

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

M0001/03



المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

Ecole nationale polytechnique

Ecole Nationale Polytechnique
Laboratoire Construction et Environnement
Mémoire de Magister

Hydraulique
Présenté par
Djallal SAIDOUNI

**Modélisation de la relation
pluie débit au pas de temps
horaire par différentes
approches**

(DPFT - Réseaux de Neurones - ACP)

Membres du jury :

Président :	A. BALI	Professeur	ENP Alger
Rapporteur :	N. DECHEMI	Professeur	ENP Alger
Examineur :	B. TOUAÏBIA	M. de Conférence	ENSH Blida
Examineur :	M. CHERRARED	M. de Conférence	Univ. A. MIRA Béjaia
Examineur :	Y. MOULOUDI	Chargé de Cours	ENP Alger
Invité :	R. TAIBI	D. Général	ANRH Alger

Année Universitaire : 2002/2003

Ecole Nationale Polytechnique 10, av. Hassen BADI – BP 182- 16200 El Harrach- Alger- ALGERIE
Tél. : 213 21 52 14 94 fax : 213 21 52 79 29

Dédicaces

*Je dédie ce travail à ma mère et mon père qui m'ont
soutenu durant les moments difficiles*

A mes deux sœurs Imène, Sana et mon frère Mohamed ;

A mes deux grandes mères et à toute ma famille ;

*A mes amis Hacène , Omar, Ahmed, Ahcene , Assia, Fateh
, Tarik et Ali.*

Remerciements

A travers ce travail, réalisé au laboratoire de construction et environnement et au sein de l'équipe gestop, je tiens à remercier vivement tous ceux qui m'ont aidé à concrétiser ce mémoire et particulièrement mon promoteur, Mr le Professeur DECHEMI, qui m'a beaucoup aidé par ses précieux conseils et ses orientations scientifiques hautement appréciables. Mes remerciements vont également à ses collaborateurs dont Mr BERMAD.

Aussi je tiens à remercier Mr le Professeur BALI d'avoir accepté de nous honorer par sa présence en qualité de président du jury.

Je remercie également les membres examinateurs, Mme TOUAIBIA, Mr CHERRARED et Mr MOULOUDI qui ont consacré leur temps précieux pour examiner avec attention le travail présenté.

Par ailleurs, je tiens à exprimer ma gratitude à tous les enseignants qui ont contribué à ma formation durant le cycle d'étude.

SOMMAIRE



INTRODUCTION

01

CHAPITRE 1 *Revue bibliographique des modèles de prévision des débits*

1.1 La prévision des débits : formulation du problème	03
1.1.1 Utilisation des prévisions météorologiques	04
1.1.2 Calibration d'un modèle de prévision	04
1.1.3 Validation d'un modèle de prévision	05
1.2 Modèles empiriques	05
1.2.1 Modèles stochastiques	06
1.3 Modèle réseau de neurone	10
1.4 Régression linéaire floue	11
1.5 Les modèles conceptuels	12
1.6 Modèles à base de règles floues	13
1.7 Modèles débit-durée-fréquence	14
1.8 Méthode de la cartographie de l'écoulement	15
1.9 Utilisation de l'imagerie satellite	16

CHAPITRE 2 *Etude de la méthode de la différence première de la fonction de transfert (DPFT)*

2.1 Introduction	18
2.2 Principes de bases	18
2.2.1 Principe de base n° 1 : formulation en différences premières	18
2.2.2 Principe de base n° 2 : les pluies efficaces comme inconnues du problème	21
2.2.3 Principe de base n° 3 : une solution alternative itérée	21
2.3 Introduction des contraintes dans l'algorithme de résolution	22
2.3.1 Contrainte sur la fonction de transfert	22
2.3.2 Contraintes sur les pluies efficaces	24
2.4 Description de la méthode de résolution	24
2.4.1 Détermination de la fonction de transfert	24
2.4.2 Déconvolution des pluies efficaces	25
2.5 Techniques numériques utilisées dans la DPFT	25
2.5.1 Identification de la fonction de transfert	26
2.5.2 Déconvolution des pluies efficaces	26
2.6 Conclusion	28

CHAPITRE 3 *Modélisation par les Réseaux de Neurones*

3.1 Introduction	31
3.2 Les réseaux de neurones	31
3.3 Définition du neurone formel	32
3.4 Historique	33
3.5 Type de réseau	34
3.5.1 Réseau à deux couches	34
3.5.2 Réseau multicouche	35
3.6 Concept et fonctionnement	36
3.6.1 Pratique	37
3.6.2 Les fonctions d'activation	38
3.7 Architecture et optimisation d'un réseau	40
3.8 Apprentissage d'un réseau de neurone	40

SOMMAIRE



3.8.1	Apprentissage supervisé	40
3.8.2	Apprentissage semi-supervisé	44
3.8.3	Apprentissage non-superviser	44
3.9	Conclusion	45

CHAPITRE 4 L'Analyse en Composantes Principales

4.1	Généralités sur l'analyse de données	47
4.2	L'analyse en composante principale	48
4.2.1	Objectif	48
4.2.2	Notion d'individus et de variables	48
4.3	Définition de l'ACP	49
4.4	Définition algébrique de l'ACP	49
4.4.1	Représentation graphique des individus	49
4.5	Définition géométrique de l'ACP	50
4.6	Terminologie	51
4.7	Les différentes méthodes	51
4.7.1	L'analyse en composantes principale notée ACP _{nn}	51
4.7.2	L'analyse en composantes principale normée notée ACP	51
4.8	Description mathématique de l'ACP	52
4.8.1	Analyse des points individus	52
4.8.2	Analyse des points variables	53
4.9	Principe d'interprétation	54
4.10	Exploitation de l'analyse en composante principale	54
4.10.1	Objectif	54
4.10.2	Interprétation numérique de l'ACP	54
4.10.3	Interprétation graphique des axes factoriels	55
4.11	Procédé de calcul de l'ACP	56

CHAPITRE 5 Application sur les données du bassin versant de l'Oued DIS (SEBAOU) Résultats et Interprétation.

5.1	Introduction	60
5.2	Présentation du bassin versant	60
5.2.1	Caractéristique physique de bassin versant	60
5.2.2	Choix des épisodes averses crues	61
5.3	Application de la méthode DPFT sur le bassin de l'oued DIS	62
5.3.1	Procédure de résolution	62
5.3.2	Présentation des résultats et interprétation	64
5.3.3	Conclusion	68
5.4	Application du modèles réseau de neurones	69
5.4.1	Introduction	69
5.4.2	Critères de validation	69
5.4.3	Présentation et interprétation des résultats	69
5.4.4	Conclusion	73
5.5	Application de l'analyse en composantes principale	73
5.5.1	ACP sur les données utilisées	73
5.5.2	Elaboration du modèle de prévision	75
5.5.3	Reconstitution	77
5.5.4	Prévision	79

SOMMAIRE



CONCLUSION GENERALE

83

Bibliographie

85

Annexes

Résumé :

Dans le cadre de ce travail nous avons essayé de trouver un moyen scientifique pour la prévision de débits en particulier ceux engendrés par les crues, un problème sur lequel les scientifiques se penchent actuellement, sans pour autant avoir trouvé un palliatif.

C'est dans cette optique que nous proposons l'étude et l'application sur des données réelles de trois techniques de prévision à savoir la différence première de la fonction de transfert, les réseaux de neurones artificiels et l'analyse en composantes principale.

Cette étude sera un moyen de mettre en évidence la capacité et la fiabilité de chacune des méthodes et d'en sortir celle qui aura un meilleur impact sur la gestion du risque.

ملخص

لقد حاولنا في إطار هذا البحث إيجاد وسيلة فعالة لدراسة توقعات منسوب الماء خاصة الذي ينجر عن الفيضانات والذي يعد اليوم مشكلا عويضا لازال العلماء يولونه عناية فائقة دون أن يتوصلوا إلى نتيجة مرضية لتفادي هذا المشكل.

ففي هذا الموضوع وفي نفس الاتجاه نقترح دراسة وتطبيق على معطيات حقيقية لثلاثة طرق فنية للتوقعات وهي: الاختلافات الأولية لوظيفة التحويل، شبكات الخلايا العصبية الاصطناعية والتحليل إلى مركبات أساسية تعتبر هذه الدراسة وسيلة لإبراز قدرة ونجاعة وفعالية كل طريقة ليتسنى اختيار الطريقة التي يكون لها الأثر الحسن في التحكم في الخطر المتوقع

Abstract :

In this work we tried to find effective for rainoff forecasting in particular those generated by sreams, problem whose scientist are currently tring to find a solution.

The study that we propose is based on the application of three methods of forecasting, the first diffrence of the transfert fonction (DPFT), the artificial neural network and the principal components analysis (PCA).

This studie will be a mean of highlighting the capacity and the reliability of each method and which of them will have a better impact on risk management.

Mots clés

- Analyse en composantes principales (ACP).
- Bassin versant.
- Crues.
- Différences premières de la fonction de transfert (DPFT).
- Modélisation pluie-débit.
- Prévision des débits.
- Réseaux de neurones.

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

CHAPITRE 1

**Revue bibliographique des modèles de
prévision des débits**

Introduction

Dans le cadre de cette revue bibliographique un recensement des modèles de prévision issus de plusieurs théories sera effectué, mais il est nécessaire de faire la distinction entre la prévision et la prédiction. La prévision concerne l'estimation des conditions futures à une date donnée ou durant un intervalle de temps spécifique, alors que la prédiction représente l'estimation des conditions futures, sans référence à un temps spécifique (LETTENMAIER et WOOD 1993). Par exemple on fait une prévision du débit de la rivière pour le lendemain à 17h00 mais on fait une prédiction du débit centenaire (période de retour de 100 ans) sur le même site. La prévision des débits horaires, journaliers, hebdomadaires et mensuels d'un cours d'eau est une tâche nécessaire pour une gestion efficace des ressources en eaux de surface, et pour permettre une réaction adéquate et rapide dans les situations de crise, comme les inondations. Il existe plusieurs approches pour effectuer ces prévisions. Certains modèles seront présentés avec leurs avantages et leurs inconvénients en ce qui concerne les problèmes liés à la prévision à court terme, moyen terme et long termes. Parmi les modèles empiriques, les modèles stochastiques paramétriques ou non paramétriques, les modèles à base de réseaux de neurones artificiels, les méthodes de régression floue. Pour ce qui est des modèles conceptuels on retiendra les modèles physiques, les modèles à base de règles floues.

1.1 La prévision des débits : Formulation du problème

Par prévision des débits, on entend le fait d'estimer le débit Q_t à un temps futur τ en n'utilisant que l'information disponible au temps $t < \tau$, correspondant normalement au temps présent (WERY, 1990). Cette information comprend les valeurs mesurées du débit lui-même, depuis un temps précédent $t - \tau$ jusqu'au temps t , ainsi que d'autres séries temporelles X_1, X_2, \dots, X_K . Il s'agit généralement d'informations météorologiques, par exemple les précipitations et la température. On note $x_t = (x_{1,t}, x_{2,t}, \dots, x_{K,t})'$ la transposée du vecteur des valeurs prises par chacune de ces variables au temps.

1.1.1 Utilisation de prévisions météorologique

Une prévision du débit doit généralement tenir compte non seulement de l'information hydrologique et météorologique passée, mais aussi des prévisions météorologiques (AFFANE et OUGLAOUENE, 1997). Formellement, il s'agit d'information incertaine sur la valeur des variables X_1, X_2, \dots, X_K à des temps postérieurs à l'instant présent. Les prévisions météorologiques doivent être traitées comme des variables explicatives distinctes, puisque leurs caractéristiques statistiques sont fort différentes de celles des observations météorologiques.

1.1.2 Calibration d'un modèle de prévision :

Pour obtenir une prévision des débits à partir des observations passées et des prévisions météorologiques, on suppose l'existence d'un modèle f reliant ces observations et ces prévisions des débits Q au temps τ :

$$\hat{Q}_\tau = f(Q_t, Q_{t-1}, \dots, Q_{t-r}, \hat{x}_{t-1}, \hat{x}_{t-2}, \dots, x_t, x_{t-1}, \dots, x_{t-r}, \hat{\varepsilon}_\tau | \hat{\theta}) \quad (1.1)$$

$\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m)'$ et $\hat{\varepsilon}_\tau$ sont respectivement les estimateurs de θ et ε_τ

Un modèle de prévision doit être ajuster à l'aide d'observations passées. De cette période de calibration, le modèle doit extraire l'information reliée au comportement du système modélisé, en éliminant le bruit. Si la souplesse du modèle de prévision est trop grande, le bruit risque d'être modélisé comme s'il s'agissait d'un comportement prévisible : sur la période de calibration le modèle pourrait alors donner un très bon ajustement, puisqu'il reproduirait même les variations aléatoires du débit. Par contre les prévisions effectuées par un tel modèle seraient alors, en comparaison, particulièrement mauvaises. On doit donc en vertu du principe de parcimonie chercher à minimiser le nombre de paramètres d'un modèle de prévision, tout en permettant qu'il soit assez souple pour modéliser convenablement le système (MASSON, 1971).

Une façon de vérifier si la modélisation permet d'extraire convenablement l'information disponible consiste à observer l'auto-correlation des écarts entre les débits observés et modélisés. Une forte auto-correlation indique que ce que l'on a

considéré comme bruit aléatoire comporte une information pouvant être incorporée dans le modèle (CEMAGREF, 1989), (JATON, 1982).

1.1.3 Validation d'un modèle de prévision

Une étape essentielle de la construction d'un modèle de prévision consiste à évaluer la précision des prévisions. Il faut réserver une portion des données disponible pour cette phase de validation. Ces données ne seront pas utilisées lors de la phase de calibration. Lors de la validation, on peut calculer différentes statistiques des prévisions, par exemple leur variance et leur biais. La période de validation doit être assez grande pour ces statistiques soient suffisamment précises pour être utiles. Ainsi, il importe de déterminer les distributions d'échantillonnage des statistiques utilisées comme critères de qualité d'un modèle de prévision.

Pour construire un modèle de prévision robuste, il est possible d'optimiser le nombre de paramètres en observant l'évolution de ces critères de qualité, toujours calculés sur une période de validation, en fonction de la complexité du modèle (BASTIN, 1984), (RODRIGUEZ, 1989). On peut alors déterminer à partir de quel moment l'augmentation du nombre de paramètres entraîne une détérioration des critères de validation au lieu d'une amélioration des prévisions.

1.2 Modèles empiriques

La caractéristique principale d'un modèle empirique est qu'on ne demande pas à sa structure de ressembler à celle du système modélisé, dans la mesure où le modèle permet de bien reproduire le comportement du système. Un modèle empirique peut être une boîte noire : dans la mesure où il reproduit des prévisions précises, il importe peu de comprendre sa structure interne. Cependant, dans les cas les plus simples, il peut être possible d'interpréter de façon physique les paramètres et les fonctions de transfert d'un modèle empirique. On distingue dans cette catégorie quatre grands types de modèles : Les modèles stochastiques, les modèles à base de réseaux de neurones artificiels, les modèles à base de régression floue et les modèles de systèmes chaotiques.

1.2.1 Modèles stochastiques

La plupart des modèles utilisés de nos jours sont des modèles stochastiques. Selon FELLER (1968) : il s'agit de tous les modèles utilisant la théorie des probabilités pour représenter l'incertitude.

a- Régression linéaire multiple

La régression linéaire multiple est un outil de modélisation très puissant mais parfois mal utilisé. En effet, le modèle de régression linéaire repose sur un ensemble d'hypothèses qui ne sont pas toujours vérifiées en pratique. Le modèle peut s'écrire ainsi pour la prévision des débits :

$$Q_t = \sum_{i=r}^t (\varphi_{t-i} Q_{t-i} + \lambda'_{t-r} x_{t-r}) + \varepsilon_t \quad (1.2)$$

φ_{t*} : coefficient d'ajustement associé à Q

λ'_{t*} : vecteur de coefficients d'ajustement associé au vecteur de variables x_{t*}

ε_t : bruit normal.

La régression est un outil relativement simple d'utilisation pour la prévision à moyen terme (débits mensuels). Cependant, les hypothèses du modèle de régression sont trop restrictives pour la prévision à court terme (pas de temps horaire). Pour les débits hebdomadaires, un modèle de régression risque d'être très limité dans son pouvoir prévisionnel en raison de son grand nombre de paramètres.

b- Méthode de Box-Jenkins

La famille des modèles de Box-Jenkins, dans sa formulation la plus générale, permet de présenter un vaste ensemble de séries temporelles. Il s'agit d'un modèle linéaire, tout comme le modèle de régression présenté précédemment (BOX et JENKINS, 1970). Une hypothèse souvent plus contraignante que la linéarité est la stationnarité : Les modèles Box-Jenkins ne permettent que de modéliser des séries dont les propriétés statistiques n'évoluent pas dans le temps. Cependant, il est en général possible d'effectuer une transformation sur les séries de données pour les rendre stationnaires. Par exemple la transformation $z_t = Q_t - Q_{t-1}$, qui correspond à la première différence du débit est souvent satisfaisante. Lorsque cette première différence n'est pas stationnaire, on peut considérer la deuxième différence, notée :

stationnaire, on peut considérer la deuxième différence, notée :

$$z_t = (Q_t - Q_{t-1}) - (Q_{t-1} - Q_{t-2}) = Q_t - 2Q_{t-1} + Q_{t-2} \quad (1.3)$$

Un modèle Box-Jenkins comprend deux parties : une composante autoregressive (AR) et une composante à moyenne mobile (MA).

Modèle autorégressif (AR)

Un modèle autorégressif linéaire d'une transformation z_t du débit Q_t , relie la valeur de z_t au temps t à ses valeurs précédentes de la façon suivante :

$$z_t = \gamma + \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \dots + \varphi_p z_{t-p} + \varepsilon_t = \gamma + \sum_{i=1}^p \varphi_i z_{t-i} + \varepsilon_t \quad (1.4)$$

Où γ, φ_i ($i = 1, 2, \dots, p$), étant des coefficients de régression, ε_t est un résidu normal $N(0, \sigma)$. Un modèle à p termes (en plus de ε_t et γ), est dénoté $AR(p)$.

La valeur de p , que l'on nomme l'ordre du modèle, doit être fixée a priori, alors que les autres paramètres sont estimés par calibration. En général il est nécessaire d'essayer plusieurs valeurs de p et observer la qualité des ajustements ainsi obtenus avant de faire un choix final. Il existe plusieurs méthodes permettant de choisir une valeur de p donnant un ajustement adéquat, pour cela il est recommandé d'utiliser les tests statistiques (BOX et al., 1976).

