O., ROUSSELLE, J., LEFEBVRE, M., LAUZON, N., RIBEIRO, J., (2002) : Modélisation de l'incertitude sur les séquences futures de débits en rivière. *Journal des Sciences Hydrologiques*, 47(3), pp. 367–385.

SHAMSELDIN, A. Y., (1997) : Application of a Neural Network Technique to Rainfall-Runoff Modelling, *Journal of Hydrology*, Vol. 199, p. 272-294.

SPEARMAN, C., (1904) : General intelligence, objectively determined and measured. *Amer. Journal of Psychology*, 15, pp. 201- 293.

TOUZET, C., (1992) : Les réseaux de neurones artificiels.

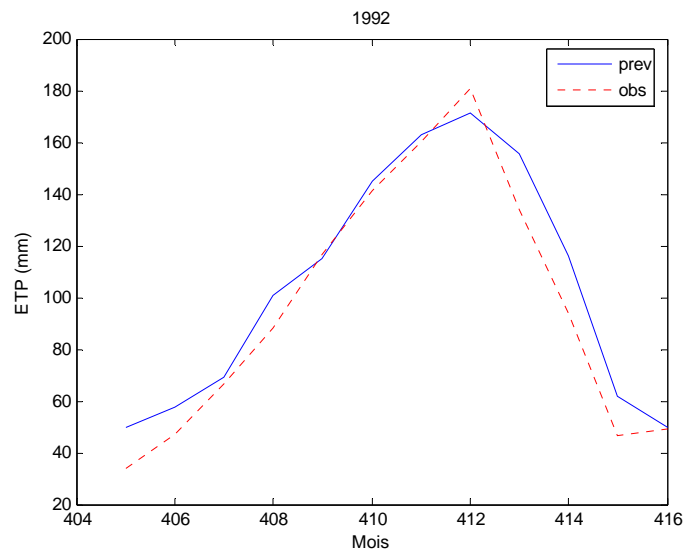
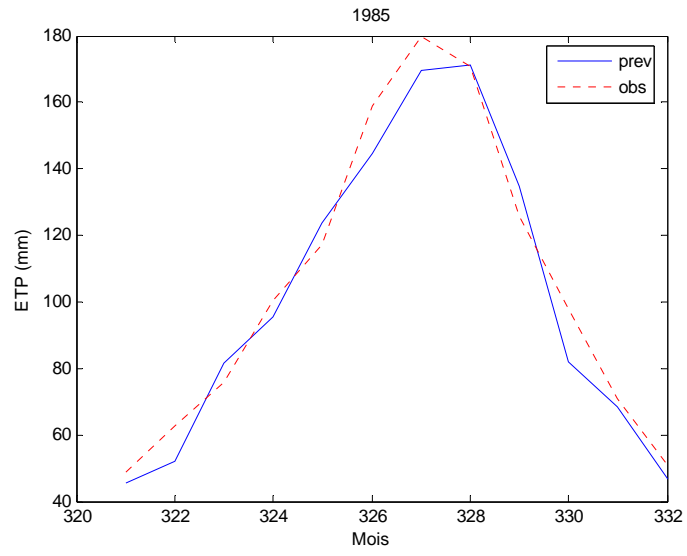
YAHIAOUI, G., ARIAT, J.L., (2002) *Techniques avancées pour le traitement de l'information*. Edition Cépaduès, Paris.

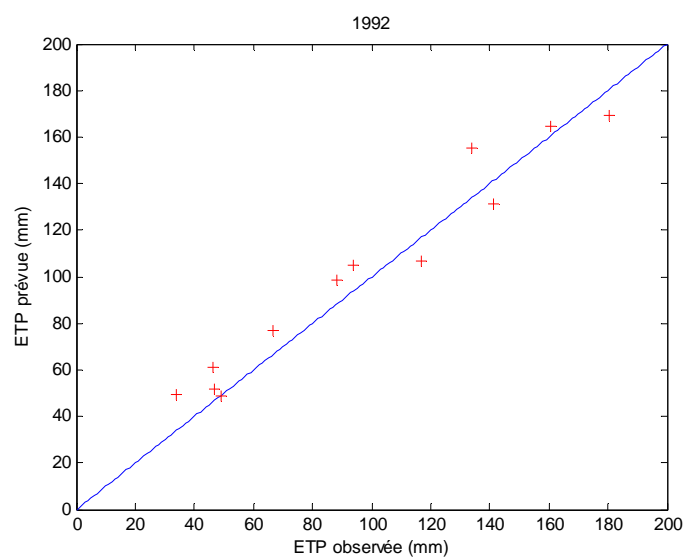
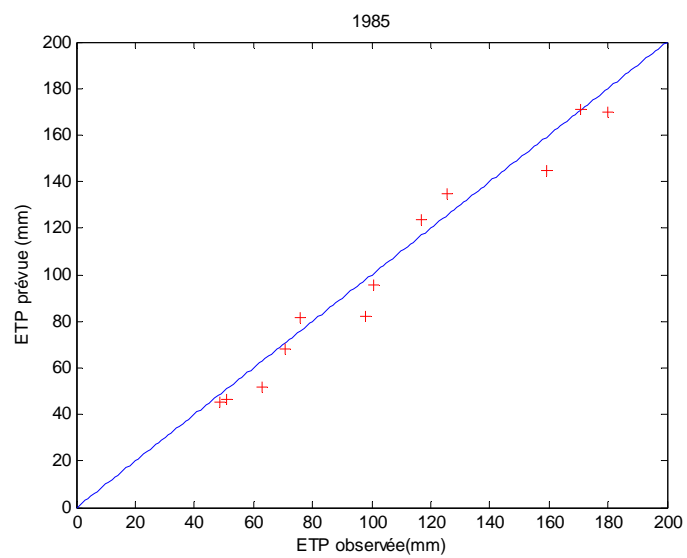
ZEALAND, C., BURN, H., SLOBODAN, S., SIMONOVIC, P., (1999) : Short term streamflow forecasting using artificial neural network . *J. of hydrology* 1999 N° 214. Pp 32 – 48.

ANNEXES VI

ANNEXE VI-1

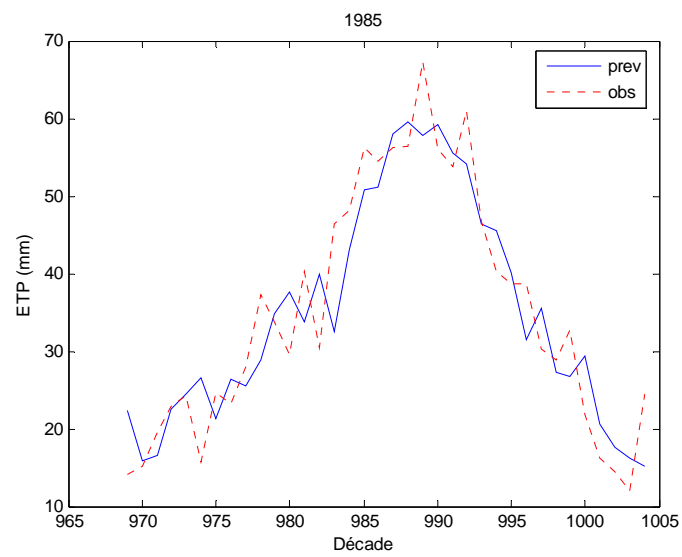
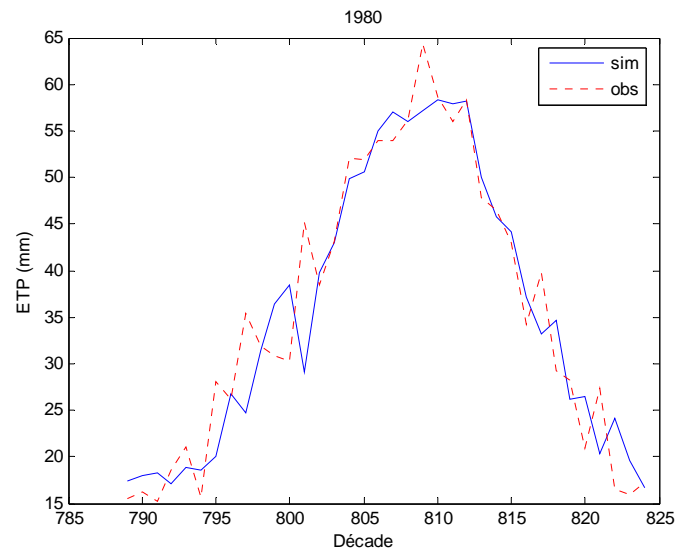
Prévision des ETP mensuelles par les réseaux de neurones

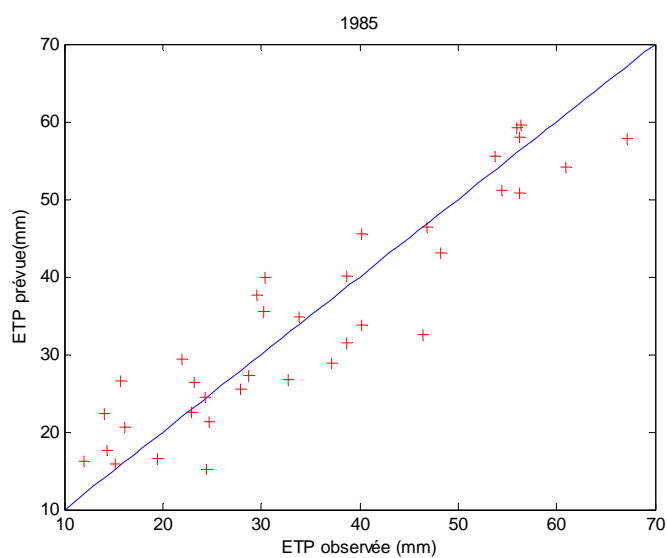
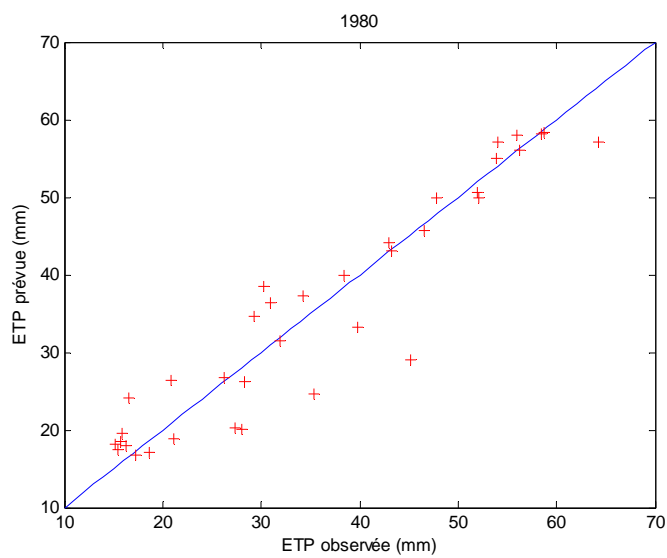




ANNEXE VI -2

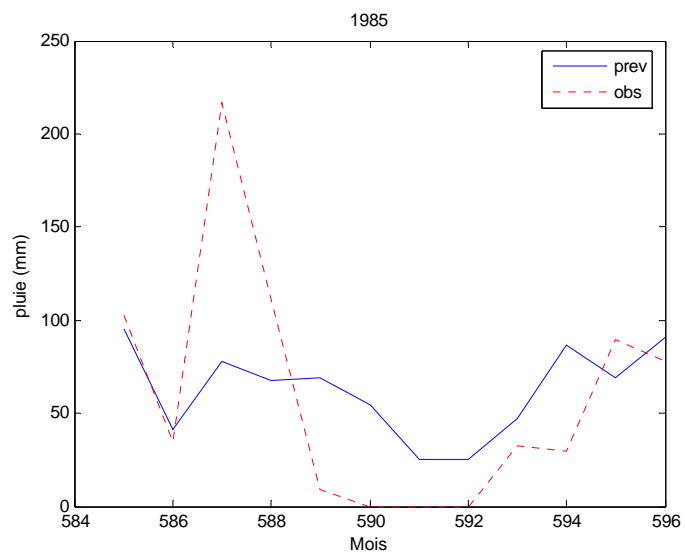
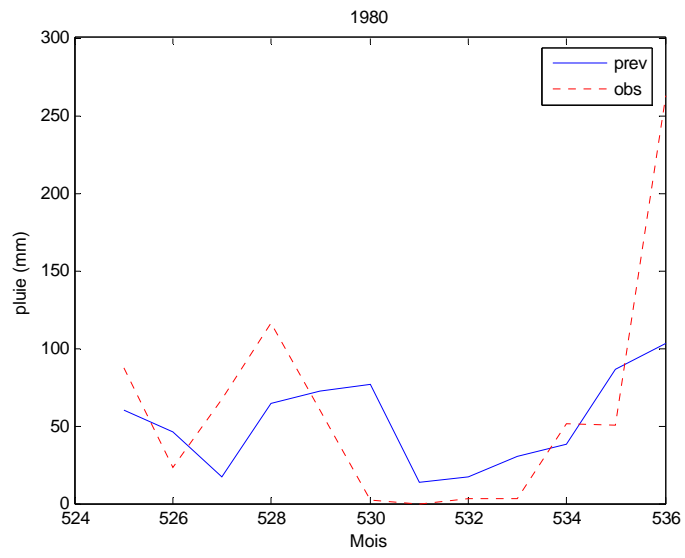
Prévision des ETP décadaires par les réseaux de neurones

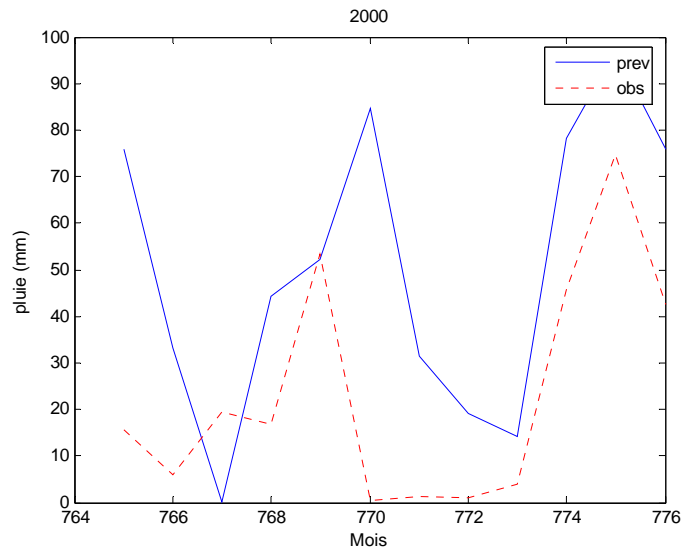
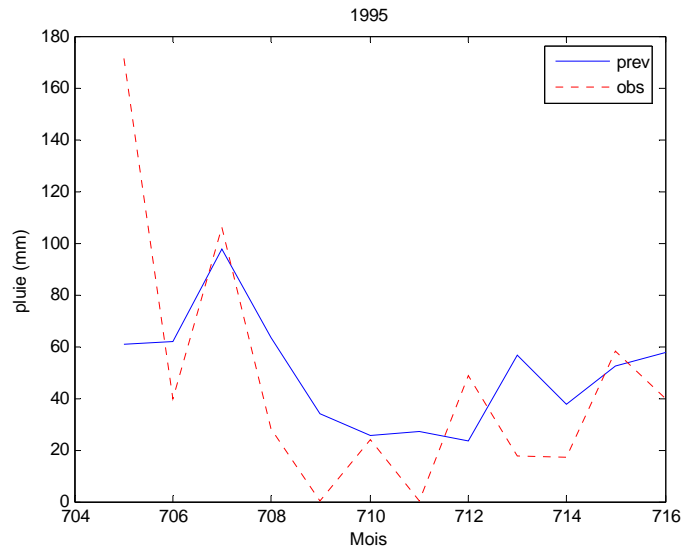


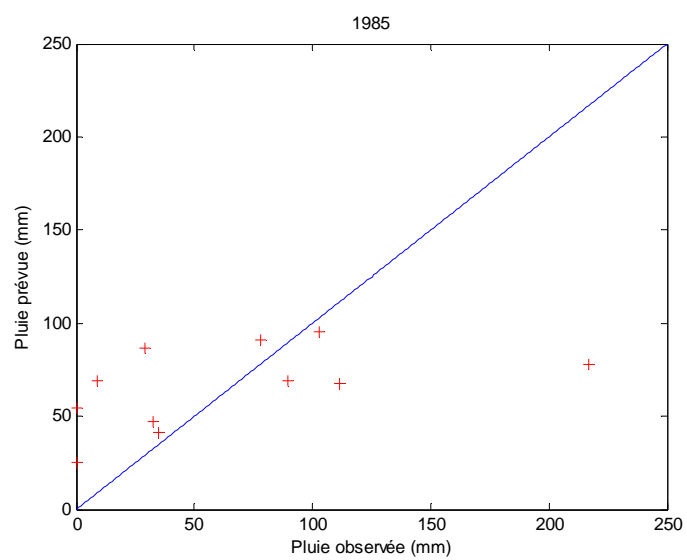
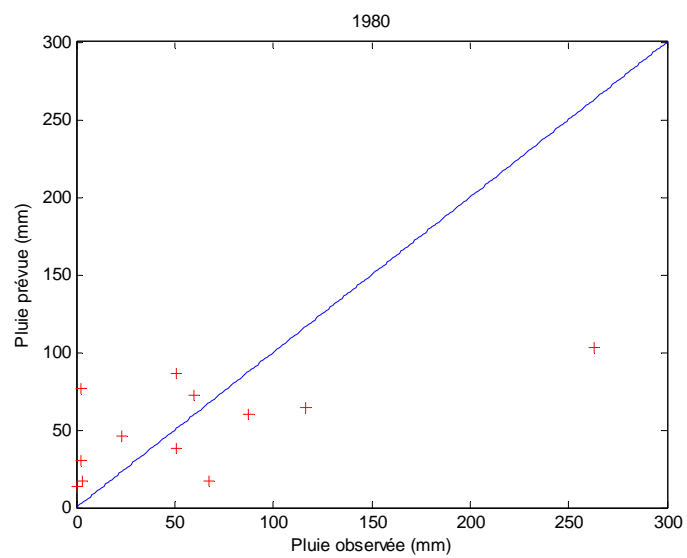