Modèle à moyenne mobile (MA)

Un modèle à moyenne mobile d'une transformation z_t du débit Q_t , relie la valeur z_t au pas de temps t aux erreurs ε_t précédentes de la façon :

$$z_t = \gamma + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t = \gamma \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (1.5)$$

Où, θ_i ($i = 1, 2, \dots, q$) est un coefficient de régression, et ε_t le résidu normal $N(0, \sigma)$.

Un modèle comportant q termes, est dénoté $MA(q)$. La valeur de q , que l'on nomme l'ordre du modèle, doit être fixée a priori, alors que les autres paramètres sont estimés par calibration. Comme dans le cas d'un modèle AR, différentes valeurs de q peuvent être essayées avant d'effectuer un choix final.

Modèles ARMA, ARIMA et PARMA

Un modèle $ARIMA(p, q)$ est une généralisation du modèle $AR(p)$ et $MA(q)$. il peut être représenté par l'équation suivante :

$$\begin{aligned} z_t &= \gamma + \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \dots + \varphi_p z_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \\ &= \gamma + \sum_{i=1}^p \varphi_i z_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \end{aligned} \quad (1.6)$$

Pour une écriture moins condensée, des opérateurs ont été utilisés : l'opérateur B de décalage arrière (backshift) : $B^n(z_t) = z_{t-n}$

Un autre opérateur est également utilisé, il s'agit de l'opérateur de différence $\nabla^n = (1 - B)^n$.

Pour $n = 2$ on obtient $\nabla^2(z_t) = (1 - 2B + B^2)(z_t) = z_t - 2z_{t-1} + z_{t-2}$, qui correspond à la seconde différence de z_t . La $n^{\text{ième}}$ différence de Q_t s'écrit alors $z_t = \nabla^n Q_t$ et le modèle ARIMA peut être résumé ainsi :

$$\varphi_p(B)z_t = \nabla^n Q_t = \gamma + \theta_q(B)\varepsilon_t \quad (1.7)$$

$$\text{Où } \varphi_p(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \text{ et } \theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (1.8)$$

Malgré la souplesse des modèles ARIMA, ces derniers n'utilisent comme entrées que les valeurs précédentes de la variable z_t et des erreurs ε_t . Il est souvent intéressant d'introduire d'autres variables explicatives comme la température et les précipitations, cette introduction de nouvelles variables est assurée par une fonction appelée *fonction de transfert*.

Fonctions de transfert de type ARMAX :

L'introduction d'une variable explicative dans un modèle de Box et Jenkins, une fonction de transfert est utilisée. Le modèle ARMAX est une bonne représentation (ARMAX pour ARIMA avec entrées auxiliaires). Une mesure x_t est introduite par exemple des précipitations au temps t dans le modèle ARIMA de la transformation z_t du débit Q_t . La première étape lors de la construction de ce type de modèle consiste à rendre x_t

stationnaire par une transformation : $z_t^{(x)} = x_t - x_{t-1}$. Ensuite, on ajoute au modèle *ARMA* une composante additive qui est fonction de $z_t^{(x)}$:

$$z_t = \gamma + \sum_{i=1}^p \varphi_i z_{t-i} + \sum_{i=1}^q \theta_i \eta_{t-i} + \eta_t + C \left[z_{t-b}^{(x)} + \sum_{i=1}^s w_i z_{t-b-i}^{(x)} \right] \quad (1.9)$$

Où $\gamma, \theta_i, \varphi_i, w_i$ et C sont des coefficients à estimer, et η_t est bruit. L'ordre d'un tel modèle est donné par p, q et b . le paramètre s représente le nombre de termes précédents de $z_t^{(x)}$ que l'on doit conserver pour calculer z_t . Le paramètre b indique le délai (en unités de temps) que prend une valeur $z_t^{(x)}$ pour affecter z_t . De façon générale, le bruit η_t n'a pas une distribution normale $N(0, \sigma)$. Il est parfois nécessaire de le modéliser lui-même à l'aide d'un modèle *ARMA* :

$$\eta_t = \gamma + \sum_{i=1}^{p_n} \varphi_{\eta,i} \eta_{t-i} + \sum_{i=1}^{q_n} \theta_{\eta,i} \varepsilon_{t-i} + \varepsilon_t \quad (1.10)$$

Par conséquent, l'introduction de variables explicatives dans le modèle suit le même procédé. Cependant chaque ajout de variables engendre un modèle lourd ; avec un nombre important de paramètres.

c- Méthode des K plus proches voisins

Cette méthode est une approche non paramétrique intuitive, mais qui repose sur de solides bases dans le domaine de la prévision des séries temporelles. Comme les modèles *ARMA*, elle s'applique aux séries stationnaires. Pour pouvoir estimer la sortie z reliée à un ensemble d'entrées $y_t = (z_t, z_{t-1}, \dots, z_{t-r}, x_t, x_{t-1}, \dots, x_{t-r})'$, on détermine dans l'historique les k voisins les plus proches de la situation actuelle, selon une distance pondérée. L'ensemble des observations de la sortie du système correspondant à ces k ensembles de paramètres d'entrées voisins de la situation actuelle forment alors une distribution prévisionnelle de z .

Si on prend l'exemple de la prévision d'un débit hebdomadaire, on pourrait chercher dans l'historique des situations présentant une tendance semblable du débit, des prévisions météorologiques semblables, un stock de neige (pour l'hiver) ou des conditions d'humidité du sol semblable (pour l'été). On pourrait alors prédire le débit à partir de ce qui a été observé des conditions similaires. Il faut noter que cette méthode

nécessite un temps de calibration du modèle considérable si l'historique est de taille conséquente.

Dans cette approche, la valeur de k ainsi que le poids associé à chaque entrée du vecteur y_t , sont des coefficients qui doivent être ajustés. Il s'agit donc d'une méthode essentiellement basée sur les données, puisque pour chaque prévision on doit chercher dans tout l'historique. Cette méthode. La méthode des k plus proche voisins a été appliquée pour la prévision des débits journaliers en utilisant les données des précipitations, par Karlsson et Yakowitz en 1987. D'autres auteurs démontrent que la méthode donne des résultats comparables à ceux issus d'un modèle *ARMAX* pour la prévision des débits

1.3 Modèle réseaux de neurones

Les réseaux de neurones artificiels ont été mis au point par McCulloch et Pitts en 1943 se basant sur le fonctionnement du cerveau humain (JODOUIN, 1994). De ce fait, ce type de modèles relevant du domaine de l'intelligence artificielle a acquis une notoriété particulière malgré qu'il s'agit d'un modèle empirique.

La forme générale du modèle réseau de neurone est sous la forme :

$$g(u_1, u_2, \dots, u_m) = \delta\left(\sum_{j=1}^m w_j u_j\right) \quad (1.11)$$

Où w_j est le poids accordé à l'entrée u_j et δ est une fonction quelconque, nommée fonction d'activation. En pratique, on retrouve trois types de fonction d'activation différente : la fonction identité $\delta(z) = z$, la fonction binaire $\delta(z) = 0$ si $z < h$, $\delta(z) = 1$ si $z > h$ (h étant le seuil d'activation), et la fonction sigmoïde

$$\delta(z) = \{1 + \exp(-z + h)\}^{-1} \quad (1.12)$$

Il existe plusieurs structures de réseaux de neurones mais la plus utilisée et celle du perceptron multicouche, ce dernier comporte une couche de neurones pour capter les entrées, une ou plusieurs couches intermédiaires, ainsi qu'une couche de neurones artificiels pour émettre les sorties liées au modèle.

Pour augmenter la fiabilité d'un modèle réseau de neurone il faut choisir la bonne configuration. Le choix de la configuration a une influence sur la qualité de la prévision, en général il est préférable d'utiliser une seule couche cachée et un nombre de

neurones dans cette couche du même ordre de grandeur que le nombre de neurones dans la couche d'entrée (DAVALO et NAÏM, 1993). Ceci vient du fait qu'il a été prouvé que les réseaux de neurones avec une telle configuration représente des approximateurs universels.

1.4 Régression linéaire floue

La régression floue est un outil récent issu des théories des sous-ensembles. Proposée par TANAKA en 1983, la régression floue a été étudiée plus en par BARDOSSY en 1990. C'est un outil efficace lorsque la régression statistique ne peut s'appliquer soit à cause de la non vérification des hypothèses du modèle de régression statistique soit le nombre d'observation est insuffisant pour permettre de valider les hypothèses du modèle. La régression floue repose sur la théorie du nombre flou, c'est un concept qui permet de représenter l'incertitude sur une quantité réelle X . il s'agit d'une fonction $\mu: R \rightarrow [0,1]$ qui associe à chaque nombre réel x un degré d'appartenance $\mu(x)$ compris entre 0 et 1, avec la contrainte que $\mu(x) = 1$ pour au moins une valeur de x .

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

$$Y = A_0 + \sum_{k=1}^K A_k (x_k - \bar{x}_k) \quad (1. 13)$$

Où $A_k, k = 1, 2, \dots, K$ sont des nombres flous et $\bar{x} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k\}$ est un point de l'espace choisi, pour lequel on estime que le modèle est plus précis. 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 celles de la régression linéaire classique :

- La corrélation entre les observations d'une même variable explicative x_k n'influence pas la validité du modèle.
- La variance n'a pas besoin d'être homogène ; une structure particulière d'erreur peut être modélisée explicitement.

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 grand mais avantageuse quand ces derniers sont petits.

Malgré ses avantages la méthode ses fondements de base ont été remis en question par plusieurs scientifiques.

1.5 Les modèles conceptuels

Contrairement aux modèles empiriques, les modèles conceptuels reflètent, de par leur structure interne, une conception du système modélisé. En pratique cette structure peut être vue comme un ensemble de contraintes qui limite le champ d'investigation lors de la calibration du modèle, Alors un nombre réduit sera utilisé pour avoir des prévisions précises.

1.5.1 Modèles physiques

Ce type de modèle est basé sur une ou plusieurs lois physiques fondamentales : conservation de l'énergie, de l'entropie, de la masse, et la loi de la gravitation. Leur application est réalisée en divisant le bassin en unités élémentaire, sur lesquels les équations décrivant ces lois physiques sont assez simple et peuvent être résolues sans difficulté apparente. Les plus utilisées sont : la méthode des différence finies et celle des éléments finis. Ces méthodes sont utilisées massivement dans l'hydrogéologie. en modélisation hydraulique et encore assez peu en hydrologie de surface.

Pour la prévision l'utilisation d'un modèle physique peut s'avérer utile parce que le débit dépend généralement, pour un bassin versant de taille moyenne, des conditions existant en amont et beaucoup moins des précipitations qui pourraient survenir (Lettenmaier et Wood, 1993) où ils présentent aussi une revue des méthodes de prévisions des débits qui sont basées sur la modélisation de la transformation précipitations débits, et des méthodes basées sur le laminage des volumes d'eau (onde dynamique, équation de diffusion, réponse impulsionnelle, Muskingum...).

La limite des approches physiques, entre autre de leur complexité, est le nombre de paramètres que font introduire. En prenant les unités élémentaires il faut considérer

l'ensemble de tous les paramètres qui doivent être mesurés ou estimés. Ajouter à cela la complexité des processus hydrologiques au sein du bassin versant, un ensemble d'équations ne pourra pas décrire ces processus d'une manière fiable. Même pour les petits bassins très homogènes, une modélisation physique, malgré souhaitée, reste un objectif lointain.

1.6 Modèles à base de règles floues

Les modèles à base de règles floues sont couramment utilisés dans le domaine du contrôle, mais leurs applications dans d'autres domaines furent récentes (BARDOSSY, 1990). Un tel modèle est un ensemble de règles SI-ALORS qui relient les entrées du modèle à ses sorties.

Deux exemples de règles pour la prévision des débits :

SI il y a une pluie *importante* en un temps *court*.

ALORS il y aura une *forte* augmentation du débit.

SI il y a une pluie *faible* en un temps *long*.

ALORS il y aura une *faible* augmentation du débit.

Les termes, court, fort, faible et long sont des quantificateurs imprécis, qui montrent que chaque règle est incertaine.

Les modèles à base de règles floues sont très utiles lorsque l'on dispose de connaissances générales sur les processus physiques du système à modéliser, sans que ces connaissances soient sous la forme d'équations précises. Ce type de modèle permet de donner un ordre de grandeur pour la prévision, mais peut aussi conduire à un modèle plus précis, puisqu'il a été montré que les modèles à base de règles floues forment aussi une classe d'approximateurs universels, tout comme les réseaux de neurones (KOSKO, 1992). Cette approche a été utilisée pour quantifier la prévision de la demande en eau en tenant compte des prévisions météorologiques, et pour prévoir la température moyenne journalière (Bardossy, 1990)

1.7 Méthodes d'estimation de la distribution régionale des débits :

1.7.1 Modèles débit-durée-fréquence :

Ces modèles notés QdF, permettent la reconstitution du régime des crues à partir d'une technique de normalisation qui présente des analogies avec la méthode de l'indice de crue (DALRYMPLE, 1960) (GALEA et PRUDHOMME, 1997). La procédure d'estimation est la suivante :

- a- Estimation de trois paramètres locaux, représentatifs du bassin versant : la durée caractéristique de crue D , le débit décennal de pointe noté $QIXA10=Q(T=10 \text{ ans})$ et le gradex des pluies $a_p(d)$. les deux premiers paramètres, respectivement des descripteurs de la fonction de transfert et de production des crues, sont estimés directement à partir de mesure hydrologiques lorsqu'elles existent, ou autrement à l'aide de régressions linéaires établies à l'échelle nationale ou régionale et utilisant des variables explicatives facilement estimables comme la superficie du bassin versant et des valeurs caractéristiques de pluies (CEMAGREF, 1989).
- b- Choix d'un modèle de référence QdF adimensionnel. Pour différentes valeurs de durées d , le rapport $a_p(d)/QIXA10$, obtenu en première étape, est comparé à deux limites $L1(d/D)$ et $L2(d/D)$. Suivant la position de ce rapport, nous choisissons respectivement un des trois modèles de référence suivant :

$$T \leq 10 \text{ ans} : \frac{Q(T, d)}{QIXA10} = A(d/D) + B(d/D) \cdot \text{Log } T \quad (1. 14)$$

$$T > 10 \text{ ans} : \frac{Q(T, d)}{QIXA10} = \frac{Q(10, d)}{QIXA10} + C(d/D) \cdot \text{Log} \left[1 + \frac{B(d/D)}{C(d/D)} \cdot \frac{T-10}{10} \right] \quad (1. 15)$$

- c- Reconstitution du régime des crues par dénormalisation du modèle de référence. Le quantile local de crue est alors estimé en dénormalisant la distribution adimensionnelle du modèle de référence (OUARDA et LANG, 1999) :

$$Q(T, d)_{(local)} = \left(\frac{Q(T, d)}{QIXA10} \right)_{(mod\ de\ référence)} \cdot (QIXA10)_{(local)} \quad (1. 16)$$

Cette régionalisation comporte plusieurs aspects originaux :

- L'objectif de la méthode est de reconstituer la distribution des crues, pas seulement pour le débit de pointe, mais aussi pour les valeurs des débits calculés sur une durée d quelconque. La variable Q du débit étudié est relative soit à un débit moyen maximum sur la durée d, débit moyen VCXd(T), soit) un débit dépassé de façon continue pendant la durée d, débit seuil maximum QCXd(T), (OUARDA et LANG, 1999).
- La procédure de régionalisation intègre une information pluviométrique, par le gradex des pluies. Cette information est exploitée à deux niveaux : dans le choix du modèle de référence, par le critère $ap(d)/QIXA10$

Ajoutons aussi la possibilité de générer, à partir des courbes QdF en débit seuil QCXd, des hydrogrammes synthétiques monofréquences, fonctionnellement analogue à des crues de projet et relatifs à une période de retour.

Une procédure similaire de régionalisation par les modèles QdF a également été appliquée aux régimes d'étiages (GALEA et al., 1998), à partir des deux descripteur suivants : le débit moyen minimal VCNT(T = 2 ans, d = 1 jour) et la durée caractéristique d'étiage issue des courbe de tarissement.

1.8 Méthode de la cartographie de l'écoulement :

Un autre aspect de la régionalisation concerne la restitution d'informations locales sous forme cartographique. L'objectif est davantage d'appréhender la variabilité spatiale des écoulements, que d'estimer ceux-ci en tout points. Une méthode classique, pour l'étude de la ressource en eau, consiste à cartographier le débit moyen annuel ou mensuel sous forme de lame d'eau écoulée. Elle présente cependant l'inconvénient de faire l'hypothèse implicite de l'homogénéité spatiale des processus d'écoulement. une fois la caractéristique de débit normée par la superficie du bassin versant.

La cartographie est réalisée avec une technique de krigeage utilisant la distance d (OUARDA et LANG, 1999) :

$$d^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + \lambda \cdot (\text{Log}S_2 - \text{Log}S_1)^2 \quad (1. 17)$$

$$\text{Où } \lambda = \text{Max}_{i,j} [(x_j - x_i)^2 + (y_j - y_i)^2] / \text{Max}_{i,j} [(\text{Log}S_j - \text{Log}S_i)^2] \quad (1. 18)$$

Cette technique de cartographie permet par exemple de visualiser le cycle annuel des écoulement et de mettre en évidence, pour chaque période de l'année, les zones présentant la plus forte production.

1.9 Utilisation de l'imagerie satellite :

Grâce à sa vision synoptique et répétitive, la télédétection satellitale permet l'extrapolation spatiale et temporelle d'indicateurs relatifs aux états de surface.

Ces approches de type agrégatif (de l'unité du m² jusqu'à la taille du bassin versant) butent le plus souvent sur la difficulté de prendre en compte les mécanismes hydrologiques à des niveaux d'organisation qui sont supérieur à celui de la cellule élémentaire définie par le modèle (quelques m²). Ces démarches conduisent alors à établir des fonctions de transition destinées à corriger ces sauts d'échelle. Une des voies possibles pour permettre le passage de l'échelle locale à l'échelle globale est de rechercher un niveau de description hydrodynamique des états de surface intermédiaire entre l'unité élémentaire et le bassin versant, autrement dit de réduire le saut de l'échelle. VINE et al. (1997) proposent une approche désagrégative, en utilisant la télédétection satellitale comme outil de segmentation de l'espace, à partir du découpage du bassin versant en zones homogènes, suivant leur état de surface et le taux de végétation. Un coefficient d'écoulement $(Ce)_j$ est affecté à chaque zone homogène n° j. Pour un bassin versant n° i, de superficie S_i , constitué de sous-bassins homogènes de superficies S_{ij} , le modèle hydrologique transforme la pluie p_i en débit q_i , en additionnant les écoulement de chaque zone homogène :

$$q_i = \sum_j (Ce)_j \cdot (S_{ij} / S_i) \cdot p_i \quad (1. 19)$$

$$\text{avec } : (Ce)_j = (Ce_{ref})_j \cdot \left\{ S_{ij} / (S_{ref})_j \right\}^{n-1} \quad (1. 20)$$

Le coefficient d'écoulement Ce est modulé en fonction de la superficie, à l'aide de l'exposant $(n-1)$, pour tenir compte du laminage et de l'infiltration lors du transfert amont-aval. Après une phase de calibrage des paramètres du modèle, la procédure permet de reconstituer les écoulements à partir des mesures de pluie et de l'information spatiale issue de la télédétection. Cette technique permet de progresser

dans le domaine de la régionalisation hydrologique, en exploitant la connaissance des états de surface obtenus par imagerie satellitale.

Conclusion

Les modèles cités précédemment, décrivent la modélisation pluie-débit au pas de temps journalier. Au pas de temps horaire, l'approche devient plus complexe du fait de la non-continuité des données, et de la difficulté de simuler certains épisodes, ajoutant à cela le problème de certaines crues complexes.

C'est dans ce contexte, que nous avons entamé cette modélisation au pas de temps horaire, en prenant en considération trois types de modèles, le modèle DPFT, les réseaux de neurones, et une nouvelle méthode de modélisation basée sur l'analyse en composante principale.

CHAPITRE 2

**Etude de la méthode de la différence
première de la fonction de transfert
(DPFT)**

2.1 Introduction

La description de la relation pluie-débit à l'aide d'un modèle capable de reproduire le phénomène naturel d'une manière rigoureuse s'avère nécessaire pour effectuer une prévision sur les débits, c'est dans cette optique que nous proposons une méthode de prévision qui s'inspire de la méthode classique de l'hydrogramme unitaire proposé par Sherman en 1932 (JATON, 1982).

Le modèle étudié comme celui de l'hydrogramme unitaire décompose la relation pluie débit en deux parties ou sous-modèle :

- Un sous-modèle de production caractérisé par une fonction de production non linéaire, cette dernière permet la transformation des pluies brutes tombant sur un bassin versant en pluies efficaces qui, par la suite se transformera en débit de ruissellement qui atteindra l'exutoire.
- Une fonction de transfert caractérisée par sa linéarité, assurant la transformation des pluies efficaces en débit.

La méthode dont l'étude sera faite Est connue sous le nom de **DPFT** : *Différence Première de la Fonction de Transfert*.

La méthode est basée sur les études réalisées par *NEWTON et VINYARD* en 1967, leurs travaux se sont soldés par l'élaboration d'un algorithme itératif et alterné qui permet une identification directe d'une fonction de transfert moyenne ainsi que la détermination des pluies efficaces.

On définit la *DPFT*, comme la fonction dont les coefficients sont les différences premières des coefficients de la fonction de transfert, les coefficients de la *DPFT* peuvent être de nombre infini, c'est pour cette raison qu'on va se limiter à un nombre bien défini *K*, on dira alors que la *DPFT* est tronquée à une longueur *K*. Cette longueur est appelée *Mémoire du Système*.

2.2 Les principes de base de la méthode

2.2.1 Principe de base n° 1 : Formulation en différence première Comme son nom l'indique la *DPFT* est exprimée sous forme de différences de débits sur un pas de temps reliant le présent et le futur (NALBANTIS, 1987).

L'idée de travailler en différences premières de débits à deux origines :

1) *Opérationnelle* : Il est souvent nécessaire, et plus significatif de travailler en variation de débit entre le pas de temps futur et celui observé.

2) *Théorique* : Pour un processus stochastique avec des caractéristiques statistiques variable dans le temps. La dérivée d'ordre 1 et 2 fournissait un processus beaucoup plus régulier (Box et Jenkins, 1970), sur lequel on pouvait travailler pour revenir ensuite au processus initial.

Sur un intervalle de temps $j-1, j$ la relation pluie-débit est donnée par :

$$Q_{dj} = A_1 PE_j + A_2 PE_{j-1} + \dots + A_{k-1} PE_{j-k+2} \quad (2.1)$$

$$Q_{dj-1} = A_1 PE_{j-1} + A_2 PE_{j-2} + \dots + A_{k-1} PE_{j-k-1} \quad (2.2)$$

Où :

Q_{dj} : Débit de ruissellement direct au pas de temps j ;

Q_{dj-1} : Débit de ruissellement direct au pas de temps $j-1$;

PE_j : Pluie efficace exprimée en terme de volume d'eau cumulé sur l'intervalle de temps $(j-1, j)$;

$A_1, A_2, A_3, \dots, A_{k-1}$: Coefficients de la fonction de transfert.

Le temps est mesuré en nombre de pas de temps $j = 1, 2, 3, \dots$

La pluie et le débit sont exprimés dans la même unité (m³/h). Les pluies sont multipliées par le coefficient : $R = 10 S/36$, S étant la surface du bassin versant en km².

Les coefficients de la fonction de transfert sont alors adimensionnels.

Sachant que le débit total est la somme du débit de base, et du débit de ruissellement direct, on peut écrire :

$$Q_j = Q_{dj} + Q_{bj} \quad (2.3)$$

Où : Q_j : Débit total au pas j ;

Q_{bj} : Débit de base au pas j ;

Q_{dj} : débit de ruissellement direct au pas j .

On considère maintenant les différences premières du débit total, appelées aussi variations de débit entre le pas $J-1$ et J :

$$q_j = Q_j - Q_{j-1} = (Qd_j - Qd_{j-1}) + (Qb_j - Qb_{j-1}) \quad (2.4)$$

On substitue Q_{dj} et Q_{dj-1} par leurs expressions données en (3.1) et (3.2) on a :

$$q_j = Q_j - Q_{j-1} = A_1 PE_j + A_2 PE_{j-1} + \dots + A_{k-1} PE_{j-k+2} + 0 - 0 - A_1 PE_{j-1} - A_2 PE_{j-2} - \dots - A_{k-2} PE_{j-k+2} - A_{k-1} PE_{j-k+1} + Qb_j - Qb_{j-1} \quad (2.5)$$

Cette équation s'écrit aussi :

$$q_j = (A_1 - 0) PE_j + (A_2 - A_1) PE_{j-1} + (A_3 - A_2) PE_{j-2} + \dots + (A_{k-1} - A_{k-2}) PE_{j-k+2} + \dots + (0 - A_{k-1}) PE_{j-k+1} + Qb_j - Qb_{j-1} \quad (2.6)$$

Plus formellement on peut écrire :

$$a_i = A_i - A_{i-1} \quad i = 1, 2, 3, \dots, k \quad (2.7)$$

Avec : $a_1 = A_1$ et $A_k = 0$

a_i : Coefficients de la DPFT.

$$q_j = a_1 PE_j + a_2 PE_{j-1} + a_3 PE_{j-2} + \dots + a_k PE_{j-k+1} + Qb_j - Qb_{j-1} \quad (2.8)$$

avec : $a_1 = A_1$ et $a_k = -A_{k-1}$

Ou de manière condensée :

$$q_j = \sum_{i=1}^k a_i PE_{j-i+1} + (Qb_j - Qb_{j-1}) \quad (2.9)$$

En général, la variation du débit de base d'un pas de temps à l'autre est négligeable par rapport au débit de ruissellement (NALBANTIS, 1987). Donc l'équation (2.9) deviendra alors :

$$q_j = \sum_{i=1}^k a_i PE_{j-i+1} \quad (2.10)$$

Finalement, on a abouti à une équation de convolution discrétisée entre les pluies efficaces et la Différence Première de la Fonction de Transfert, cette formulation, comme il a été cité plus haut, est analogue à celle de hydrogramme unitaire, mais en faisant intervenir les différences la DPFT est plus stable.

2.2.2 Principe de base n°2 : les pluies efficaces sont considérées comme inconnues du problème.

On a souvent considéré que la seule inconnue du problème est la fonction de transfert (méthode de l'Hydrogramme Unitaire), car les pluies efficaces étaient susceptibles d'être approchées par des modélisations plus ou moins grossières.

La DPFT à l'avantage de déterminer en premier lieu la fonction, et en second lieu les pluies efficaces.

2.2.3 Principe de base n°3 : une solution alternative itérée

On a utilisé d'abord une approximation des pluies efficaces pour estimer la fonction de transfert, ensuite, on considère cette fonction de transfert approchée pour résoudre le problème inverse, et estimer le deuxième jeu d'inconnues qui représente les pluies efficaces, en espérant d'une part que le processus itératif converge, et d'autre part qu'il converge surtout vers un optimum.

Ce principe se résume dans les étapes suivantes :

- à l'itération (i), on identifie la DPFT (i) moyenne à l'aide des pluies efficaces déconvoluées ou estimées à l'itération (i-1)
- avec cette DPFT (i) moyenne on réestime les pluies efficaces PE (i) à l'aide du système (2.10) :

$$q_j = \sum_{i=1}^k a_i PE_{j-i+1}$$

L'initialisation de ce système se fait par l'approche suivante :

$$PE(0) = PB \quad (2.11)$$

Avec :

PB : Pluie brute ;

PE : Pluie efficace.

L'écriture des systèmes d'équations sera :

$$\{q\} = [a] \{PE\} \quad (2.12)$$

$$\{q\} = \begin{Bmatrix} q_1 \\ q_2 \\ \cdot \\ \cdot \\ q_n \end{Bmatrix}; \{PE\} = \begin{Bmatrix} PE_1 \\ PE_2 \\ \cdot \\ \cdot \\ PE_m \end{Bmatrix}; [a] = \begin{bmatrix} a_1 & 0 & 0 & \cdot & 0 \\ a_2 & a_1 & 0 & \cdot & 0 \\ a_3 & a_2 & a_1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_n & a_{n-1} & a_{n-2} & \cdot & a_{n-m+1} \end{bmatrix}$$

avec :

$$\text{et : } \{q\} = [PE] \{a\} \quad (2. 13)$$

$$\{a\} = \begin{Bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_n \end{Bmatrix}; [PE] = \begin{bmatrix} PE_1 & 0 & 0 & \cdot & 0 \\ PE_2 & PE_1 & 0 & \cdot & 0 \\ PE_M & PE_{M-1} & PE_{M-2} & \cdot & PE_{M-k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & PE_M \end{bmatrix}$$

Où : n : Nombre total de débit ;
 m : Nombre de pluies efficaces ;
 k : Mémoire du système.

2.3 Introduction des contraintes au niveau de l'algorithme de résolution

Dans le but d'assurer la fiabilité des résultats, ainsi que le bon fonctionnement de la méthode ,on introduit des conditions sur la DPFT et les pluies efficaces qui sont les inconnues du problème, certaines de ces conditions sont introduites directement dans le calcul, ce type de contrainte est appelé *contraintes directes*. Il y a un autre type de contraintes qui sont introduites entre les différentes étapes de calcul, il est appelé *contraintes indirectes*.

2.3.1 Contraintes sur la fonction de transfert

a- Contrainte de positivité

Durant la détermination des coefficients de la fonction de transfert certains coefficients peuvent être négatifs, ce qui n'a pas de signification en hydrologie, cette contrainte peut être expliquée par le fait que des coefficients négatifs vont engendrer une diminution du débit total en diminuant le débit de ruissellement donc, toute valeur négative ne sera pas acceptée, et elle sera prise égale zéro.

$$A_i > 0 \quad i = 1, 2, \dots \quad (2.14)$$

A_i : Valeur de la fonction de transfert.

b- Contrainte sur la forme de la fonction de transfert :

Si on trace la fonction de transfert on observe des oscillations au niveau de l'extrémité de la fonction (au niveau des dernières valeurs). Pour mettre fin à ces oscillations on introduit une contrainte sur la forme de la fonction de transfert en ajustant la fonction suivant une exponentielle décroissante, cette technique est connue sous le nom de *lissage* de la fonction de transfert (Rodriguez, 1989). Cette courbe exponentielle est de la forme :

$$A_j = A_L \text{EXP} (-\alpha) \quad j > L \quad (2.15)$$

Avec :

A_j : Coefficient de la fonction de transfert au pas j ;

A_L : Premier coefficient de la branche de la FT à lisser ;

α : Coefficient du lissage exponentiel.

c- Norme de la fonction de transfert : *conservation de la masse.*

Si la pluie et les débits sont exprimés dans le même système d'unité, la loi de conservation de la masse implique que la fonction de transfert, qui est la réponse à une pluie efficace unitaire, doit avoir une surface égale à l'unité (Nalbantis, 1987). Cela se traduit par :

$$\sum_{i=1}^k A_i = 1 \quad (2.16)$$

Cette contrainte peut être introduite de deux façons :

- Entre les itérations \Rightarrow *contrainte indirecte.*
- Au cour de l'itération \Rightarrow *contrainte directe.*

2.3.2 Contraintes sur les pluies efficaces

En ce qui concerne les pluies efficaces, on notera les contraintes suivantes :

a- Contrainte de positivité des pluies efficaces

Il est clair que la pluie efficace n'est qu'une fraction de la pluie brute exprimée en termes de volume et qui alimente le ruissellement direct, par conséquent, toute valeur négative de

celle-ci n'a aucun sens en hydrologie. Donc, les pluies efficaces seront bornées par la valeur zéro.

Cette contrainte peut se traduire par l'expression suivante :

$$PE \geq 0 \quad (2.17)$$

PE : Pluie efficace.

b- Contrainte sur le coefficient de ruissellement

A partir de la définition du coefficient de ruissellement et du fait que celui-ci doit être inférieur ou égal à un, sachant aussi que ce coefficient est donné par :

$$C_{rj}^L = PE_j^L / PB_j^L \leq 1 \quad (2.18)$$

C_{rj}^L : Coefficient de ruissellement ;

PE_j^L : Pluie efficace sur l'intervalle [j-1, j] pour l'épisode L ;

PB_j^L : Pluie brute sur l'intervalle [j-1, j] pour l'épisode L.

On peut déduire que : $PE_j^L \leq PB_j^L \quad (2.19)$

2.4 Description de la méthode de résolution

La résolution se fait par itération, dans ce qui suit, nous donnons une idée générale sur la résolution de l'algorithme comportant la détermination de la fonction de transfert ainsi que la déconvolution des pluies efficaces.

2.4.1 Détermination de la fonction de transfert

On résout en premier lieu le système (2. 13) permettant de déterminer les coefficients de la DPFT en initialisant l'algorithme par les pluies brutes étant donné que les pluies efficaces sont les inconnues du problème.

La fonction de transfert issue de cette résolution doit vérifier les différentes conditions annoncées plus haut.

$$\{q\} = [PE] (0) . \{a\} = PB . a \Rightarrow a^* \quad (2.20)$$

2.4.2 Déconvolution des pluies efficaces

Maintenant, il suffit de résoudre le système (3.12). En utilisant les coefficients de la DPFT trouvés lors de la résolution du premier système (3.13) et après résolution, nous aurons les pluies efficaces qui doivent vérifier les contraintes notées plus haut. :

$$\{q\} = [a] \cdot \{PE\} \Rightarrow PE^* \quad (2.21)$$

Cette opération - détermination de la fonction de transfert et la déconvolution des pluies efficaces - est répétée pour chaque itération mais il faut noter qu'à partir de la deuxième itération les pluies brutes sont remplacées par les pluies efficaces déterminées dans l'itération précédente.

Globalement, on doit estimer deux jeux d'inconnues, à savoir :

1/ les coefficients de la différence première de la fonction de transfert moyenne tronquée à k pas de temps, ce qui revient à supposer une fonction de transfert de mémoire $k-1$.

2/ $m_T = \sum_{L=1}^N m_L$, N étant le nombre de pluies efficaces pouvant être considérées comme des paramètres d'entrée.

Donc, pour une première approximation $\hat{a}_{(1)}$ des coefficients de la DPFT on a :

$$\{q\} = [Pe(0)] \{a\} \rightarrow \hat{a}_{(1)} \text{ avec } PE(0) = PB$$

Pour une estimation des pluies efficaces $PE_{(1)}$ et de la différence première de la fonction de transfert pour chaque épisode séparément, l'estimation du débit sera donnée par le système (2.12) :

$$\{q_1\} = [a_{(1)}] \{PE_1\}$$

On continue la résolution en effectuant des itérations où les mêmes étapes seront réitérées jusqu'à aboutir à des estimations des coefficients de la DPFT et des pluies efficaces stables et robustes.

2.5 Techniques numériques utilisées dans la DPFT

La technique d'identification de la fonction de transfert ainsi que de la déconvolution des pluies efficaces a subi plusieurs améliorations.

2.5.1 Identification de la fonction de transfert

En premier lieu, (Guillot et Duband 1980) proposèrent un système multi-événement : $\{q\} = [PE] \{a\}$ par une régression multilinéaire. Le système utilisé est :

$$q_i = \sum_{i=1}^k a_i PE_{j-i+1}^L + \varepsilon_i^L \quad (2.22)$$

Avec

ε_i^L : Erreur d'ajustement.

On introduisant le terme de régression, on aura :

$$q_i = \sum_{i=1}^k a_i PE_{j-i+1}^L + h_0 + \varepsilon_i^L \quad (2.23)$$

Où :

h_0 : Terme constant de régression.

Dans la version qui date de 1983, Versiani a estimé que l'introduction du terme de régression peut engendrer une erreur tout au long de la crue, il a proposé un estimateur par moindres carrées :

$$a^* = [(PE)^T PE]^{-1} (PE)^T q \quad (2.24)$$

A cause du mal conditionnement de la matrice à inverser $[(PE)^T PE]^{-1}$ Versiani en 1983 a utilisé la Ridge Régression et le nouvel estimateur de la fonction de transfert était :

$$a^* = [(PE)^T PE + \tau_k G_L]^{-1} (PE)^T q \quad (2.25)$$

τ_k : Coefficient de la Ridge Régression ;

G_L : Matrice diagonale formée par la diagonale de $(PE)^T PE$.

NALBANTIS en 1987, propose une identification par le modèle ARX, appliqué sur les débits mais il a suggéré le recours aux moindres carrés pour l'estimation de la DPFT.

2.5.2 Déconvolution des pluies efficaces

Dans le but d'avoir des pluies efficaces robustes et bien stables plusieurs techniques ont été élaborées par différents auteurs.