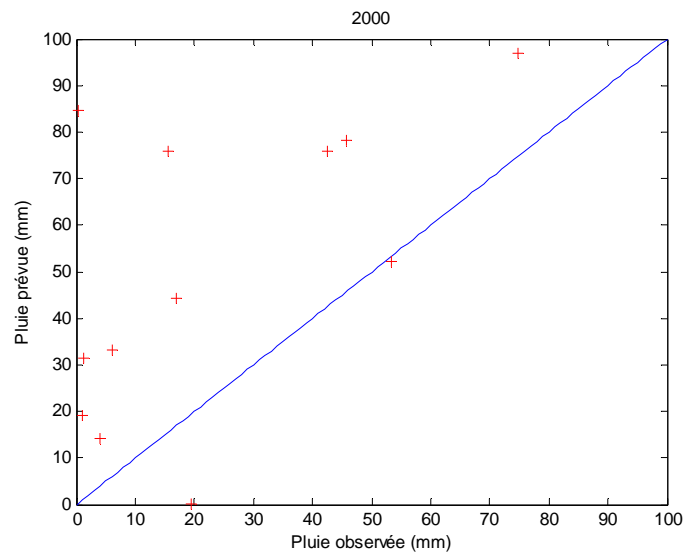
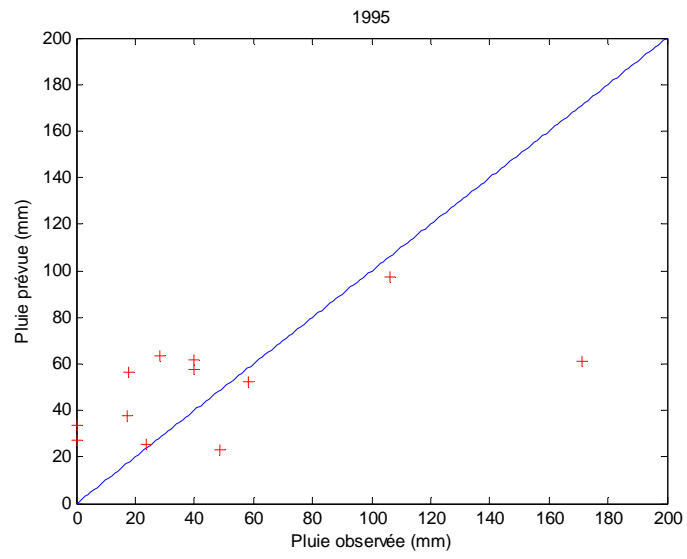
ANNEXE VI-3

Prévision de la pluie mensuelle par les réseaux de neurones



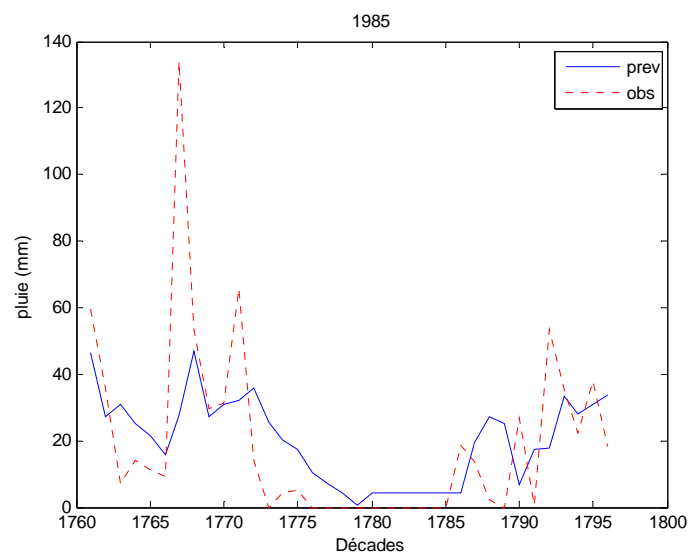
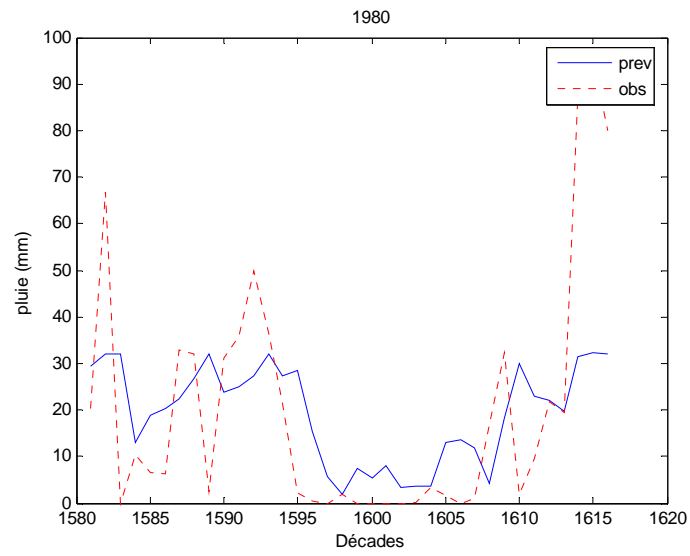


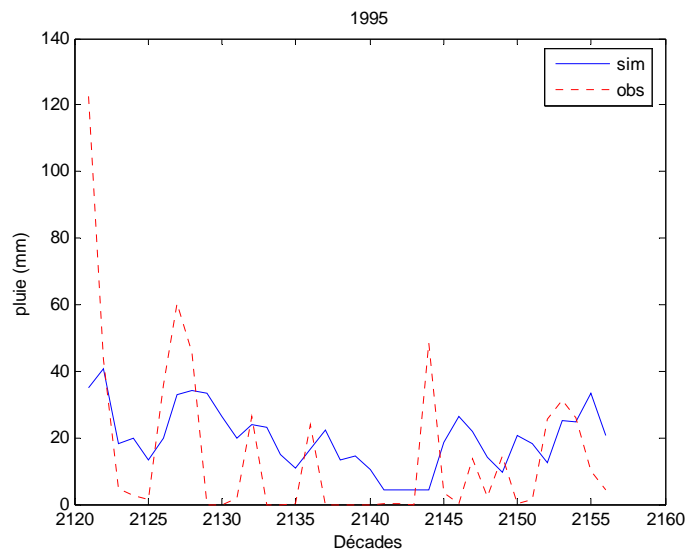
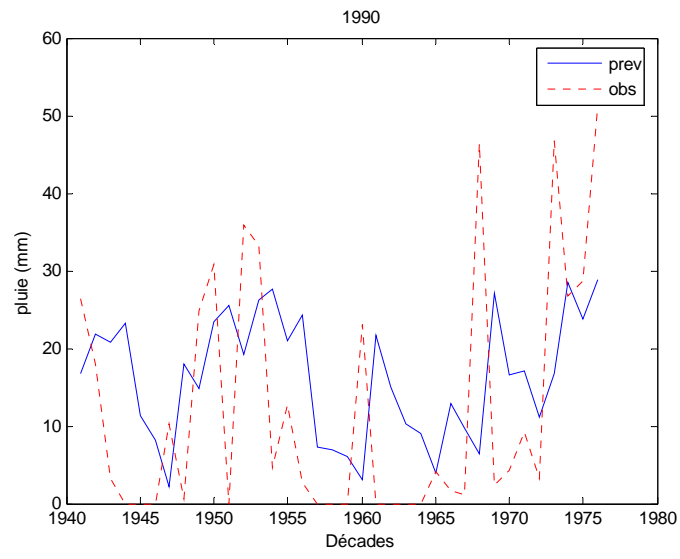