L'estimateur des pluies efficaces est donné par :

$$PE^L = [(a)^T a]^{-1} (a)^T q^L \quad (2.26)$$

Cet estimateur est influencé par le mauvais conditionnement de la matrice à inverser $[(a)^T a]^{-1}$. Dans sa version de (1983) et pour faire face à ce problème de mal conditionnement de la matrice inverse, Versiani a proposé un estimateur par la ridge régression dont l'expression est la suivante :

$$PE^L = [(a)^T a + \tau_K G_L]^{-1} (a)^T q^L \quad (2.27)$$

G_L : Matrice diagonale formée par la diagonale de $(a)^T a$;

τ_K : Coefficient de la ridge régression.

NALBANTIS (1987) a tenu compte de certaines contraintes sur les PE en utilisant la programmation linéaire , il a aussi travaillé par simulation mais il a opté pour la ridge régression proposée par Versiani pour la stabilité de ses résultats.

Rodriguez (1989), proposa une nouvelle technique en utilisant la programmation quadratique (algorithme de moindres carrés sous contraintes), son travail est basé sur l'optimisation du critère de l'erreur quadratique du système de déconvolution noté plus bas et l'application des contraintes sur les PE.

$$q^L = a * PE^L + \varepsilon^L \quad (2.28)$$

Où :

ε^L : erreur d'ajustement.

La contrainte $PE \geq 0$ peut s'écrire matriciellement sous la forme suivante :

$$0_m \leq I_m PE \quad (2.29)$$

0_m : vecteur nul de dimension " m "

I_m : matrice identité d'ordre " m "

m étant le nombre de pluies efficaces.

On peut aussi écrire : $PE \leq C_{conf} PB$ sous la forme matricielle :

$$C_{conf} PB \geq I_m PE \quad (2.30)$$

Cette technique peut diminuer les oscillations des PE lorsque les données sont entachées d'erreurs.

Rodriguez propose aussi de supprimer les pluies efficaces pour lesquelles les pluies brutes sont nulles, ce qui augmente le degré de liberté du système.

Finalement, le principe de la méthode se résume dans l'identification de la DPFT sur un jeu multi-événement. Pour se faire, il faut fixer la mémoire du système ou la taille de la fonction de transfert, utiliser la technique la plus simple possible pour la détermination des coefficients de la DPFT afin d'éviter les très lourds calculs (moindres carrées) et application des différentes contraintes citées auparavant.

En ce qui concerne la déconvolution des pluies efficaces les différentes étapes peuvent se résumer comme suit :

- Déconvolution directe des PE par le système (2. 12).
- Utilisation de la ridge régression.
- Application à chaque itération des contraintes propres aux pluies efficaces.

2.6 Conclusion

La DPFT est plus qu'une simple méthode, elle représente une approche basée sur les hypothèses de linéarité et d'invariance temporelle qui découle de la méthode de "l'hydrogramme unitaire" mais avec l'avantage de la détermination simultanée de la fonction de transfert et des pluies efficaces.

La méthode DPFT utilise les variations du débit total entre deux pas de temps ce qui assure une bonne stationnarité de la variable et diminue l'autocorrélation entre les coefficients de la DPFT comparés à ceux de la fonction de transfert, la formulation en différences permet aussi de négliger la variation du débit de base et par conséquent de s'affranchir de la séparation des hydrogrammes et de ne prendre que la réponse rapide.

L'algorithme de résolution procède par itération (figure 2-1), en résolvant alternativement un système multi-événement permettant l'identification de la fonction de transfert, en ayant cette estimation de la DPFT, on résout le système de déconvolution ou le système inverse qui estime une série de pluies efficaces, le démarrage du processus de résolution repose sur l'approche : $PE = PB$. Le processus est réitéré jusqu'à ce que les estimations de la DPFT et des pluies efficaces soient stables et robustes.

Le nombre d'inconnues à estimer étant assez élevé, l'introduction de plusieurs contraintes dans l'algorithme s'avère nécessaire afin de guider notre choix des valeurs à utiliser. Parmi celles-ci nous pouvons citer la contrainte de positivité de la fonction de transfert, la positivité des pluies efficaces et les contraintes imposées sur le coefficient de ruissellement.

La méthode DPFT ne nécessite donc que la mesure des " pluies brutes " et la mesure de " débit " pour la détermination des fonctions de transfert et des pluies efficaces.

CHAPITRE 3

Modélisation par les réseaux de neurones

3.1 Introduction :

L'identification des processus physiques dans le cycle hydrologique pouvant avoir lieu dans un bassin versant et l'attribution à chacun d'eux une description analytique permettant la prévision nous amène à des expressions complexes utilisant un grand nombre de paramètres rendant ainsi leur exploitation lourde et difficile.

Les processus hydrologiques existants au niveau d'un bassin versant sont généralement non linéaires, le recours à des modélisations simples et à faible nombre de variables capables de traiter la non-linéarité s'avère nécessaire.

C'est dans ce contexte que nous proposons une méthode de modélisation pluie débit basée sur l'utilisation des réseaux de neurones dont les performances dans le domaine de la modélisation non linéaire ont été déjà prouvées (DONNA RIZO et DOUGHERTY , 1994), ainsi que dans le domaine de la prévision des séries temporelles. En particulier, ils ont eu un succès retentissant pour dans le domaine des sciences de l'eau, on retrouve des applications à la prévision (FRENCH et al., 1992)

3.2 Les réseaux de neurones :

Relevant du domaine de l'intelligence artificielle, les réseaux de neurones sont une métaphore des structures cérébrales : Des assemblages de constituants élémentaires, qui réalisent chacun un traitement simple voire simpliste, mais dont l'ensemble fait émergé des propriétés globales dignes d'intérêt. Chaque constituant fonctionne indépendamment des autres, de telle sorte que l'ensemble est un système parallèle, fortement interconnecté. L'information détenue par le réseau de neurones est distribuée à travers l'ensemble des constituants. Enfin un réseau de neurone ne se programme pas pour réaliser telle ou telle tâche. Il est entraîné sur des données acquises, grâce à un mécanisme *d'apprentissage* qui agit sur les constituants du réseau afin de réaliser au mieux la tâche souhaitée.

Parmi les tâches particulièrement adaptées au traitement par réseau de neurones, on trouve l'*association* (association entre odeur et une expérience donnée), la *classification*, la *discrimination* et l'*estimation*. Toutes ces tâches sont très différentes de celles traitées par l'informatique traditionnelle. Tout l'art de la technique des réseaux de neurones consiste aujourd'hui à modéliser ces divers éléments ; Il faut ainsi proposer des modèles mathématiques du neurone réel et de la jonction synaptique, des schémas d'interconnexion des neurones, et enfin des règles permettant d'agir sur les différents paramètres de cette

modélisation. Cette démarche a été engagée en 1943 par McCulloch et Pitts qui ont proposé un premier modèle de neurones et par D. Hebb qui a énoncé une règle de modification des connexions synaptiques (JODOUIN, 1994)

3.3 Définition du neurone formel :

Le neurone de McCulloch et Pitts est un dispositif binaire, qui reçoit des stimulations par des entrées, et les pondérées grâce à des valeurs réelles appelées coefficients synaptiques, poids synaptiques ou simplement synapses. Ces coefficients peuvent être positifs, et l'on parle alors de synapses excitatrices ou négatifs pour des synapses inhibitrices. Le neurone calcule ainsi une somme de ses entrées pondérée par les coefficients, et prend une décision en comparant à un seuil fixe : si la somme pondérée des entrées dépasse le seuil, la sortie produite vaut +1, sinon la sortie vaut -1. Un schéma du neurone formel est proposé sur la figure 3.1

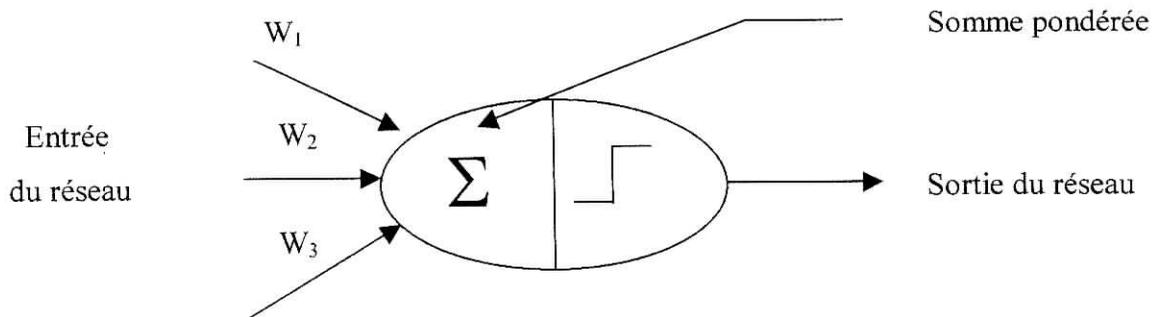


Figure 3.1 : Représentation schématique d'un réseau de neurones

Le neurone reçoit les entrées x_1, x_2, \dots, x_N et calcule le potentiel ou la somme pondérée des entrées $x_1 W_1 + x_2 W_2 + \dots + x_N W_N$. Ensuite, une décision est prise pour calculer la sortie s en fonction du seuil θ :

Si $p > \theta$, alors $s = +1$

Sinon $s = -1$

Ce qui revient à tester si la différence $(p - \theta) > 0$. Il est donc équivalent de remplacer le seuil par une entrée fixe, de valeur -1 , avec un poids variable. Ceci est représenté par une fonction à seuil, appelée fonction de décision ou d'activation.

Les variations possibles de ce modèle viendront du choix du nombre d'entrées, de la valeur des poids ; ou de la fonction d'activation. Le nombre d'entrées dépend essentiellement des problèmes à aborder : En reconnaissance de forme par exemple, on pourra avoir autant d'entrées que de pixels dans l'image ; en classification de données on aura autant d'entrées que de mesures du processus observé. La fonction de décision est un élément du modèle qui dépend en partie du problème et de la sortie voulue soit en valeur réelle ou binaire. Dans le cas de la sortie réelle, une autre fonction couramment utilisée il s'agit de la fonction sigmoïde dont la caractéristique entrée/sortie est proche à celles d'un neurone réel.

Les derniers paramètres du modèle sont les poids ou les coefficients synaptiques W , qui dépendent du problème à résoudre. Ce sont eux qui construisent le modèle de résolution, en fonction des informations données au réseau : il faut donc trouver un mécanisme qui permette de faire les calculs à partir des grandeurs que l'on peut acquérir sur le problème. C'est le principe fondamental de l'*apprentissage*.

3.4 Historique :

C'est en 1943 que le premier neurone artificiel fût inventer par Mc Culloch et Pitts. Ce modèle est inspiré des architectures et des fonctions du système nerveux humain, plus exactement de la cellule nerveuse (neurone).

Cette découverte était comme une étincelle poussant les chercheurs à s'intéresser et s'approfondir dans ce nouveau domaine de recherche et c'est en 1950 que Rosenblatt a présenté son nouveau modèle de réseau de neurones appelé *Perceptron*, ce dernier était la réplique du modèle initial, présenté par Mc Culloch et Pitts mais avec la possibilité d'apprentissage des différentes couches de neurones (YEDDOU, 1998), ce qui constituait une nouveauté à cette époque surtout lorsqu'il s'agit de résoudre les problèmes non linéaires.

En 1969, Papert et Minsky ont démontré les limites du réseau de Rosenblatt et la nécessité de plusieurs couches identiques pour modéliser un phénomène physique le plus fidèlement possible (DAVALO, 1993). Les deux chercheurs ont reproché au perceptron l'inexistence d'une méthode d'apprentissage de plusieurs couches, ce qui mettra à l'ombre l'idée des réseaux de neurones.

Les recherches dans ce domaine furent réactivées durant les années 70, mais cette fois-ci les études portées sur les réseaux interconnectés, mais l'inexistence de moyens capables de réaliser des calculs très poussés obligea les chercheurs à abandonner ce concept.

Au début des années 80 une découverte vient propulser les réseaux de neurones permettant ainsi la résolution de vieux problèmes posés par Papert et Minsky. Cette découverte est due à un groupe de chercheurs de l'université de Stanford, leurs études se sont soldées par la mise en œuvre d'un algorithme d'apprentissage qui porta le nom de *Backpropagation* (JODOUIN, 1994). Désormais il est possible d'utiliser les réseaux multicouches sans se soucier de leurs apprentissages. En ce qui concerne le domaine de l'hydrologie, la modélisation avec les réseaux de neurones a été introduite au début des années 90. Les études réalisées portaient généralement sur la prévision des débits d'écoulement soit souterrain au niveau des aquifères (DONNA RIZO, 1994) soit superficiel, ainsi que la prévision spatio-temporelle de la précipitation au niveau des bassins versants (FRENCH et al., 1992) et la gestion d'un réservoir en temps réel (Sakakima et al., 1992).

3.5 Type de réseaux :

On distingue deux types de réseaux de neurones :

3.5.1 Réseaux à deux couches :

a- Perceptron :

Se basant sur les connaissances neurophysiologiques de l'époque et sur les travaux connexionnistes naissants, Rosenblatt proposa un modèle neuromimétique de la perception (YEDDOU, 1998).

Les modèles classiques de neurones comportent une seule couche de neurones. Les applications possibles avec ce type de modèle sont donc très limitées. Cependant on peut rassembler plusieurs neurones dans une même structure, ayant en commun les mêmes entrées, et que l'on nomme une couche de neurones.

Au sein d'une même couche, chaque neurone agit indépendamment des autres, et en particulier, il ne reçoit aucune connexion en provenance des autres neurones appartenant à la même couche. Cette structure permet de résoudre et de traiter des problèmes pour lesquels la sortie est attendue n'est pas une seule valeur scalaire, mais un vecteur de K valeurs scalaires.

Le modèle est composé d'une aire sensorielle qui lui sert d'entrée, une ou plusieurs aires d'association puis une aire de réponse où se situe la sortie du modèle. Chaque aire est

constituée d'une couche de neurones formels. Le Perceptron fonctionne de la façon suivante : une donnée perceptuelle est présentée au réseau en activant les neurones de l'aire sensorielle, l'activation se propage alors aux aires d'association puis éventuellement à l'aire de réponse, où l'on peut noter la réponse du réseau (JODOUIN, 1994).

b- Adaline :

Le modèle Adaline (ADaptive LINar Element) a été proposé par B. Widrow, chercheur américain à Stanford en 1960. Il travaillait sur les systèmes adaptatifs, et cherchait comment on pourrait construire un système capable de trouver une solution aux problèmes en ne connaissant, à chaque instant, qu'une partie des données du problème. Contrairement. Par exemple, construire un système qui apprenne à classer des pièces de voiture à partir d'une dizaine d'exemple, et qui soit capable de poursuivre le classement pour des pièces jamais rencontrées auparavant.

L'adaline est un modèle de neurone formel, semblable au modèle classique. Sa sortie est calculée comme une somme de ses entrées, pondérées par les coefficients synaptiques (poids). Les entrées sont à valeurs réelles, la sortie également et la fonction de décision est la fonction identité.

3.5.2 Réseaux multicouches :

Après avoir construit des couches de neurones, on généralise encore une fois en construisant des modèles à composition de couches. C'est une succession de couches de neurones, reliés entre elles par des coefficients synaptiques. Le modèle du neurone est non linéaire, de type sigmoïde ou, plus généralement, n'importe quelle fonction continue et dérivable.

- **Le Perceptron multicouche :**

Bien que les règles d'apprentissage soient relativement simples, les réseaux à deux couches demeurent limités au calcul de fonction très simple, d'où l'intérêt d'utiliser des réseaux plus évolués, contenant en outre des neurones cachés, d'où le perceptron multicouche mis au point par Rosenblatt en 1958 ce nouveau réseau une généralisation du modèle de l'Adaline. Il consiste en une succession de couches d'unités avec des fonctions de décisions (activation) différentiables, reliées entre elles par des coefficients synaptiques.

Acquérant une plus grande capacité de calcul, l'apprentissage de ces modèles devient plus difficile, c'est ce qui a conduit les chercheurs à être pessimiste quant à l'avenir du

connexionnisme . Tout le problème de l'apprentissage sur une telle structure vient du fait que la sortie que l'on désire associer à une entrée n'est disponible que pour les unités de sortie. Les unités dites internes, c'est-à-dire celles qui sont placées entre les entrées et la couche de sortie, n'ont aucune information sur le but à atteindre. Une méthode a donc été proposée par P. Werbos puis par D. Rumelhart et Le Cun, qui consiste à fixer un but aux unités internes à partir de l'erreur mesurée en sortie du réseau, et véhiculée par les poids qui relient les couches entre elles. En réalité cette méthode est une simple application de la méthode de la descente du gradient, et comme on constate que le calcul du gradient de l'erreur instantanée se fait de proche en proche dans le sens rétrograde des connexions, on a appelé cette méthode *rétro-propagation du gradient*.

Propriétés :

Le perceptron multicouches connaît de nombreuses applications, essentiellement grâce sa propriété d'approximation. Il a été montré qu'un réseau à une couche cachée est capable d'approximer n'importe quelle fonction avec la précision souhaitée, à condition que le nombre de neurones cachés soit suffisant. C'est la propriété essentielle de ce réseau, qui confère la capacité d'un approximateur universel. Néanmoins, cette propriété ne permet pas de choisir, pour un problème donné, le nombre d'unités optimal dans une couche cachée. Cette question est encore l'objet de recherches théoriques.

3.6 Concept et fonctionnement :

Les réseaux de neurones sont des modèles mathématiques et informatiques formés par la connexion connue sous le nom de liens synaptiques reliant plusieurs unités appelées neurones, cette technique est inspirée de la cellule nerveuse humaine, chaque neurone est caractérisé par un état appelé activation ou activité du neurone (DAVALO et NAÏM, 1993). Le neurone est animé par une fonction d'activation qui prend en considération l'état des neurones en amont et leurs poids de liens synaptiques.

Une fois l'activation déterminée, elle est propagée le long des liens synaptiques vers les neurones en aval, cette propagation représente le mode de calcul d'un réseau de neurone. (une entrée est introduite dans les neurones d'entrée sera propagée jusqu'au neurones de sortie).

Dans un réseau neuromimétique, les neurones n'agissent pas séparément mais le réseau à un comportement collectif qui est fonction de sa connectivité, donc de ses liens synaptiques ;

car la fiabilité du réseau repose sur l'exactitude de ses poids de liens. Pour la détermination des poids un processus appelé apprentissage est appliqué au réseau.

3.6.1 Pratique :

Le réseau de neurones a été développé à partir de trois couches connues respectivement sous le nom de : couche d'entrée, couche cachée et couche de sortie.

Chaque couche est constituée de plusieurs nœuds appelés neurones, les couches sont interconnectées entre elles par des jeux de poids.

Les nœuds appartenant à la première couche reçoivent l'entrée à partir de l'extérieur du modèle (entrée initiale). Cependant, les neurones des couches cachées et de sortie reçoivent les entrées à partir des interconnexions.

Les nœuds des couches cachées et de sortie transforment les entrées en sorties appelées "*firing rate*" en utilisant une fonction de transformation associée à chaque nœud connue sous le nom de "*fonction sigmoïde*" :

$$F(x) = \frac{1}{1 + \exp[-\beta(x - \theta)]} \quad (3.1)$$

X : est l'entrée des nœuds.

F(x) : est la sortie des nœuds (firing rate).

β : c'est le gain.

θ : le biais.

β et θ peuvent être : 1/ les mêmes partout.

2/ différents pour chaque couche.

3/ différents pour chaque couche en chaque nœud.

La fonction associée aux nœuds de la couche d'entrée est la fonction "*Identité*".

Dans le réseau de neurone qui sera utilisé dans cette étude nous avons retenu la deuxième proposition, c'est à dire celle où les valeurs de θ et β sont les mêmes à l'intérieur de chaque couche et différent d'une couche à l'autre.

En présentant un jeu de données aux nœuds de la couche d'entrée "*I_i*"

$i = 1, \dots, k$. ces valeurs sont multipliées par les poids des interconnexions $(Winput)_{ij}$, $j = 1, \dots, h$, où $(Winput)_{ij}$ sont les poids des interconnexions reliant le $i^{ème}$ nœud d'entrée au $j^{ème}$ nœud caché et "*h*" étant le nombre de nœuds cachés.

Le produit : $I_i \cdot (W_{input})_{ij}$ est sommé sur l'indice " i " et deviennent les entrées des nœuds cachés, comme le montre l'équation suivante :

$$(H_{input})_j = \sum_{i=1}^{k,h} I_i (W_{input})_{ij} \quad (3.2)$$

Où : $(H_{input})_j$ est l'entrée du " $j^{ème}$ " nœud caché.

(W_{input}) est le poids des interconnexions reliant le $i^{ème}$ nœud d'entrée au $j^{ème}$ nœud caché.

Chaque entrée des nœuds cachés est transformée à l'aide de la fonction sigmoïde définie plus haut pour donner les sorties de nœuds cachés $(H_{output})_j$, de là chaque sortie est multipliée par le poids interconnecté $(W_{output})_{jn}$, $n = 1, \dots, m$, entre le $j^{ème}$ nœud caché et chaque " m " nœuds de sortie.

L'entrée des " m " nœuds de sortie $(O_{input})_n$ sera la somme des produits exprimée comme suit :

$$(O_{input})_n = \sum_{j=1}^{h,m} (H_{input})_j (W_{output})_{jn} \quad (3.3)$$

Ces entrées $(O_{input})_n$ sont transformées à travers la fonction sigmoïde pour donner les sorties finales des " m " nœuds existants $(O_{output})_n$.

3.6.2 les fonctions d'activations :

a- Fonction sigmoïde exponentielle :

La fonction sigmoïde exponentielle est continue et dérivable. Vu ses nombreux avantages, ce type de fonction est généralement employé dans le Perceptron multicouches.

$$\frac{da_i}{dNet_i} = a_i(1 - a_i) \quad (3.4)$$

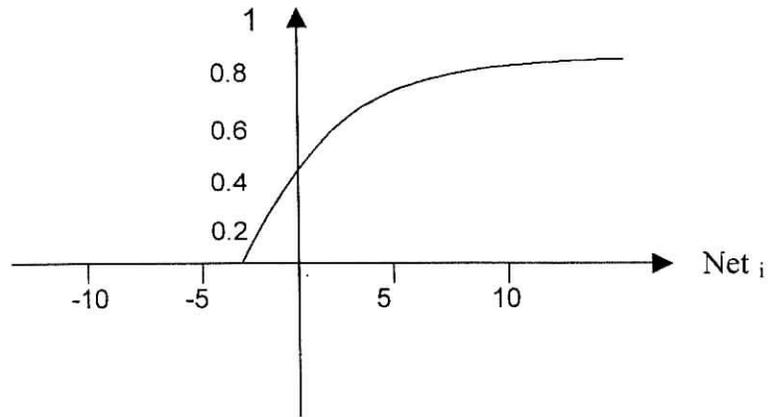


Figure 3.2 : Représentation graphique de la fonction sigmoïde exponentielle

Avec :
$$a_i = \frac{1}{1 + e^{-Net_i}} \quad (3.5)$$

b- Fonction sigmoïde tangentielle :

Autre fonction sigmoïde fort répandue. Il a été observé que l'apprentissage des réseaux multicouches converge plus rapidement quand cette fonction est employée plutôt que la sigmoïde exponentielle.

La dérivée de cette fonction est aussi simple à calculer que la précédente :

$$\frac{da_i}{dNet_i} = (1 - a_i^2) \quad (3.6)$$

Avec :
$$a_i = \tanh (Net_i) \quad (3.7)$$

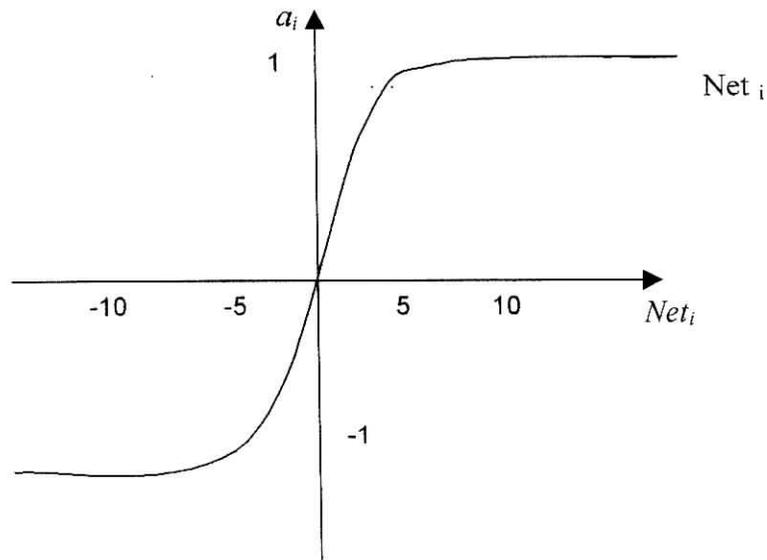


Figure 3.3 : Représentation graphique de la fonction sigmoïde tangentielle

3.7 Architecture et optimisation d'un réseau :

Dans les réseaux de neurones l'utilisation d'une seule couche cachée est généralement recommandée pour des études préliminaires. Cependant, l'utilisation de plusieurs couches cachées augmente le nombre de paramètres à estimer, ce qui rend très lourd le processus de détermination des poids du réseau, sans pour autant améliorer l'efficacité de celui-ci.

La détermination correcte du nombre de nœuds cachés est importante pour le succès de la modélisation.

Si la couche cachée comporte un nombre réduit de neurones, le modèle ne pourra pas reproduire le phénomène étudié d'une manière adéquate.

Il a été établi, d'une manière générale, qu'il n'existe pas de règle bien définie qui permet de dimensionner correctement un réseau en fonction du problème à résoudre, la meilleure stratégie pour le choix du nombre approprié de nœuds cachés et du nombre de couches est l'expérience.

3.8 Apprentissage d'un réseau de neurones :

L'apprentissage est une méthode qui consiste à déterminer les poids interconnectés, ce processus est itératif car on donne en premier lieu des valeurs arbitraires aux poids interconnectés sur l'intervalle $[-1, +1]$ ensuite on les corrigera suivant une règle bien déterminée (JODOUIN, 1994).

Il existe trois types d'apprentissage :

3.8.1 Apprentissage supervisé :

Ce type d'apprentissage nécessite la présence d'un superviseur qui présente au réseau les entrées et leur sorties désirées.

Il a pour rôle de ramener le réseau vers le comportement imposé par le superviseur ce qui se fait par la recherche du vecteur poids optimal qui assurera la meilleure approximation possible.

La technique d'apprentissage supervisé généralement utilisée est la rétropropagation de l'erreur appelée "*Backpropagation*".

- **Backpropagation :**

Cette technique a été élaborée en 1972 par Werbos dont le nom était "*The Dynamic Algorithm feed-back*".

Cette technique fut développée en 1986 à l'université de Stanford par une équipe de chercheurs (Rumelhart, Hinton et Williams (DAVALO, 1993), qui l'ont adaptée aux réseaux multicouches et l'ont appelée "*Backpropagation*" en référence de l'erreur qui se "rétropropage" à travers les couches du réseau.

Mathématiquement, cet algorithme utilise les règles de dérivations qui ne présentent aucune difficulté particulière.

Dans cet algorithme, de même que l'on est capable de propager un signal provenant des cellules d'entrées vers la couche de sortie, on peut, en suivant le chemin inverse, rétropropager l'erreur commise en sortie vers les couches internes.

On dispose d'un ensemble d'exemples qui sont des couples entrées - sorties désirées. Un exemple est présenté à l'entrée du réseau, une sortie réelle est calculée. Ce calcul est effectué de proche en proche de la couche d'entrée à la couche de sortie. Cette phase est appelée "*Propagation en avant*". Ensuite, l'erreur au niveau de la sortie du réseau est calculée, cette dernière sera rétro-propagée dans le réseau donnant lieu à une modification de chaque poids.

Ce processus est réitéré en présentant successivement chaque exemple. Si, pour tous les exemples introduits, l'erreur est inférieure à un seuil choisi, on dit alors que le réseau a convergé.

En résumé, l'apprentissage consiste à minimiser l'erreur quadratique commise sur l'ensemble des exemples, considérée comme fonction des poids, par une approximation d'une descente du gradient.

Toute la difficulté pour effectuer cette descente du gradient dans un réseau multicouche était de pouvoir calculer la dérivée de l'erreur quadratique par rapport à un poids donné. Donc l'utilisation de neurones à fonction d'activation dérivable permettra de résoudre ce problème.

- **Algorithme de la Rétropropagation de l'erreur :**

On considère un réseau de "*n*" neurones de sorties, avec un nombre quelconque de couches cachées.

On suppose qu'un neurone n'est connecté qu'à des neurones plus en aval, dans le sens entrée-sortie, mais pas forcément seulement à des neurones situés sur la couche immédiatement suivante.

Rappelons les notations utilisées :

- $\mathbf{X} = (X_1, X_2, X_3, \dots, X_n)$ est le vecteur des entrées.
- $\mathbf{Y} = (Y_1, Y_2, Y_3, \dots, Y_m)$ est le vecteur des sorties désirées.
- $\mathbf{S} = (S_1, S_2, S_3, \dots, S_m)$ est le vecteur des sorties effectivement obtenues.
- f est la fonction sigmoïde d'un neurone, f' est sa dérivée,
- O_j est la sortie du neurone "j",
- I_i est l'entrée du neurone "i",
- $E(k)$ est le pas du gradient à l'étape "k".

L'erreur commise sur un passage complet de l'ensemble des exemples est :

$$E(w) = \sum_k E^k(w) \quad (3.8)$$

Où l'on a noté $E^k(w)$ l'erreur commise sur un exemple k donné :

$$E^k(w) = (S^k - Y^k)^2 = \sum_{i=1}^m (S_i^k - Y_i^k)^2 \quad (3.9)$$

L'algorithme de rétro-propagation consiste à effectuer une descente de gradient sur "E". En général on utilise une version qui en est une approximation, où on effectue une modification de chaque poids après chaque passage d'un exemple "k", suivant la formule :

$$W_{ij}(k) = W_{ij}(k-1) - e(k) \cdot \partial E^k / \partial W_{ij} \quad (3.10)$$

Il faut donc calculer pour tous les poids "Wij" le gradient de "E^k" :

$$\text{Or : } \left\{ \begin{array}{l} \partial E^k / \partial W_{ij} = \partial E^k / \partial a_i \cdot \partial a_i / \partial W_{ij} \quad (3.11) \\ \partial a_i / \partial W_{ij} = \partial (\sum_p W_{ip} \cdot O_p) / \partial W_{ij} = O_j \quad (3.12) \end{array} \right.$$

En effet, l'indice " p " portant sur des neurones appartenant à la couche précédant celle de " i ", les sorties " O_p " de ces neurones ne dépendent pas de " W_{ij} ".

On obtient alors :

$$\partial E^k / \partial W_{ij} = \partial E^k / \partial I_i \cdot O_j \quad (3.13)$$

Ou, si l'on note :

$$d_i = \partial E^k / \partial I_i \quad (3.14)$$

$$\partial E^k / \partial W_{ij} = d_i \cdot O_j \quad (3.15)$$

Et donc la modification des poids sera :

$$W_{ij}(k) = W_{ij}(k-1) - e(k) \cdot d_i \cdot O_j \quad (3.16)$$

Pour un neurone " i " de la couche de sortie, on a :

$$d_i = \partial (\sum_j (S_j^k - Y_j^k)^2) / \partial I_i = 2 (S_i^k - Y_i^k) \cdot \partial S_i^k / \partial I_i \quad (3.17)$$

Car seul S_i^k dépend de I_i et comme $S_i^k = f(I_i)$, il vient :

$$d_i = \sum_h \partial E^k / \partial I_h \cdot \partial I_h / \partial I_i = \sum_h d_h \cdot \partial I_h / \partial I_i \quad (3.18)$$

Où l'indice " h " représente les neurones sur lesquels le neurone " i " envoie des connexions.

En effet, les entrées " I_h " des autres neurones ne dépendent pas de " I_i ".

On a donc :

$$d_i = \sum_h d_h \cdot \partial I_h / \partial O_i \cdot \partial O_i / \partial I_i \quad (3.19)$$

Or :

$$\partial I_h / \partial O_i = \partial (\sum_p W_{hp} \cdot O_p) / \partial O_i = W_{hi} \quad (3.20)$$

Car " p " portant sur les neurones envoyant une connexion vers " h ", ces neurones sont situés sur la même couche que " i " et donc les " O_p " ne dépendent pas de " O_i " (pour $p \neq i$).

Comme de plus, $O_i = f(I_i)$, on obtient finalement :

$$d_i = \sum_h d_h \cdot W_{hi} \cdot f'(I_i) \quad (3.21)$$

La règle de modification des poids à la présentation numéro k de l'exemple X est :

$$W_{ij}(k) = W_{ij}(k-1) - e(k) \cdot d_i \cdot O_j \quad (3.22)$$

Où l'on a :

$$1/ \quad d_i = 2 \cdot (S_i - Y_i) \cdot f'(I_i) \quad \text{pour la couche de sortie.} \quad (3.23)$$

$$2/ \quad d_i = \sum_h d_h \cdot W_{hi} \cdot f'(I_i) \quad \text{pour les couches cachées} \quad (3.24)$$

Où "**h**" : Neurones vers lesquels "**i**" envoie une connexion.

a- Data Adaptive Learning (FFN Pattern) :

Au passage de chaque exemple les poids sont réadaptés, le processus devient ainsi sensible à chaque exemple individuellement, ce qui le rend donc, facilement influençable par les bruits que peuvent contenir ces entrées durant l'entraînement.

Cette technique n'est donc utilisée que pour un apprentissage en temps réel c'est pour cela qu'elle est parfois appelée « *On line Optimisation Method* » (YEDDOU, 1998).

b- Block Adaptive Learning (Batch Learning) :

Dans cette méthode, la réadaptation des poids ne se fait qu'après passage de tous les exemples d'entraînement. La réadaptation est donc plus « *prudente* », le réajustement se fait suivant la moyenne de tous les exemples et la méthode est beaucoup moins sensible au bruit que peuvent contenir les exemples singuliers présentés au réseau.

C'est donc une méthode plus robuste et si l'application ne nécessite pas un apprentissage en temps réel, cette technique est préférable à la première (DAVALO et NAÏ M, 1993).

3.8.2 Apprentissage semi-supervisé :

Dans ce cas l'utilisateur ne dispose pas d'indication sur le comportement final que doit avoir le réseau.

3.8.3 Apprentissage non supervisé :

Les poids du réseau sont modifiés selon des conditions internes. C'est le réseau qui assure la correction des poids.

Remarque : pour simplifier l'apprentissage des poids des liens, aucun lien n'est permis entre les neurones d'une même couche.

3.9 Conclusion :

Les domaines d'application des réseaux de neurones sont, dans l'ensemble, liés à la perception et à l'intelligence artificielle, problèmes difficiles, dont les solutions classiques ne sont pas toujours complètement satisfaisantes et sont surtout très coûteuses en temps de calcul.

Les modèles des réseaux de neurones artificiels permettent de construire, par apprentissage, des solutions dans certains des domaines d'applications évoqués dans ce chapitre.

Il a été établi que les solutions sont souvent de qualités comparables à celles obtenues par des méthodes plus classiques. On conçoit alors, que l'enjeu pour rendre opérationnelle la technologie des réseaux de neurones est la mise en œuvre physique afin de les exploiter pleinement.