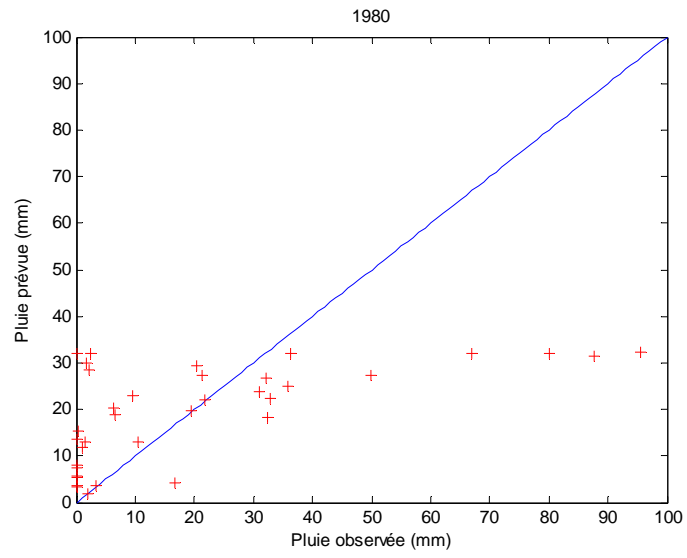
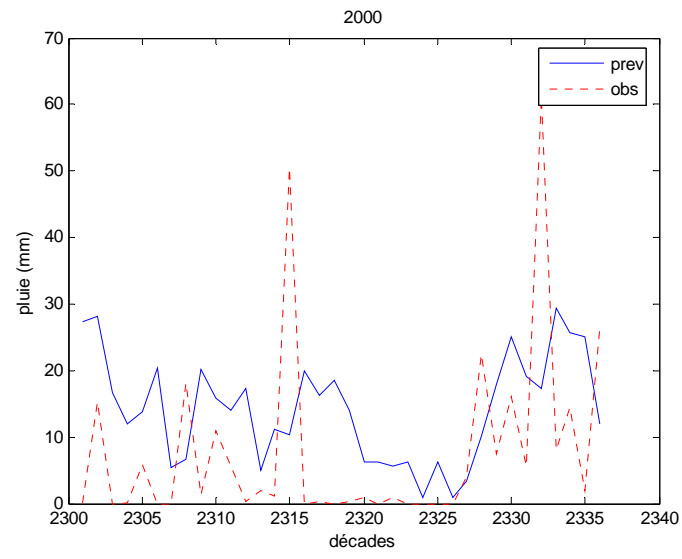


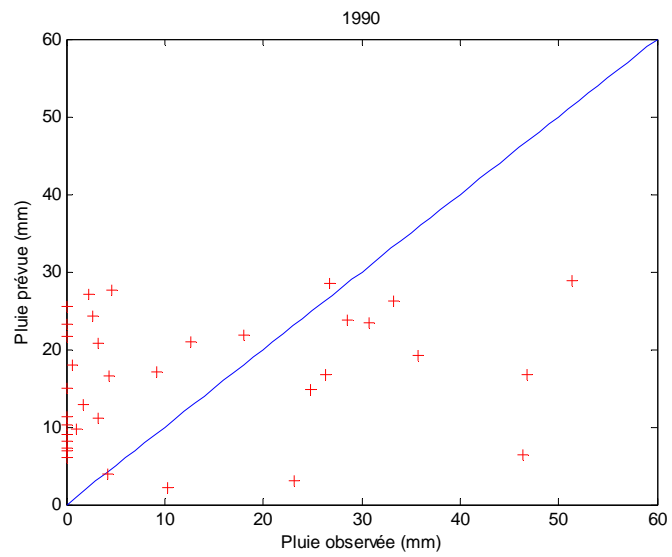
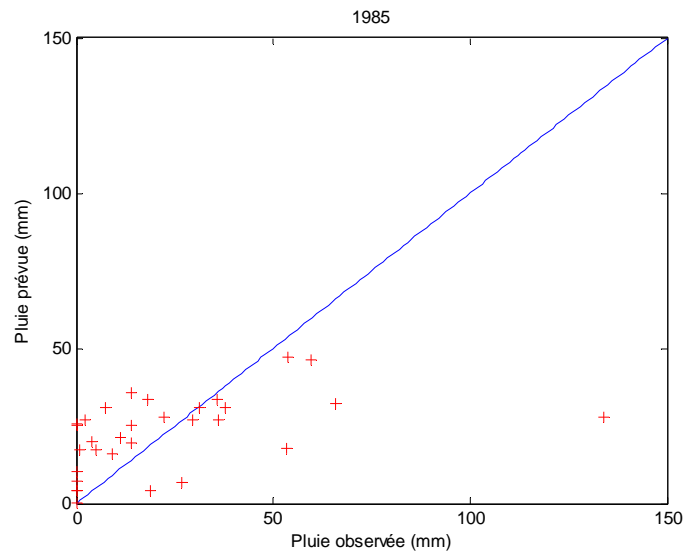
ANNEXE VI-4

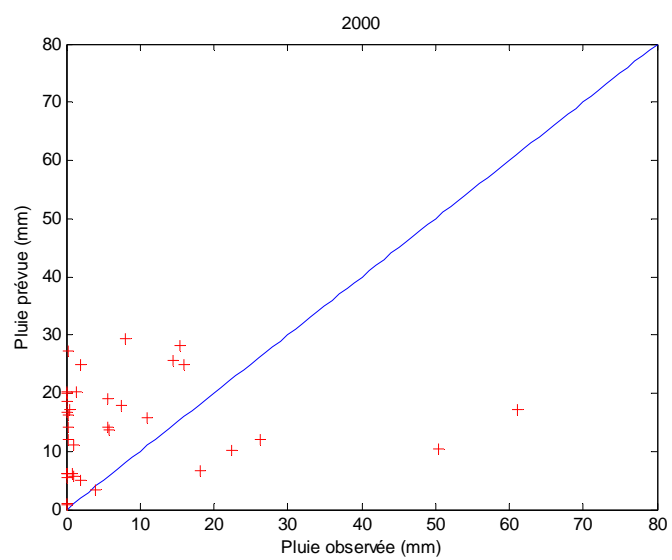
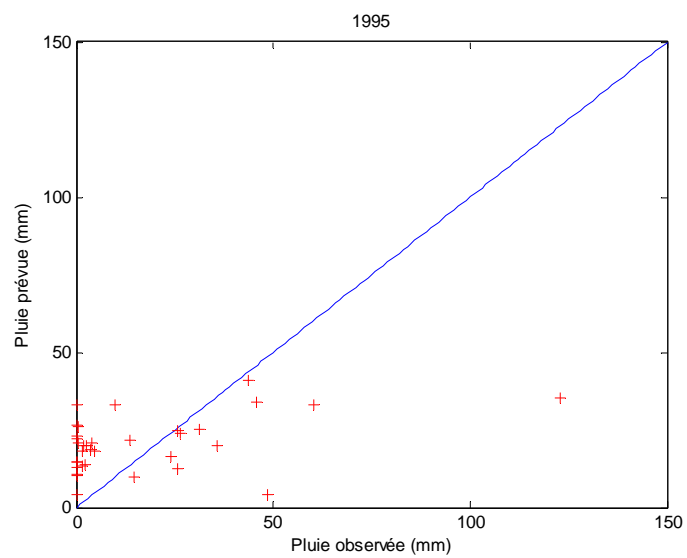
Prévision des pluies décadaires par les réseaux de neurones





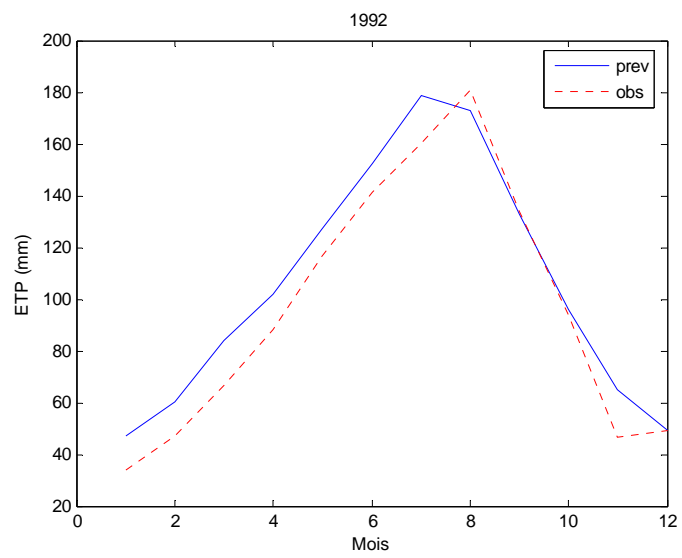
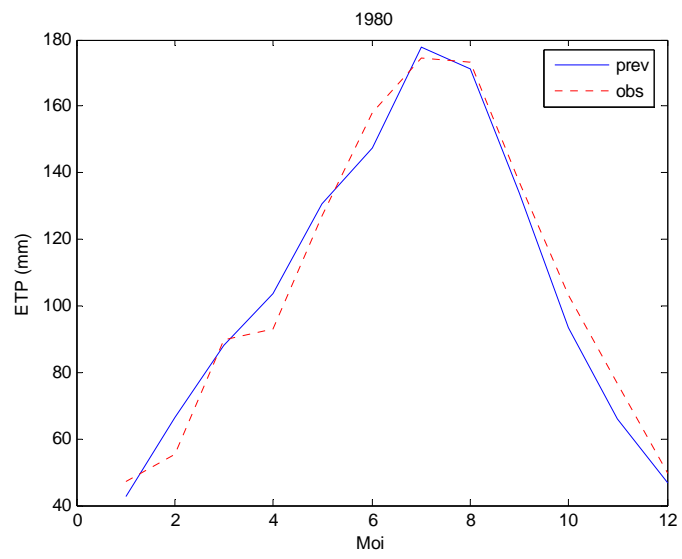