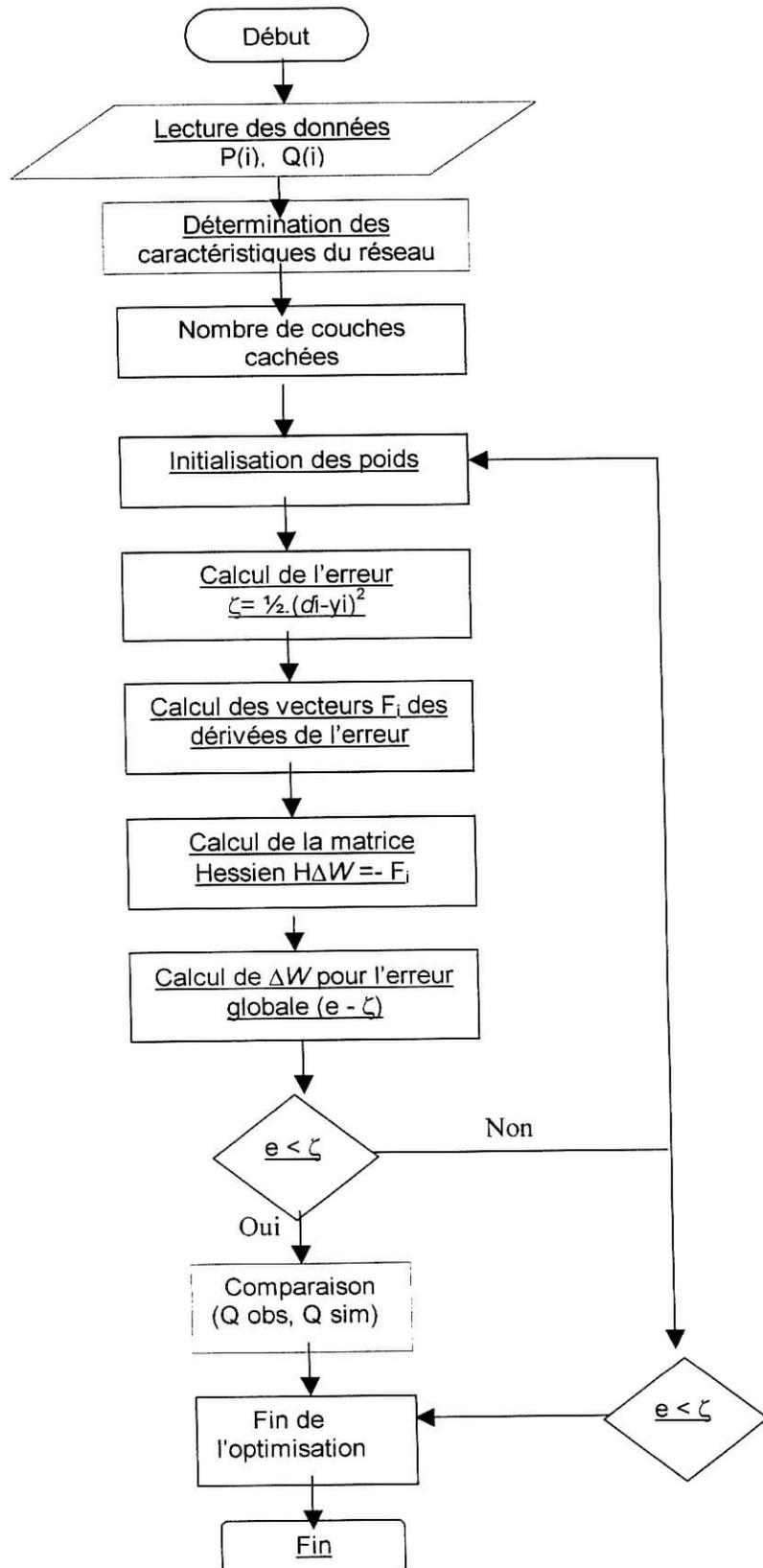


Figure 3.3 Organigramme des réseaux de neurones

CHAPITRE 4

**L'Analyse en Composantes Principales
(ACP)**

4.1. Généralités sur l'analyse des données

L'analyse des données désigne l'ensemble des méthodes à partir desquelles on collecte, organise, résume, présente et étudie des données pour permettre d'en tirer des conclusions et de prendre des décisions. Définie ainsi, l'analyse des données n'est pas une science nouvelle. Déjà, dans l'Égypte ancienne, le pharaon utilisait comme donnée un indicateur de richesse fourni par le nilomètre, sorte de puits creusé au bord du Nil qui permettait d'enregistrer la hauteur de la crue du fleuve et d'estimer ainsi la récolte sur laquelle était établi ensuite le montant de l'impôt. La collecte et l'étude des données sont importantes depuis longtemps pour bon nombre d'activités ou de professions. En réalité nous sommes tout le temps « bombardés » de données dans notre vie quotidienne, les médias nous livrent les derniers résultats de sondages en vue des élections, les gouvernements et les partis politiques se querellent sur les données du chômage, de l'inflation, de l'éducation, de la santé ; les services météorologiques parlent de prévisions en termes de données. Mais à quoi donc servent toutes ces données. Les données servent à obtenir de l'information, et l'information sert à décider, à agir. Le médecin analyse les données d'un patient pour effectuer un diagnostic et établir une ordonnance ; le chef d'entreprise analyse les données du marché en vue de vendre ses produits. Il est donc impossible d'échapper aux données. Mais pour passer des données à l'information et de l'information à la décision, il faut une méthodologie qui permettra de traduire des faits en tableaux numériques, établir des relations entre faits, prévoir de nouveaux faits.

Des méthodes sont utilisées dans l'étude de grandes masses d'information. Contrairement aux méthodes statistiques descriptives, elles peuvent réaliser un traitement plus poussé qui s'étale sur plusieurs variables. Ces méthodes permettent de mettre en évidence la vraie valeur de l'information. La simplification de grands tableaux d'information ou de données qui résultent de ces méthodes s'avère un outil de synthèse remarquable (Jumbu, 1999).

Parmi les méthodes connues dans ce domaine, on notera l'analyse factorielle à partir de laquelle plusieurs variantes se sont développées :

- L'analyse factorielle des correspondances AFC.
- L'analyse discriminante AD possédant un pouvoir descriptif et décisionnel, son objectif est de rechercher les axes qui acquièrent la meilleure discrimination du nuage de points.

La variable à expliquer est qualitative tandis que la variable explicative est mesurable.

- L'analyse de variance AV : applicable pour une variable à expliquée mesurable et une variable explicative qualitative.
- L'analyse en composantes principales ACP : permet d'étudier les tableaux de données quantitatives relatifs à un phénomène physique.

L'utilisation de l'analyse des données remonte au début du XX^{ème} siècle mais fautes de moyen de calcul puissant leur application étaient très réduite ce qui les a mis à l'ombre, c'est l'apparition des ordinateurs et leur développement très rapide qui vient propulser l'analyse des données est résoudre le problème de leur application.

Pour ce qui est du domaine de l'hydrologie et vu la quantité importante de données utilisées pour l'étude des phénomènes hydrologique l'ACP s'avère la meilleure méthode d'analyse grâce à sa capacité de concentrer le maximum d'information dans un nombre réduit de nouvelles variables.

4.2 L'analyse en composantes principales :

4.2.1 Objectif

L'objectif de l'analyse en composantes principales est de représenter graphiquement les relations entre variables quantitatives (ou assimilées à des variables quantitatives) et de visualiser en même temps les individus qui sont dans la relation avec ces variables. A partir de cette double visualisation, l'analyse pourra interpréter les résultats en fonction de sa connaissance du sujet étudié.

Interprété signifie donner sens aux axes factoriels, à des regroupements, à des singularités, à des proximités entre variables ou entre individus. C'est cette synthèse, fondée sur une exploitation des résultats d'analyse, qui permet à l'analyste de conclure, de construire d'autres hypothèses, de mettre en jeu d'autres relations entre variables jusqu'à ce que la conclusion soit suffisamment claire.

4.2.2 Notion d'individus et de variables

Dans un ensemble de données on peut distinguer : les individus et les caractères.

Le terme individu peut désigner une année d'observation ou une mesure de temps. Les individus peuvent provenir d'un échantillon appartenant à une population ou de la population elle-même.

L'individu «i » est décrit par le vecteur :

$$X_i = \{X_{ij}, j = 1, P\} \quad (4.1)$$

A chaque individu correspond un caractère ou variable, qui intervient dans le phénomène à étudier. Le caractère est représenté par le vecteur :

$$X_j = \{X_{ij}, i = 1, N\} \quad (4.2)$$

Le terme $X_j(i)$ est un nombre réel qui représente la mesure de la variable « X_j » sur l'individu « i ».

4.3. Définition de l'ACP

L'ACP consiste à décrire un ensemble d'individus et de caractères quantitatifs, elle est basée sur l'étude de la covariance ou de la corrélation entre variables ainsi que sur l'étude de la variance et la covariance de variables aléatoires.

4.4. Définition algébrique de l'ACP

L'ACP peut s'écrire sous forme matricielle $N \times P$ où on dispose de P variables pour N individus :

$$[X] = \begin{pmatrix} X_{11} & \dots & X_{1P} \\ \dots & \dots & \dots \\ X_{N1} & \dots & X_{NP} \end{pmatrix} \quad (4.3)$$

Après l'application de l'ACP, la matrice X sera remplacée par une matrice réduite, comprenant de nouvelles variables appelées Composantes Principales notées C.P.

4.4.1 Représentation graphique des individus

Les coordonnées des individus sur le premier axe sont les valeurs de la première composante principale, de même façon, les coordonnées des individus sur le deuxième axe sont les valeurs de la deuxième CP (figure 4-1)

En représentant le nuage des individus par leurs projections sur un plan, on ne conserve qu'une part de l'information totale. La variance est ici considérée comme dispersion de l'ensemble des individus et l'axe 1 est celui pour lequel la dispersion est maximale, l'axe 2 Possédant la même propriété sous contraintes d'orthogonalité au précédant et ainsi de suite.

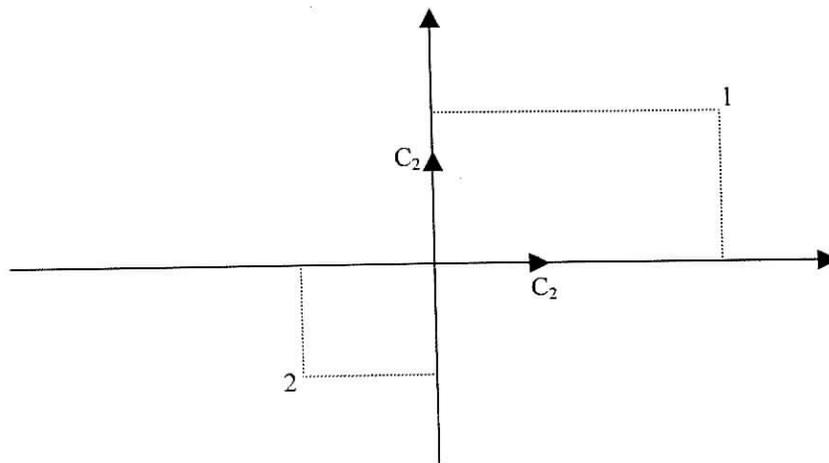


Figure 4.1 : projection de l'ensemble des individus sur un plan principal

4.5. Définition géométrique de l'ACP

Géométriquement, l'ACP consiste à déterminer un premier axe qui soit le plus proche de tous les points au sens des moindres carrés, de telle façon à avoir une somme des carrés des distances des N points à cet axe serait minimale. Cet axe est appelé axe factoriel.

Le deuxième axe est obtenu après projection des N points sur un plan orthogonal au premier axe, tel que la dispersion des projections des N points sur celui-ci serait toujours maximale, ce procédé est répété P fois.

Donc le premier axe va expliquer le maximum de la variance totale tandis que le second expliquera le maximum de la variance résiduelle non expliquée par le premier axe et c'est ainsi pour le reste des axes.

Pour l'explication des phénomènes hydrologiques, un certain nombre Q inférieur à P , d'axes factoriels peuvent donner une bonne explication sans pour autant avoir recours au reste des axes.

L'objectif de l'ACP étant de rendre la variance maximale : $\sum \text{Var}(C_j)$

$$\sum \text{Var}(C_j) = V_j [M] [R] [M] V_j \quad (4.4)$$

Avec :

C_j : Composante principale ;

$[R]$: Matrice des covariances des données ;

$[M]$: Métrique sur l'espace R^P .

4.6. Terminologie

Sous le vocable d'analyse en composantes principales, on rencontre plusieurs méthodes. On en distingue principalement deux ; l'analyse en composantes principales fondée sur la matrice des coefficients de corrélation linéaire, l'analyse en composantes principales fondée sur la matrice des covariances. Le terme de composante principale désigne la variable déduite des coordonnées factorielles associées à une analyse factorielle.

4.7. Les différentes méthodes :

4.7.1. L'analyse en composantes principales notée ACPnn

Supposons les variables d'origine très hétérogènes du point de vue moyenne. On peut analyser l'ensemble des variables indépendamment de l'effet naturel des mesures (en supposant les variables comptées dans le même système d'unités). L'ACPnn consiste à étudier un tableau de données centré X plutôt que le tableau d'origine.

$$X = \{X_{ij}; i \in I; j \in J\} \quad (4.5)$$

$$X = \left\{ X_{ij} = \frac{X_{ij} - \overline{X_j}}{\sqrt{n}} ; i \in I; j \in J \right\} \quad (4.6)$$

Où :

$\frac{1}{\sqrt{n}}$ est un coefficient de proportionnalité ;

X_{ij} : Variable centrée.

4.7.2 L'analyse en composantes principales normées notée ACP

On suppose désormais les variables d'origines non seulement hétérogènes quant à leur moyenne mais également quant à leur dispersion (les valeurs sont comptées dans des unités de mesure non comparables, ou la nature des variables n'est pas compatible). Il faut trouver un cadre de comparabilité commun. Pour ce faire, toutes les variables sont normaliser de telle sorte qu'elles soient de moyennes nulles et de variance unité : le tableau analyser est un tableau centré réduit (ou normé) :

$$X = \{X_{ij}; i \in I; j \in J\} \quad (4.7)$$

$$X = \left\{ X_{ij} = \frac{X_{ij} - \bar{X}_j}{s_j \sqrt{n}}; i \in I; j \in J \right\} \quad (4.8)$$

Dans ce cas, la matrice à diagonaliser $X'X$ est la matrice des coefficients de corrélation linéaire des variables :

$$X'X = \left\{ s_{jk}; s_{jk} = \sum_{i=1}^n \frac{(X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)}{s_j s_k n} \right\} \quad (4.9)$$

4.8. Description mathématique de l'ACP

4.8.1 Analyse des points individus

Comme c'est la forme du nuage qui nous intéresse, nous nous placerons au centre de gravité du nuage des points de base (normalisation centrée réduite). Ainsi les nouveaux axes initiaux X_1, X_2, X_3 seront translatés de O au centre de gravité G du nuage $N(I)$. On substitue à X le tableau X défini par : $\left\{ X_{ij} = \frac{(X_{ij} - \bar{X}_j)}{(s_j \sqrt{n})}; i \in I; j \in J \right\}$

L'influence des échelles de mesure sera moins sensible avec cette transformation.

- Distance entre points i de $N(I)$:

$$d^2(i; i') = \sum_{j=1}^p (X_{ij} - X_{i'j})^2 \quad (4.10)$$

Devient avec X_{ij} transformé

$$d^2(i; i') = \sum_{j=1}^p \frac{(X_{ij} - X_{i'j})^2}{(s_j^2 n)} \quad (4.11)$$

Ainsi, chaque variable aura une contribution égale dans la dispersion totale du nuage $N(I)$

- Calcul de la matrice $X'X$ à diagonaliser

$$X'X = \left\{ s_{jk} = \sum_{i=1}^n \frac{(X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)}{s_j s_k n}; j \in J; k \in j \right\} \quad (4.12)$$

$X'X$ est la matrice des corrélations entre variables. L'analyse consiste donc à chercher les vecteurs propres u_α et les valeurs propres λ_α de la matrice des corrélations empiriques.

- Coordonnées des points i de $N(I)$ sur l'axe de rang α .

Les coordonnées de i sur α sont données par construction de la façon suivante : considérons le vecteur $X u_\alpha$ $\begin{matrix} (n;p) \\ (p;1) \end{matrix}$

Les coordonnées des points de $N(I)$ sont les produits scalaires qui constituent les lignes du vecteur $X u_\alpha$ ($X u_\alpha = \lambda_\alpha^{1/2} v_\alpha$)

4.8.2 Analyse des points variables

Considérons l'espace R^n et les points variables j par rapport au repère initial O . Ils sont définis par :

$$X_j = \frac{X_{ij} - \bar{X}_j}{s_j \sqrt{n}} \quad (4.13)$$

Distance dans R^n :

$$d^2(O, j) = \sum_{i=1}^n \frac{(X_{ij} - \bar{X}_j)^2}{s_j^2 n} \quad (4.14)$$

$d^2(O, j) = 1$ (X_j est une variable centrée réduite) ; donc les variables X_j sont situées sur une sphère de rayon 1, centrée en O , origine initiale des axes.

La distance entre deux points du nuage $N(J)$ est :

$$d^2(j; k) = \frac{1}{n} \sum_{i=1}^n \left(\frac{(X_{ij} - \bar{X}_j)^2}{s_j^2} + \frac{(X_{ik} - \bar{X}_k)^2}{s_k^2} - 2r_{jk} \right) \quad (4.15)$$

$$= \frac{s_j^2}{s_j^2} + \frac{s_k^2}{s_k^2} - 2r_{jk} \quad (4.16)$$

$$d^2(j, k) = 2(1 - 2r_{jk}) \quad (4.17)$$

$r_{jk} = 1 \Rightarrow$ les points j et k sont confondus.

$r_{jk} = -1 \Rightarrow$ les points j et k sont opposés.

- Coordonnées des points variables sur les axes factoriels.

Ce sont par construction les composantes du vecteur $X' v_\alpha$ et par calcul ce sont aussi les composantes de $u_\alpha \lambda_\alpha^{1/2}$

4.9. Principe d'interprétation

Faire l'interprétation de la représentation graphique des variables, c'est donner les caractéristiques statistiques en fonction des positions de leurs points représentatifs.

La distance des variables à l'origine des axes est égale à sa norme (écart type), le cosinus de l'angle formé par des variables est égal à leurs coefficients de corrélation.

La formation de groupes de variables est généralement expliquée, soit par une relation entre les coefficients eux-mêmes, soit par les coefficients de corrélation, avec une composante principale, d'où la mise en évidence des caractères généraux de la population étudiée.

En ACP, la représentation graphique des variables est effectuée de telle manière que les points représentatifs sont sur la sphère de rayon 1, ce qui permettra une évaluation directe de qualité de la représentation de chaque variable en traçant un cercle de rayon 1 sur le plan examiné.

Ainsi on dira que la variance a été mieux reconstituée telle que son point représentatif sera proche du cercle et inversement.

4.10 Exploitation de l'analyse en composantes principales

4.10.1 Objectif :

L'exploitation de l'analyse en composantes principales ne s'arrête pas au calcul numérique des axes et à la diagonalisation d'une matrice. L'objectif de l'analyse est de pouvoir synthétiser les données en réduisant le nombre de variables à quelques facteurs significatifs et de déterminer les relations de proximité entre points individus et points variables.

4.10.2. interprétation numérique de l'ACP

L'exploitation de l'ACP consiste en premier lieu d'avoir près de soit le tableau de données ainsi que la matrice de corrélation qui permet de dégrossir les relations entre variables. Puis, sont étudiés les valeurs propres de la matrice des coefficients de corrélation. L'étude du spectre des valeurs propre montre qu'il y a deux ou trois facteurs à retenir pour l'exploitation des résultats. Par la suite les coordonnées des points individus et des points

variables dans l'espace formé par les deux ou trois axes factoriels retenus. Alors des tableaux récapitulatifs d'information sur ces points seront édités. La coordonnée d'un point j s'interprète comme un coefficient de corrélation entre l'axe factoriel et la variable.

L'interprétation d'un axe factoriel *consiste à donner un sens à cet axe*.

Ce sens est recherché à partir des éléments numériques de contribution des points individus à la variance des axes factoriels, et des contributions des points variables exprimés en termes de coefficients de corrélation. Pour favoriser le dépouillement des résultats, le travail consiste à chercher les points individus dont les contributions sont les plus fortes de coordonnées positives ou négatives.

Pour terminer l'interprétation numérique des éléments interprétatifs, l'étude de la qualité d'explication des points individus et des points variables dans l'espace factoriel retenu comme explicatif représente la phase finale de l'interprétation.

4.10.3 Interprétation graphique des axes factoriels

La représentation graphique des axes factoriels permet de saisir d'un seul coup d'œil ce que le calcul numérique ne peut fournir. Pour interpréter les axes factoriels, des représentations graphiques associés aux plans factoriels de l'espace factoriel considéré comme significatif.

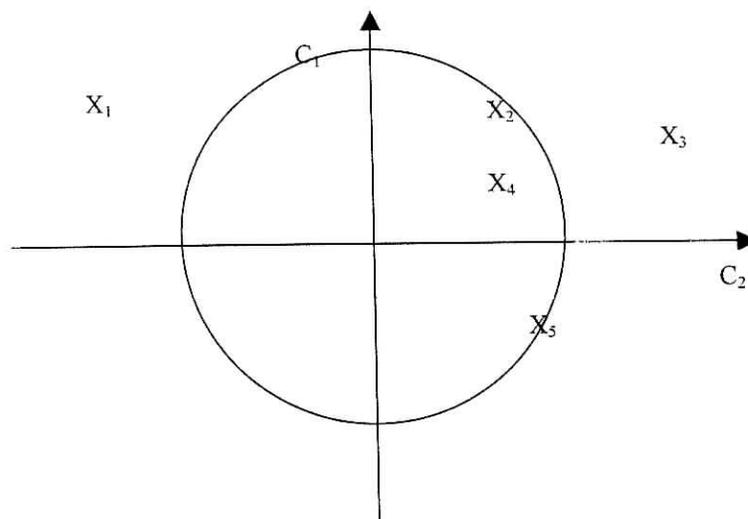


Figure 4.2 : Représentation graphique des variables dans le plan des deux premières C.P.

4.11 Procédé de calcul de l'ACP :

Calcul de la matrice R (covariance)

Elle est calculée de la manière suivante :

$$[R] = 1/n \cdot [X]^t [M] [X] \quad (4.18)$$

$$[R] = \frac{1}{n} [X]^t [M] [X] \quad (4.19)$$

[R] : Matrice des covariances ;

[M] : Métrique ;

[X] : Matrice des données ;

[X]^t : Matrice transposée.

- L'étape suivante consiste à déterminer les axes.

Comme il a été indiqué plus haut, le premier axe doit expliquer le maximum de l'information disponible, donc une variance maximale :

Après tous les calculs on obtient le vecteur V_1 qui représente le vecteur propre de la matrice $[R][M]$ va définir le premier axe principal. Ce vecteur propre sera associé à la plus grande valeur propre λ_1 de la même matrice.

Le même calcul sera réitéré pour obtenir les autres axes principaux.

Les valeurs propres obtenues seront classées par ordre décroissant, les CP résultantes suivront le même ordre. Pour avoir une CP normée d'écart type unité on divise toutes les CP sur la racine carrée des valeurs propres correspondantes.

Calcul des CP :

Les CP représente graphiquement la projection des points sur le système d'axe formé par les vecteurs propres.

$$[C] = [A]^t \cdot [X] \quad (4.20)$$

[A] : Matrice des vecteurs propres.

- **Quelques règles de sélection des axes et des points significatifs**

Dans la pratique, le nombre de variables étudiées est suffisamment grand que la lecture des tableaux et des représentations graphiques soit fastidieuse. Quelques règles simples qui permettent de faire la synthèse des principaux résultats. Les règles proposées ont été établies à

partir du dépouillement manuel de très nombreux cas de données réelles. Malgré leur empirisme, ces règles permettent à l'analyste d'extraire l'information des résultats (JUMBU,1999).

Règles de sélection des axes de premier ordre :

R1 : N est le nombre des axes α tels que $\lambda_\alpha > 1$;

R2 : N est le rang des axes α tel que ;

$$\sum_{\beta=1, N} \tau_\beta \geq p \text{ (pourcentage de variance).}$$

Le nombre p est fourni par expérience, il peut être modulé en fonction du type de données. Généralement pris égal à 80%.

Règles de sélection des axes de second ordre :

Il se peut que certaines singularités passent à travers les mailles de deux règles précédentes. Pour pallier à cette insuffisance les analystes proposent des règles pour déterminer les axes du second ordre :

R3 : soit N rang de l'axe factoriel à retenir. N est choisi tel que

Il existe au moins un élément i tel que $cor_N(i) \geq k$

Il existe au moins un élément j tel que $cor_N(j) \geq k$

K étant un nombre fixé à l'avance compris entre 0 et 1.

Cela exprime le fait qu'il existe des nuages de points $N(I)$ ou $N(J)$ qui, ont des coefficients de corrélation élevés avec un axe (k représente le carré d'un coefficient de corrélation).

A la fin de la sélection des axes factoriels à étudier, on aura N axes factoriels de 1^{er} ordre, et N axes de 2^e ordre. L'étude sera donc menée avec ces axes seulement.

b - Règles de sélection des points explicatifs des axes factoriels

Règles de sélection des points variables explicatifs des axes factoriels

Les coordonnées de j de $N(J)$ sont exprimés en termes de coefficients de corrélation. On retient alors la règle suivante qui sélectionne les variables j de $N(J)$ les plus fortement corrélés avec l'axe de rang α .

R4 : On choisit les éléments j de $N(J)$ tels que :

$$\text{cor}_\alpha(j) \geq k \quad (k \text{ étant fixé à l'avance compris entre } 0 \text{ et } 1)$$

la valeur k est déterminée par expérience. On prend par défaut $k \geq 0,5$ qui correspond à un coefficient de corrélation de 0,7.

Règles de sélection des points individus explicatifs des axes factoriels

$$\text{On à} \quad \sum_{i \in I} F_\alpha^2(i) = \lambda_\alpha. \quad (4.21)$$

L'axe α est sous dépendance des éléments i dont la contribution à la variance λ_α est la plus forte ; on prendra l'une ou l'autre des règles suivante

R6 : On choisit les éléments de i de $N(i)$ tels que la somme des contributions relatives des points i et $N(i)$ à l'axe α soit au moins égale à p , les points i étant triés par ordre décroissant de la variance (p est un pourcentage de variance fixé à l'avance ; par défaut pris à 80%).

R7 : On choisit les éléments i de $N(i)$ tels que la contribution de chaque élément i soit au moins égale à la moyenne des valeurs des contributions.

Les étapes généralement suivies dans une étude par l'analyse en composantes principales, peuvent être résumé dans l'organigramme présenté dans la figure 4-2.

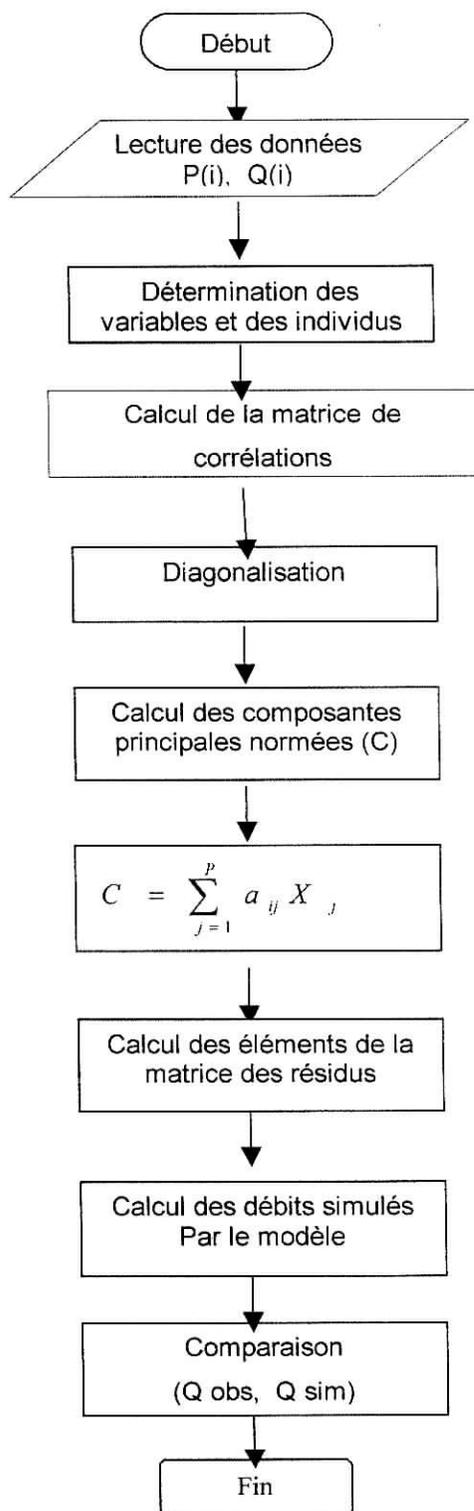


Figure 4.3 : Organigramme du modèle de l'ACP

CHAPITRE 5

**Application sur les données du bassin versant de
l'oued DIS (SEBAOU)
Résultats et interprétation**

5.1 Introduction :

Après une étude approfondie des trois méthodes et de techniques de résolution, nous procédons à leur l'application sur des données réelles, qui nous permettra de juger les performances et la fiabilité de ces méthodes ainsi que la qualité des résultats qu'elles génèrent.

Ce chapitre comportera deux parties : la première sera consacrée à la présentation du bassin versant et les événements averse-crue, la deuxième portera sur l'application des méthodes DPFT, Réseaux de Neurones et ACP sur les données du bassin de l'oued DIS ainsi que l'interprétation des résultats obtenus.

5.2 Présentation du bassin versant de Oued DIS :

Le bassin versant utilisé pour cette étude est celui de l'oued DIS, petit affluent du SEBAOU, il est situé dans la grande Kabylie à environ 25 km au nord-est de la ville de TIZI-OUZOU. (Figure 5.1)

5.2.1 Caractéristiques physiques du bassin de Oued DIS :

Les caractéristiques du bassin de l'Oued DIS peuvent être résumés dans ce qui suit :

Surface :	100km ²
Périmètre :	42 km
Coefficient de compacité de Gravélius :	$K = (0,28 P/\sqrt{S}) = 1,18$
Longueur du cours d'eau principal	19 km
Longueur du rectangle équivalent :	13,7 km
Indicedepente(I.P) :	0,26
Densité de drainage (Dd) :	3,4km / km ²
Coefficient de torrentialité :	24,3
Altitude Max :	1278 m
Altitude Min :	130 m
Altitude Moy :	476 m

Le bassin de l'Oued DIS est assez accidenté à l'amont de la station hydrométrique de FREHA. Il est formé de deux versants boisés présentant des pentes raides, ce qui favorise l'écoulement d'une part et d'autre part l'interception des précipitations par les végétaux.

Le climat de la région est tempéré, méditerranéen.

Le bassin est équipé de trois pluviomètres dont un couplé à un pluviographe (AZAZGA) et une station hydrométrique située à FREHA. Cette dernière est mise en service en 1973 et permet le contrôle des hauteurs d'eau, les données de débits fournies par l'ANRH de BLIDA s'étalent sur 15 ans entre 1985 et 1997.

On ce qui concerne les données de pluies au niveau de la station pluviométrique de AZZAZGA fournies aussi par l'ANRH de BLIDA ces dernières existent depuis 1973 jusqu'à 1997. Les tests d'homogénéité ont été effectués par l'agence qui nous a fourni ces données.

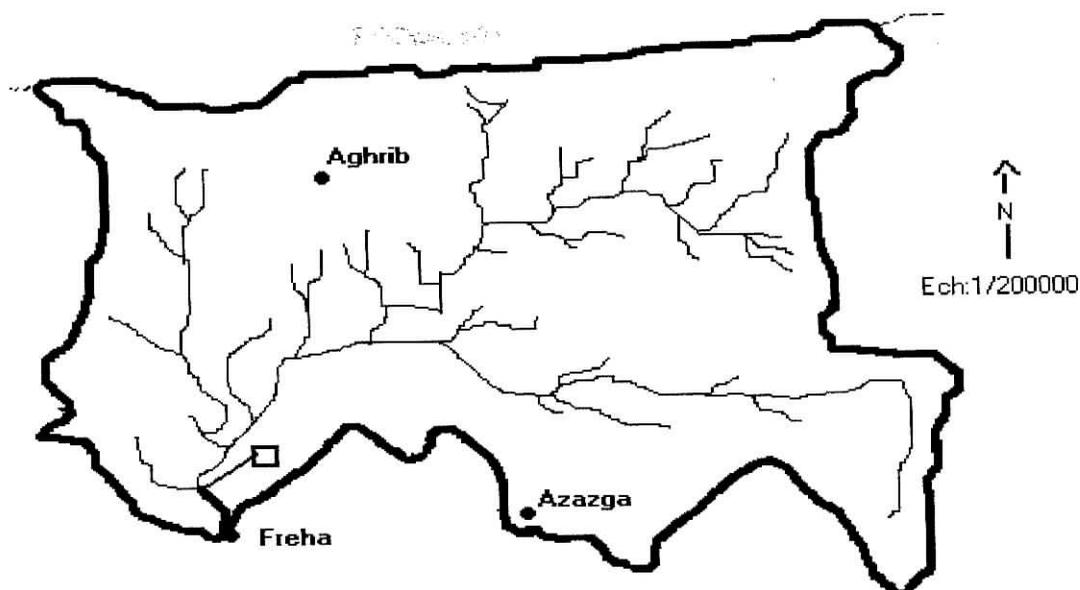


Figure 5 . 1 : le bassin versant de l'oued DIS

5.2.2 Choix des épisodes averse-crue :

Bien que les données de pluies et de débits soient disponibles et s'étalent sur des périodes considérables, nous n'avons pas pu faire sortir un nombre important d'événements averse-crue ; car dans la plupart des cas rencontrés nous avons eu affaire à des crues complexes ou multiples. Ces dernières nécessitent une décomposition très minutieuse pour pouvoir séparer les crues successives, et déterminer leur début et leur fin, en plus, il fallait mettre en évidence l'averse qui a généré chaque crue. Quelques épisodes ont été éliminés à cause du grand décalage entre les pluies et les débits, c'est à dire que les valeurs des débits n'étaient pas significatives et ne reflétaient pas celles enregistrées pour les précipitations.

Néanmoins et à défaut de crues simples (crue à un seul pic) nous étions contraints de prendre quelques épisodes contenant au maximum deux crues composées pour la constitution de l'échantillon d'étude. Enfin, la taille de l'échantillon retenu pour notre étude est égale à quinze (15) épisodes.

5.3 Application de la méthode DPFT sur le bassin de l'Oued DIS :

5.3.1 Procédure de résolution :

La méthode est appliquée selon la démarche citée au chapitre 3.

On utilisera :

- Nombre d'itération : 5
- Identification des coefficients de la DPFT par moindres carrés.
- Application de la contrainte de positivité sur les coefficients de la DPFT identifiés.
- Lissage exponentiel de l'extrémité de la fonction de transfert.
- Normalisation de la fonction de transfert.
- Déconvolution des pluies efficaces.
- Application de la contrainte : pluies efficaces inférieures aux pluies brutes.

Les paramètres que nous avons fait varier sont :

- La mémoire d'identification de la fonction de transfert.
- Le coefficient de la ridge régression.
- Point du début de lissage de la fonction de transfert.

Pour une comparaison plus rigoureuse entre les débits calculés par la méthode DPFT et les débits observés, nous allons estimer le coefficient de corrélation ainsi que le critère de comparaison de Nash pour l'échantillon de calage et de test.

Le coefficient de corrélation est donné par :

$$\rho_{xy} = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_y} \quad (5. 1)$$

Où :

ρ_{xy} : Coefficient de corrélation entre les débits calculés et observés ;

$$Cov(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) \quad (5. 2)$$

σ_x et σ_y : Ecart type de la série des débits observés et celle des débits calculés ;

x et y : Le débit observé et calculé.

Le critère de comparaison de Nash est défini par la formule suivante :

$$T(\%) = 100 \cdot \left(1 - \frac{\sum_{i=1}^n (Q_i - \bar{Q})^2}{\sum_{i=1}^n (Q_i - \bar{Q})} \right) \quad (5. 3)$$

Avec :

T : Coefficient de Nash en % ;

Q_i : Débit observé en (m^3/s) ;

\hat{Q}_i : Débit calculé en (m^3/s) ;

\bar{Q} : Moyenne des débits observés en (m^3/s).

Les coefficients de corrélation et de Nash sont noté respectivement "**CR**" et "**R²**"

a- Identification de la fonction de transfert :

Le calage du modèle DPFT consiste à déterminer la fonction de transfert. Le nombre d'épisodes pluie - débit utilisé pour le calage est égal au deux tiers (2/3) de la taille de l'échantillon c'est à dire dix (10) crues, ce qui constitue un nombre relativement faible par rapport aux recommandations de Nalbantis, qui préconise l'utilisation de vingt (20) à cinquante (50) épisodes (NALBANTIS, 1987). Néanmoins, il est à souligner que les quinze(15) épisodes que nous avons utilisés ont été judicieusement choisis de façon à éviter ceux mal configurés.

Par ailleurs, les caractéristiques de la fonction de transfert retenue après avoir varié les paramètres cités précédemment sont les suivants :

- La mémoire d'identification : $K=14$ pas de temps ;
- Un lissage effectué à partir du 9^{ième} point avec un coefficient de lissage $\alpha = 0.45$

b- Déconvolution des pluies efficaces :

Les pluies efficaces sont obtenues à partir du système de déconvolution :

$$\{q\} = [a] \{PE\} \tag{5. 4}$$

$$\{q\} = \begin{Bmatrix} q_1 \\ q_2 \\ \cdot \\ \cdot \\ q_n \end{Bmatrix}; \{PE\} = \begin{Bmatrix} PE_1 \\ PE_2 \\ \cdot \\ \cdot \\ PE_m \end{Bmatrix}; [a] = \begin{bmatrix} a_1 & 0 & 0 & \cdot & 0 \\ a_2 & a_1 & 0 & \cdot & 0 \\ a_3 & a_2 & a_1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_n & a_{n-1} & a_{n-2} & \cdot & a_{n-m+1} \end{bmatrix}$$

La déconvolution consiste à déterminer les pluies efficaces à partir de la fonction de transfert calculée à l'itération (i-1) et du débit observé.