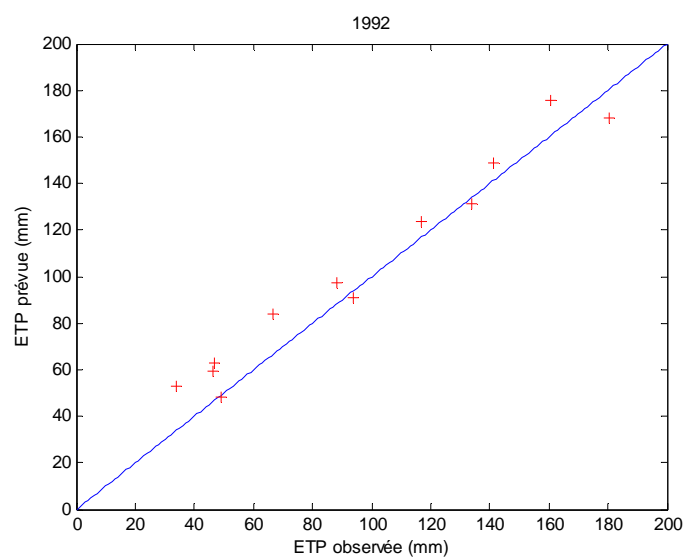
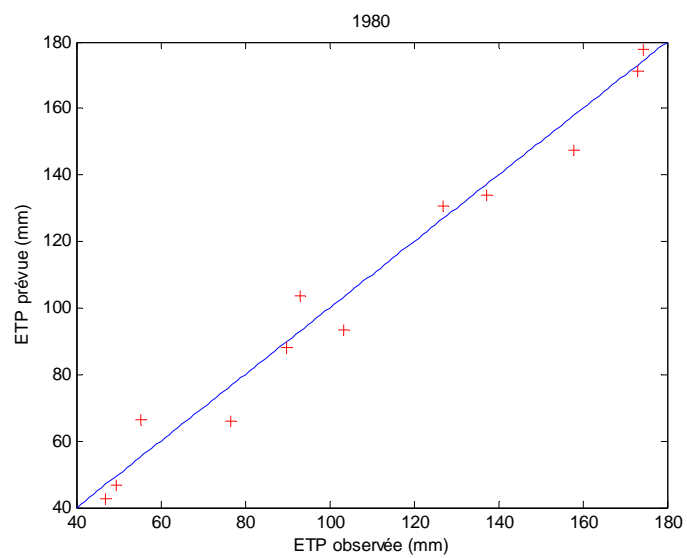




ANNEXE VI-5

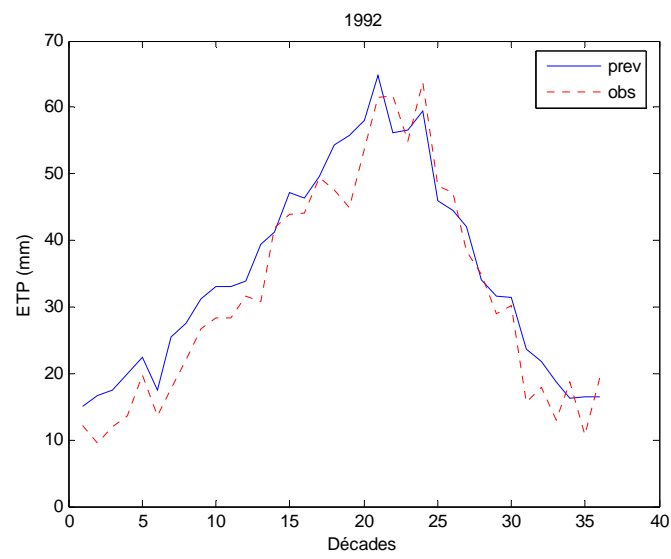
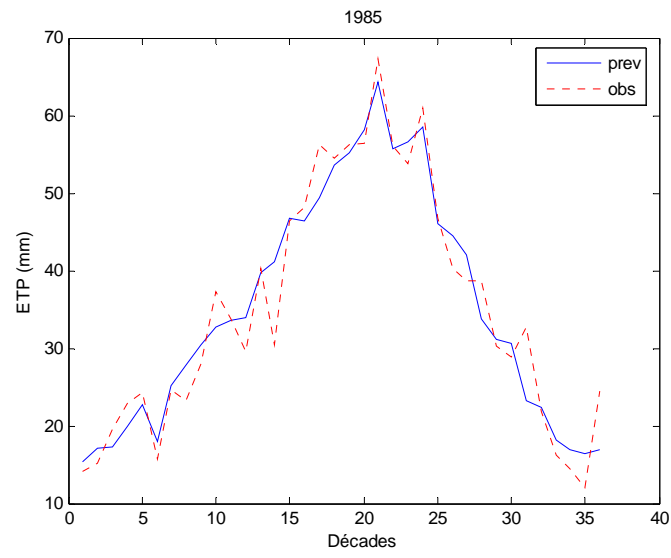
Prévision de l'ETP mensuelle par l'Analyse en Composantes Principales

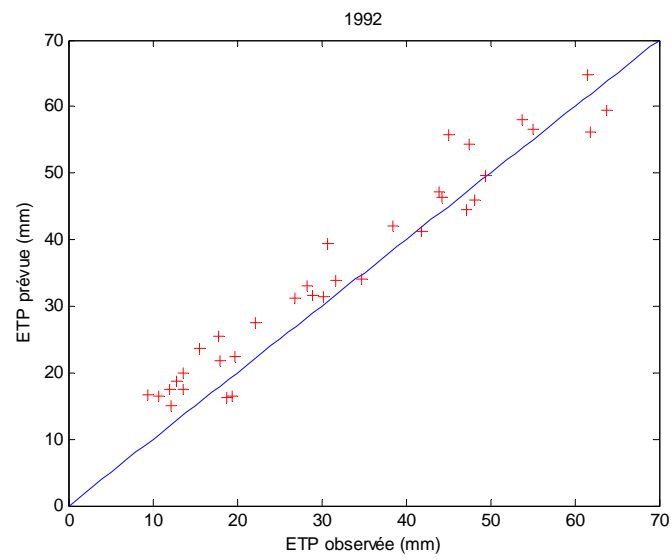
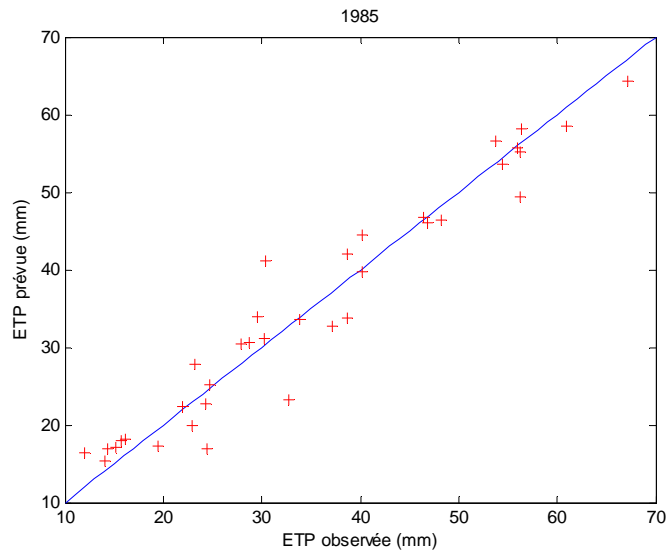




ANNEXE VI-6

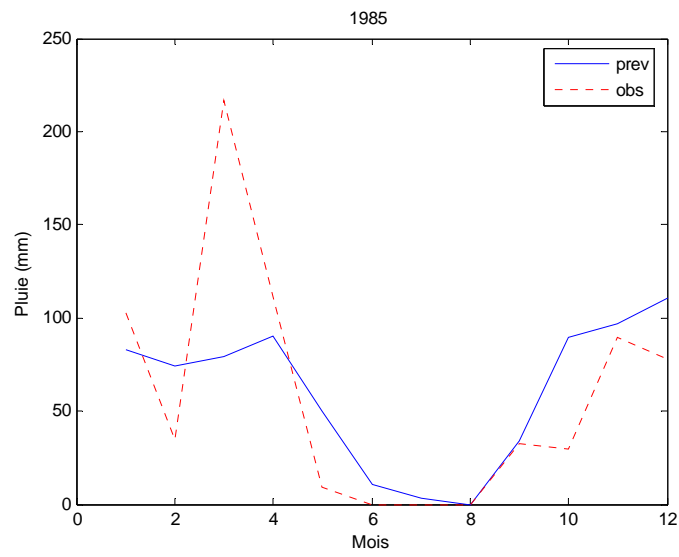
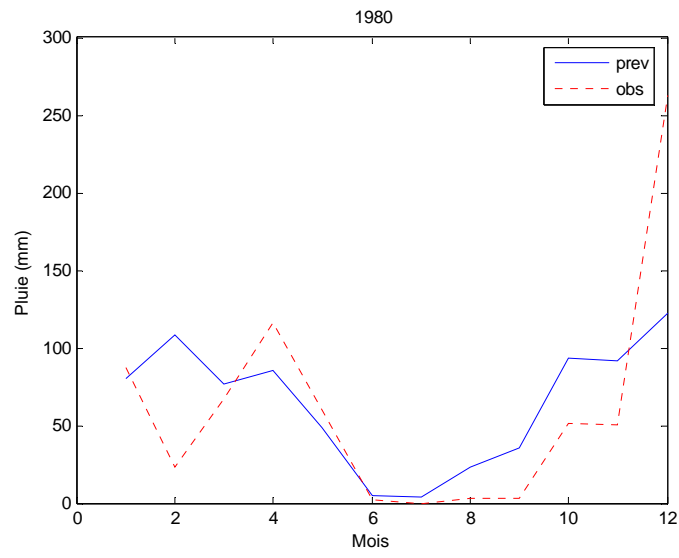
Prévision de l'ETP décadaire par l'Analyse en Composantes Principales

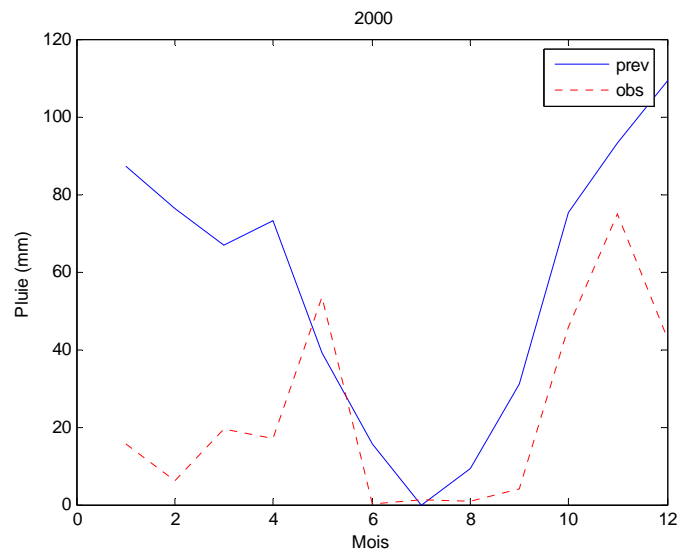
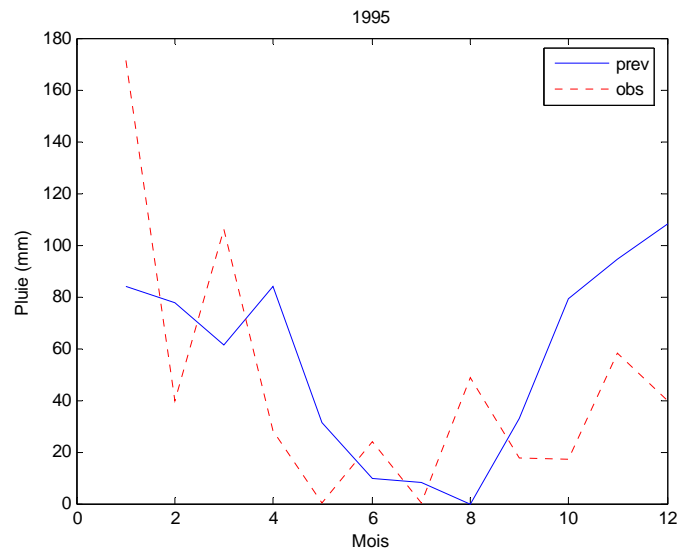


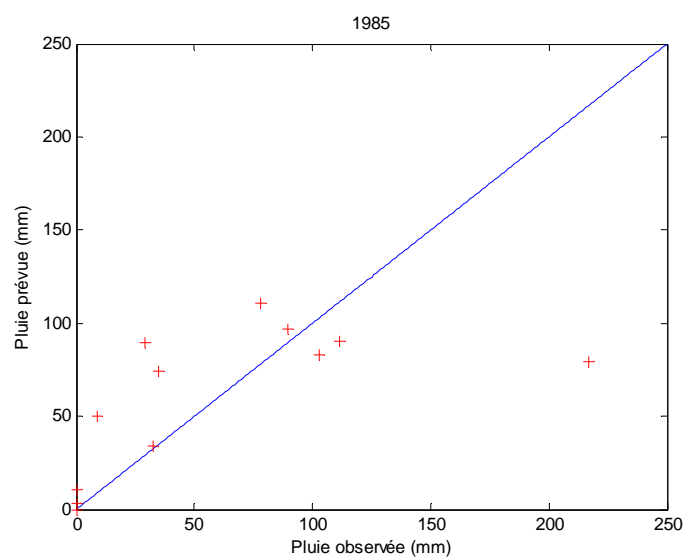
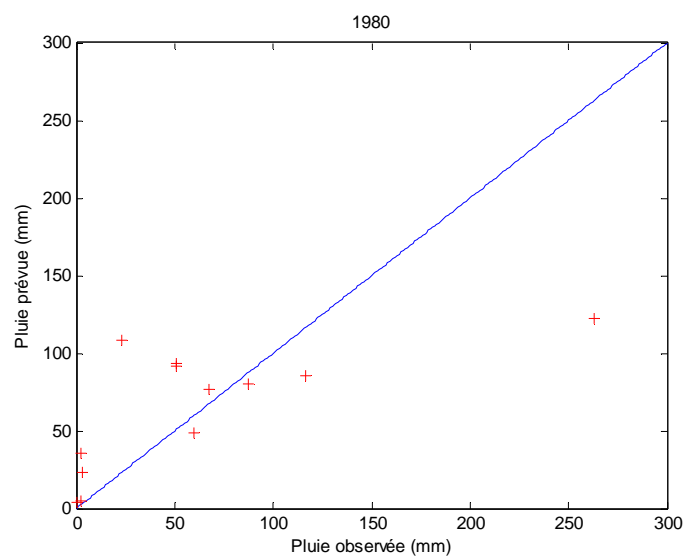


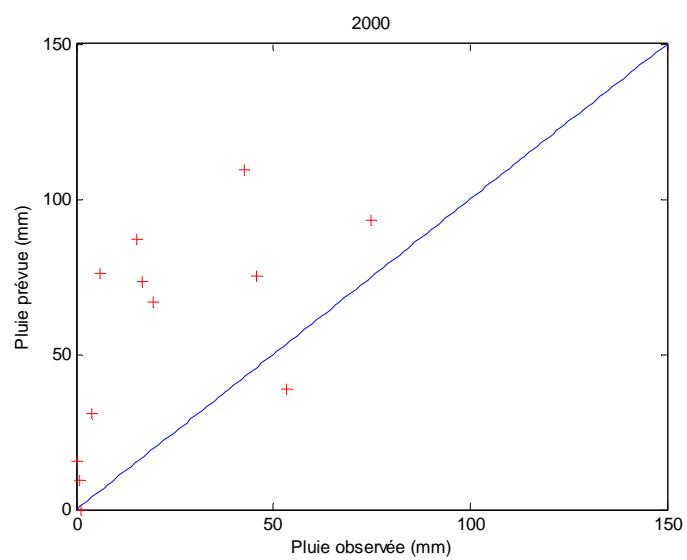
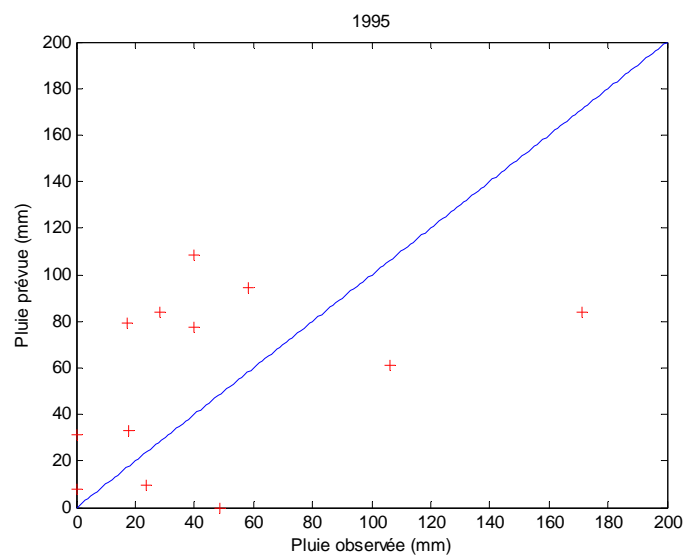
ANNEXE VI-7

Prévision des pluies mensuelles par l'Analyse en Composantes principales



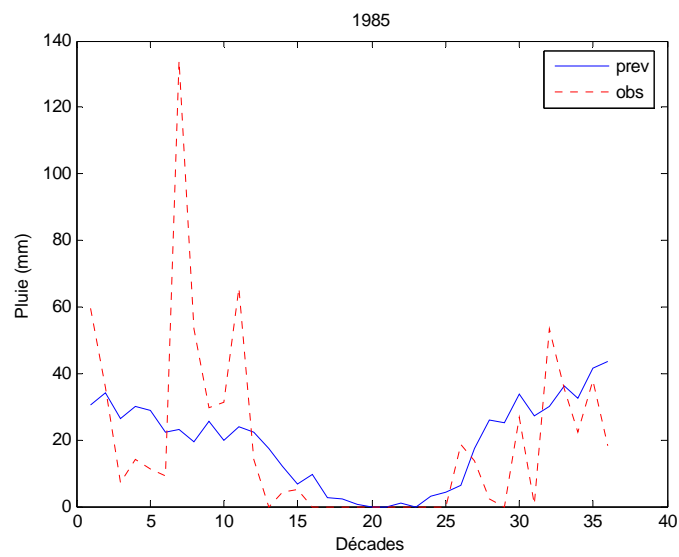
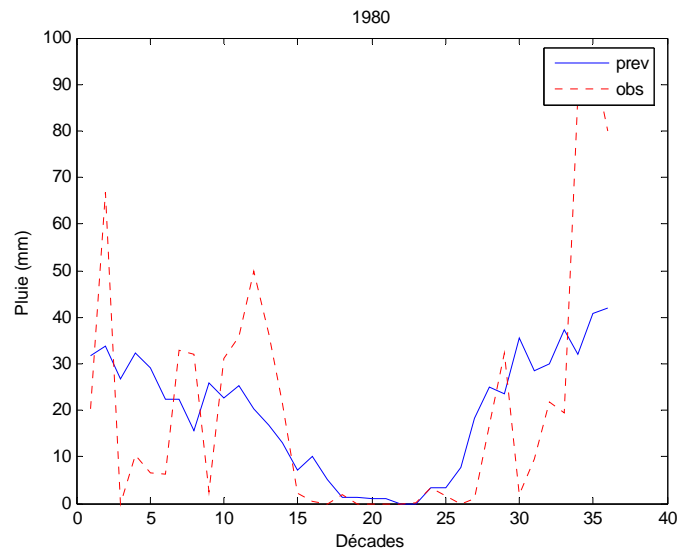


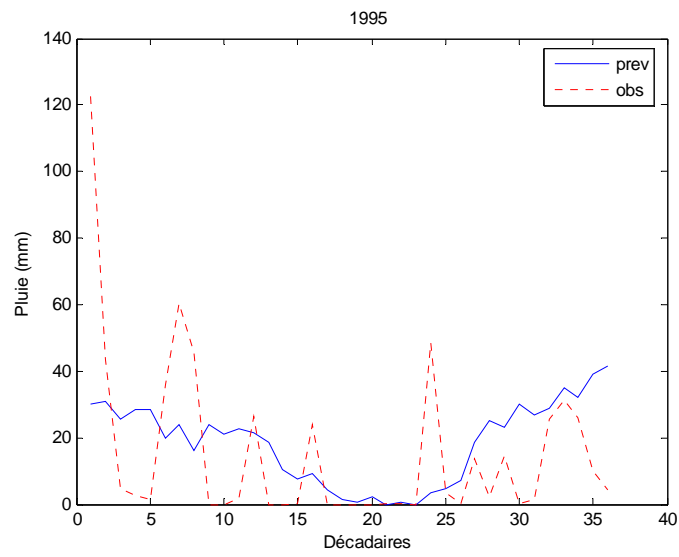
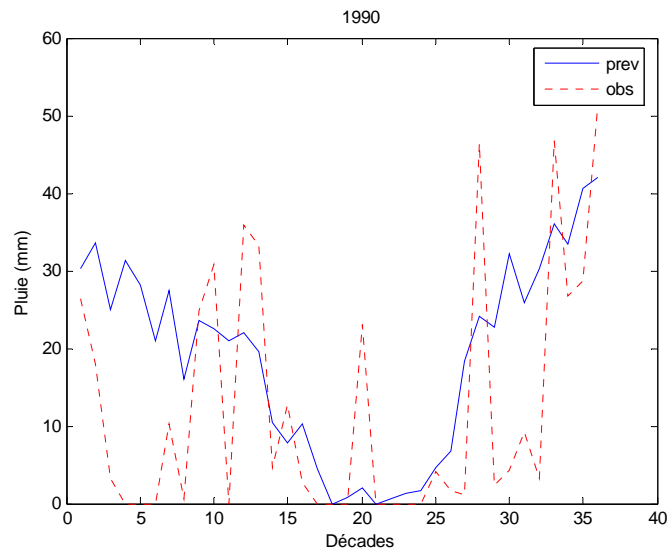


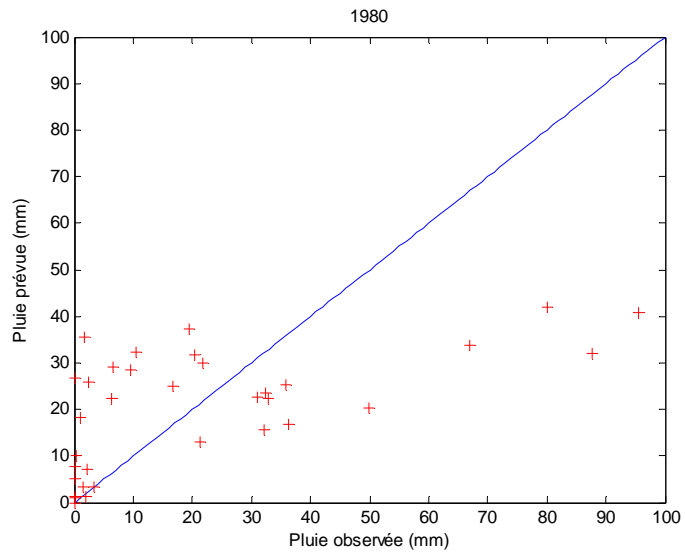
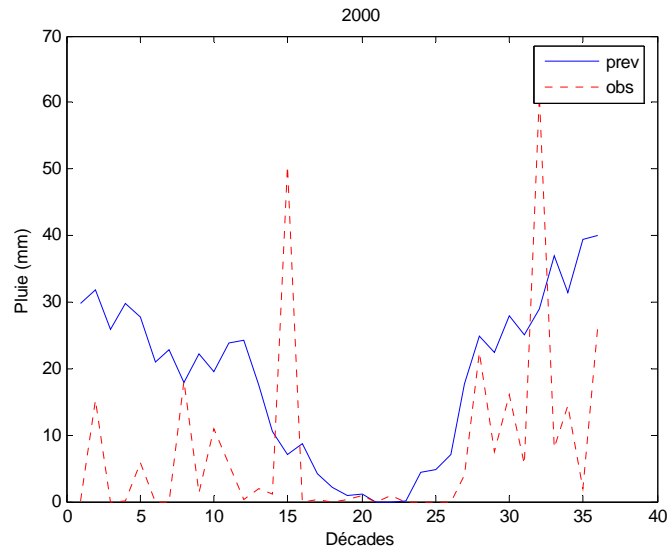


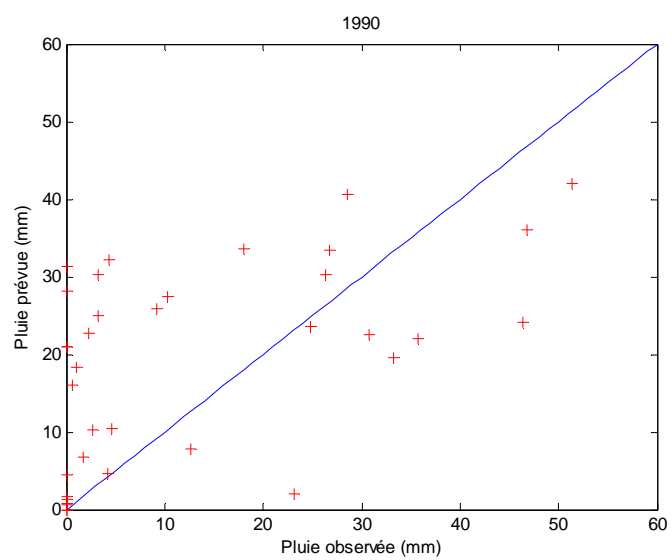
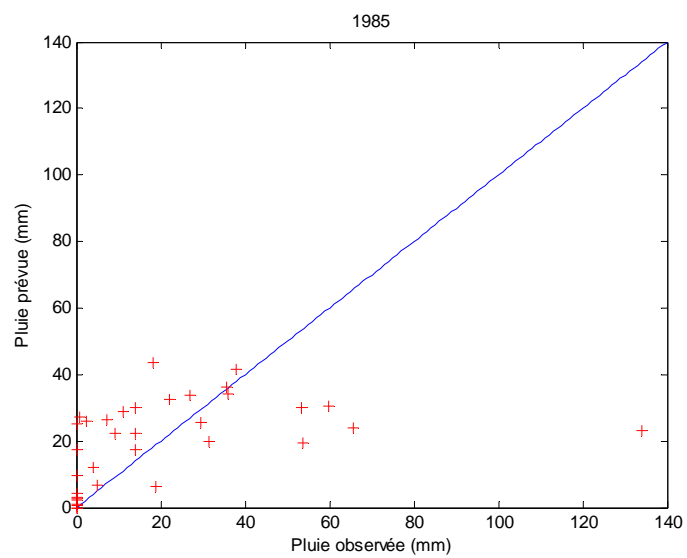
ANNEXE VI-8

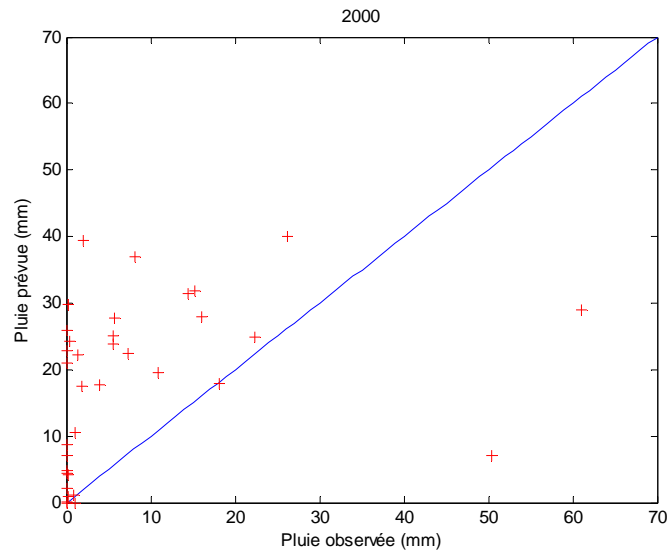
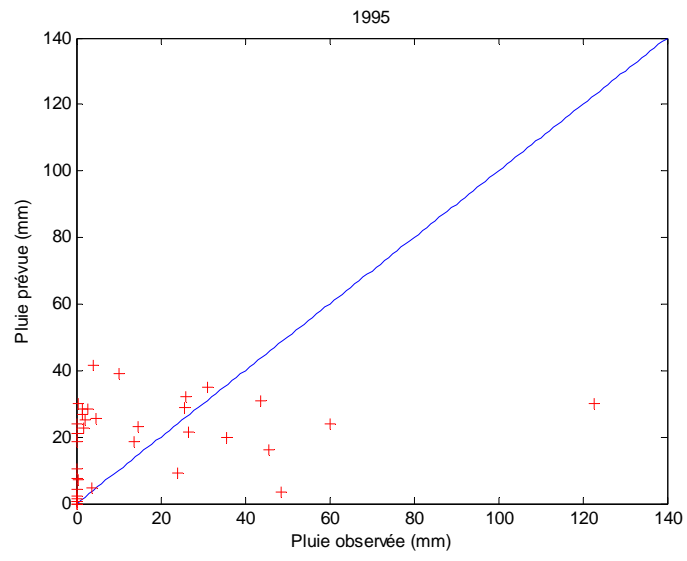
Prévision de la pluie décadaire par l'Analyse en Composante Principales





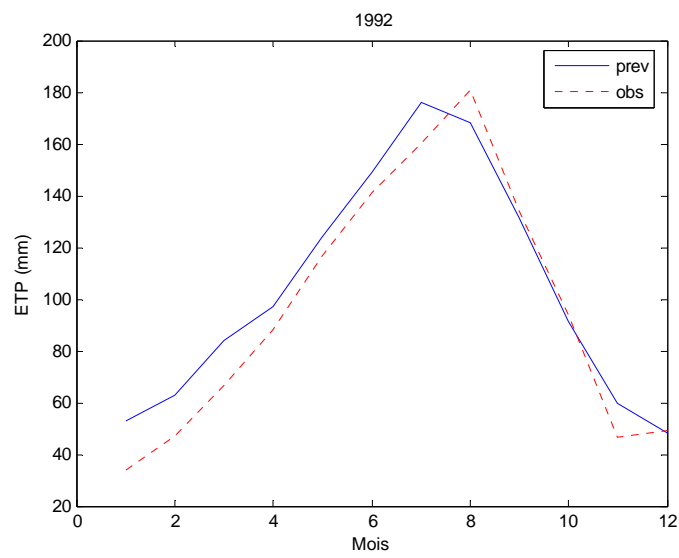
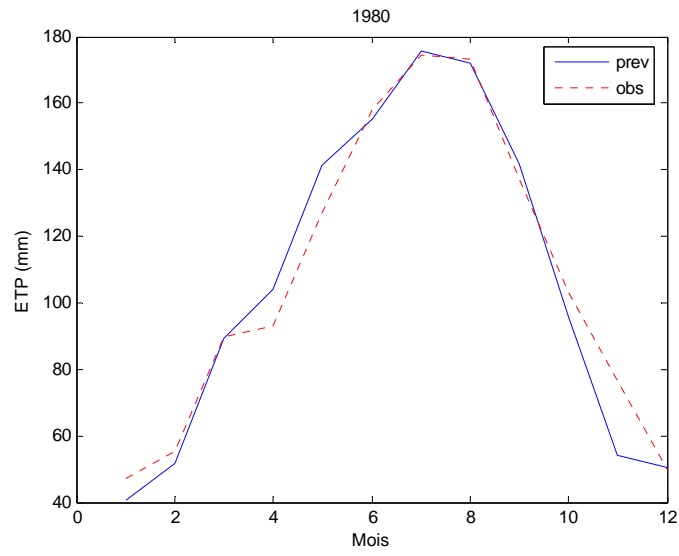


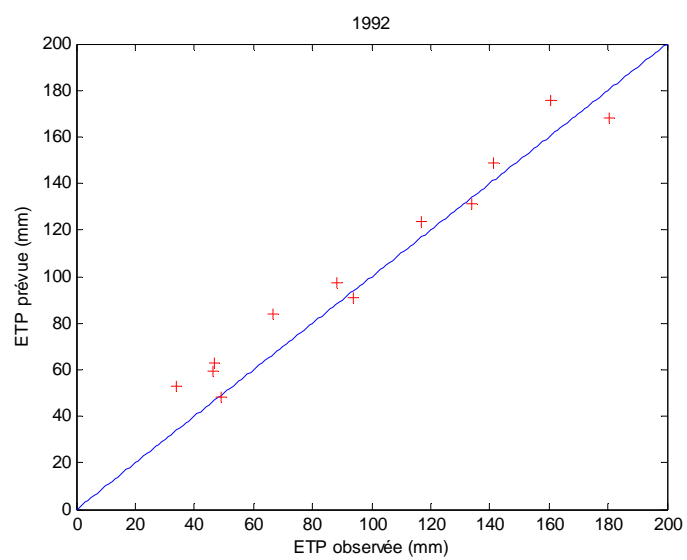
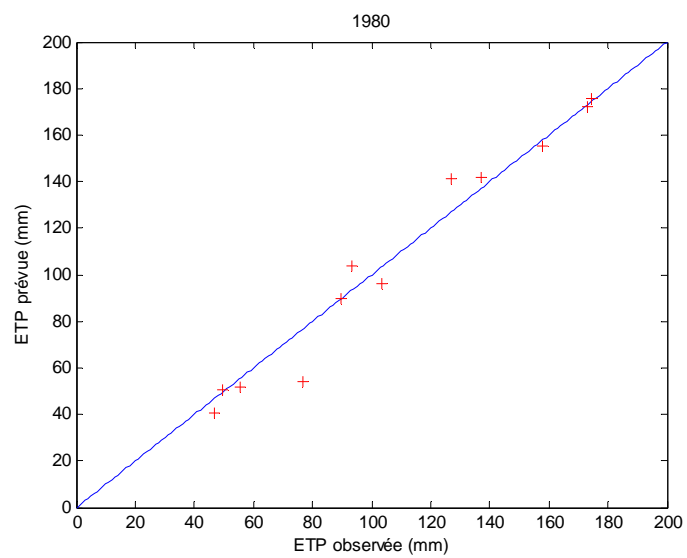




ANNEXE VI-9

Prévision de l'ETP mensuelles par le modèle combiné





ANNEXE VI-10

Prévision de la pluie mensuelle par le modèle combiné