Le coefficient de ridge régression a été pris égal à 10^{-1} . Il a été remarqué que pour des valeurs supérieures à 10^{-1} le système de déconvolution était perturbé, et la corrélation entre débit observé et calculé est très faible, par contre, les valeurs inférieures à 10^{-1} étaient sans effet sur le système de déconvolution.

5.3.2 Présentation des résultats et interprétation :

Pendant la période de **calage**, dix épisodes averse-cruve ont été utilisés afin de caler le modèle, ce qui représente les deux tiers de l'échantillon d'origine. Les résultats obtenus pour les crues : 1, 2, 3, 4, 5, 6, 7, 8, 9 et 10 donnent un coefficient de corrélation compris entre 0.70 et 0.98 et un coefficient de Nash variant entre 34 et 99.2% (tableau : 5-1-a et 5-1-b), les faibles valeurs des deux coefficients sont essentiellement dues à l'existence de crues complexes dans l'échantillon de calage, (figure 5-2).

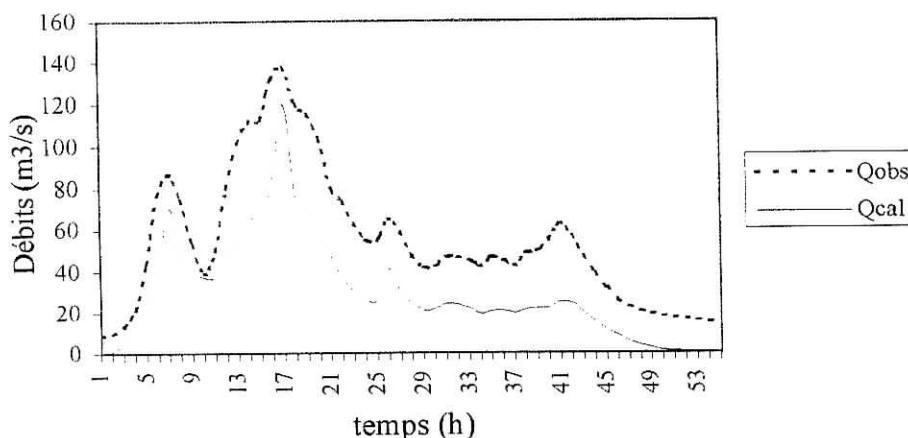


Figure 5. 2 : Reconstitution de la crue n°7 (calage)

Tableau 5-1-a : Paramètres statistiques du **calage** de la DPFT

Crue	Crue n° 1	Crue n° 2	Crue n° 3	Crue n° 4	Crue n° 5
Coeff de Corr.	0.94	0.97	0.98	0.96	0.97
Coeff. Nash (%)	71.3	99.2	90	91	91.8

Tableau 5-1-b : Paramètres statistiques du **calage** de la DPFT

Crue	Crue n° 6	Crue n° 7	Crue n° 8	Crue n° 9	Crue n°10
Coeff de Corr.	0.85	0.70	0.91	0.92	0.75
Coeff. Nash (%)	50	35	70	70.5	34

Après le calage du modèle, il nous a été possible de tracer la fonction de transfert et de voir son allure. La figure 5.3 montre l'évolution de la fonction de transfert à chaque itération où il est possible de constater sa stabilité à partir de la 3^{ème} itération et que son allure est presque confondue avec celle identifiée à la 4^{ème} et la 5^{ème} itération.

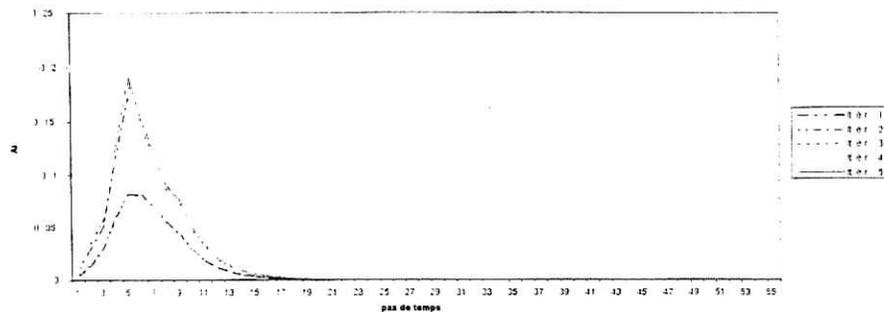


Figure 5. 3 : Calage de la fonction de transfert

Pour vérifier cette constatation, nous avons utilisé la contrainte de forme relative à la fonction de transfert qui exige que : $\Sigma FT_{\text{lissé}} = 1$.

En dressant le tableau 5-2 nous constatons effectivement que l'algorithme de résolution commence à converger à partir de la 3^{ème} itération avec une norme de 0,982 jusqu'à la convergence finale à la 5^{ème} itération avec une légère augmentation de la $\Sigma FT_{\text{lissé}}$ atteignant une valeur de 0,999. la convergence du processus itératif au bout de la cinquième itération a été observée par plusieurs auteurs (NALBANTIS, 1987), (RODRIGUEZ, 1989), (CHAMBAZ, 1990).

Tableau 5-2 : Valeur de $\Sigma FT_{\text{lissé}}$ à chaque itération.

Itération (i)	1	2	3	4	5
$\Sigma FT_{\text{lissé}}$	0,516	0,923	0,982	0,995	0,999

Après avoir identifié la fonction de transfert moyenne pour tous les épisodes ainsi que des pluies efficaces, il nous est possible de déterminer les débits issus de la résolution du système

$$\{q\} = [a] \{PE\} \quad (5. 5)$$

Le travail consiste à utiliser le tiers (1/3) restant des épisodes comportant des crues n'ayant pas participées au calage du modèle pour une prévision pluie-débit. La reconstitution s'est avérée acceptable pour les crues : 11 et 13 portées respectivement sur les figures : 5-4 et 5-5.

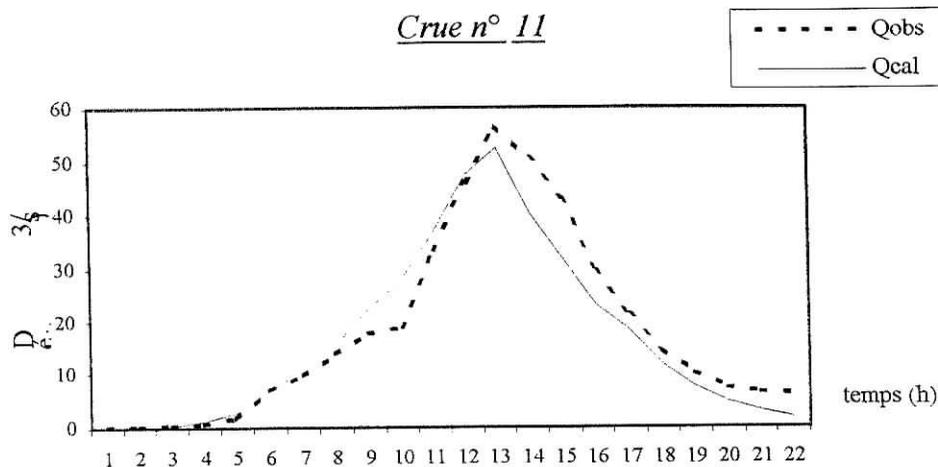


Figure 5-4 : Reconstitution de la crue n° 11 (période de test).

Tableau 5-3 : Paramètres statistiques de la prévision DPFT (période de test).

Crue	Crue n° 11	Crue n° 12	Crue n° 13	Crue n° 14	Crue n° 15
Coeff Corr	0.97	0.89	0.97	0.86	0.30
Coeff Nash (%)	92.8	65	89	55	86.5

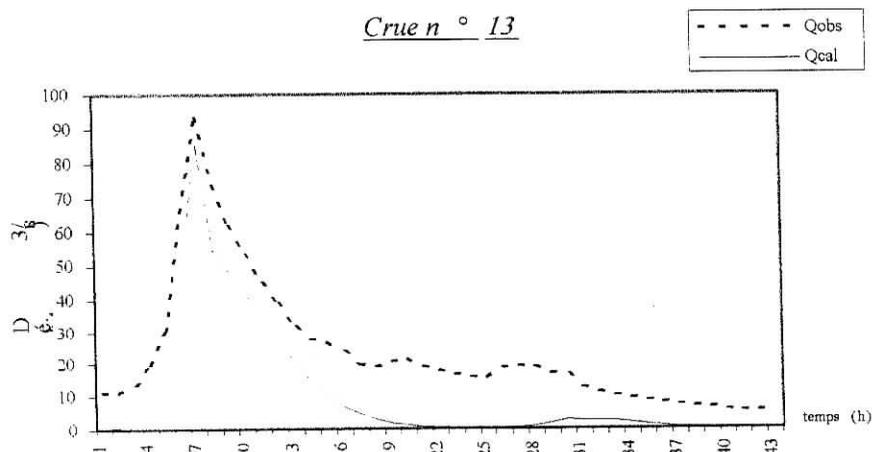


Figure 5-5 : Reconstitution de la crue n° 13 (période de test)

Nous remarquons que l'allure générale de l'hydrogramme issu des débits calculés est bien respectée, surtout au début de la crue, avec une légère sous-estimation des pics qui peut être expliquée par le fait que la méthode ne reproduit pas assez bien les pics.

Les valeurs du coefficient de corrélation et du coefficient de Nash relatifs au tableau 5-3 nous permettent de confirmer les remarques précédemment citées.

Effectivement, nous avons obtenu un coefficient de corrélation variant entre 0.89 et 0.97 pour les crues simples possédant un seul pic, ceci nous amène à affirmer nos constatations quant à la capacité de la méthode DPFT de reconstituer ce type de crue. Par ailleurs, les valeurs du coefficient de Nash qui varient entre 65 % et 92.8% montre que les débits calculés sont plus ou moins proches des valeurs réelles.

D'autre part, la reconstitution est moins bonne pour les crues : 12 et 14, représentées respectivement sur les figures : 5-6 et 5-7.

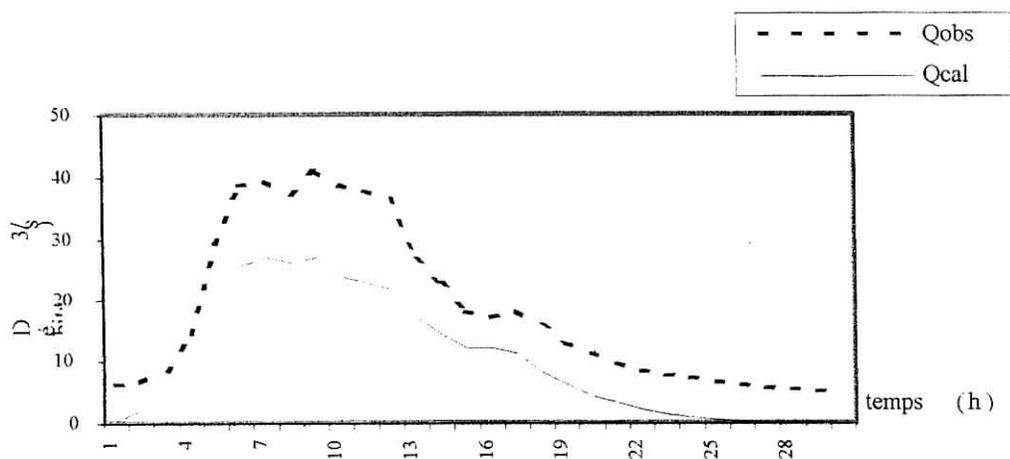


Figure 5-6 : Reconstitution de la crue n°12 (période de test)

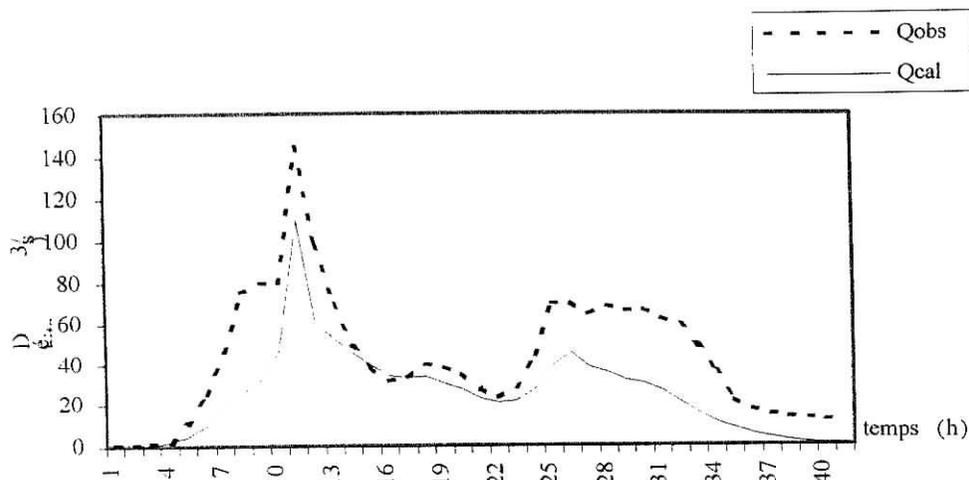


Figure 5-7: Reconstitution de la crue n°14 (période de test)

Nous remarquons que la forme générale de l'hydrogramme des débits observés est assez bien reproduite mais les valeurs de débits calculés étaient un peu loin des valeurs réelles. Cette sous estimation est très probablement due au fait que les crues testées sont du type composite possédant plus d'un pic. Effectivement nous avons pu constater un faible coefficient de corrélation compris entre 0.30 - 0.86 et un coefficient de Nash variant entre 55% et 55% (tableau 5-3) ce qui conforte nos remarques relatives à la mauvaise reconstitution des crues complexes.

5.3.3 Conclusion :

L'étude pratique de la méthode consiste à son application sur des données réelles dans le bassin versant de l'Oued DIS. Ladite étude a comporté une première partie concernant l'identification d'une fonction de transfert moyenne (calage du modèle) pour tous les épisodes pluie-débit, cette étape s'est soldée par la détermination d'une fonction vérifiant la contrainte de conservation de masse avec une $\Sigma A_i = 0.999$, ce qui nous permet de conclure du bon choix de la fonction.

Une deuxième partie intermédiaire qui consiste en la déconvolution des pluies efficaces. La combinaison entre ces deux parties nous conduit à l'étape la plus importante de cette étude qui réside dans la détermination des débits issus de la fonction de transfert et des pluies efficaces (teste du modèle), ce qui nous permet de reconstituer les crues. Cette reconstitution, selon nos constatations, et les critères de comparaison que nous avons utilisés, est acceptable

pour les crues simples avec des coefficients de corrélation et de Nash proches de 90% et moins appréciable pour les crues complexes où les mêmes coefficients étaient nettement plus faibles.

Finalement, nous pouvons conclure que malgré les difficultés rencontrées lors de la constitution de l'échantillon qui était de petite taille, que la méthode aurait été plus fiable si nous disposions d'un échantillon assez important.

5.4 Application du modèle Réseau de Neurones :

5.4.1 Introduction :

L'application du modèle des réseaux de neurones et sa validation se fera sur des événements averse crue déjà utilisés dans l'application de la DPFT sur des données pluies-débits à partir des stations d'AZAZGA pour les mesures de pluies et de FREHA pour celles des débits, constituant ainsi un échantillon de 15 crues.

5.4.2 Critères de validations :

La validation et la performance du modèle sont jugées à l'aide du coefficient de corrélation "*CR*" et celui de Nash "*R²*", qui ont déjà été utilisés pour la *DPFT*.

Il a été constaté que ces deux coefficients de corrélation peuvent donner des résultats différents selon la position des valeurs extrêmes par rapport à la majorité des observations. En effet, "*R²*" est sensible aux valeurs extrêmes même si l'erreur moyenne est faible, il peut donc être pénalisé par quelques fortes erreurs. Ce qui n'est pas le cas du coefficient "*CR*" qui est moins influencé par des valeurs extrêmes (LEK, DIMOPOULOS, DERRAZ, 1996).

5.4.3 Présentation et interprétation des résultats :

Dans cette étude un réseau de neurones à trois couches a été utilisé pour la modélisation de la relation pluie-débit. Le modèle en question est constitué d'une couche d'entrée, comportant un neurone d'entrée, une couche cachée, comportant cinq neurones cachés et une couche de sortie comportant un neurone de sortie. Le choix du nombre de couches ainsi que du nombre de neurones a été décidé après avoir effectué différentes combinaisons. Nous avons effectivement pu constater, à travers plusieurs essais, que l'augmentation du nombre de couches cachées ou celui des neurones ne conduisait pas à une amélioration des résultats, au contraire il rendait le réseau plus difficile à caler et son temps d'apprentissage plus long, confirmant ainsi ce qui a été constaté par d'autres auteurs (MASTER, 1993), (FORTIN, 1997), (FRENCH, 1992). L'échantillon utilisé pour le calage représente deux tiers de

l'échantillon initial donc dix crues, le tiers restant (cinq crues) est utilisé pour le test du modèle.

Interprétation :

Le calage du modèle réseaux de neurones consiste à la détermination des poids interconnectés, entre les couches composant le réseau, pour cela, un échantillon de dix crues a été utilisé. Le calage du réseau nous a permis d'avoir une reconstitution des crues avec des coefficients de corrélation et de Nash très appréciables (tableaux : 5-4-a, 5-4-b), les figures 5-8 et 5-9 nous permet de remarquer ce fait.

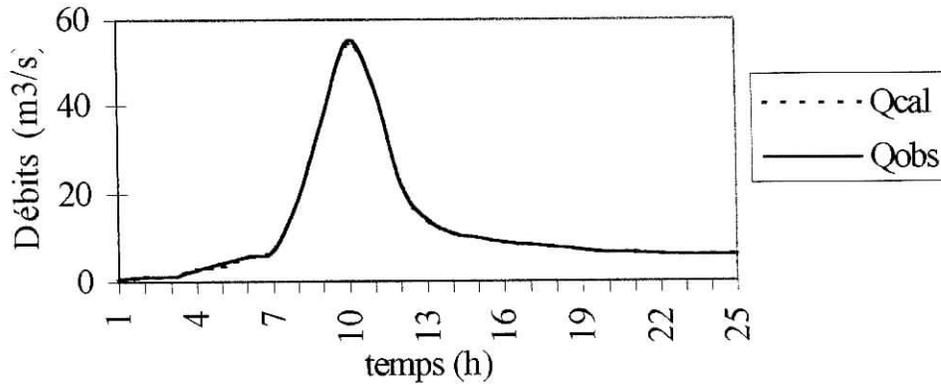


Figure 5-8 : Reconstitution de la crue n° 1

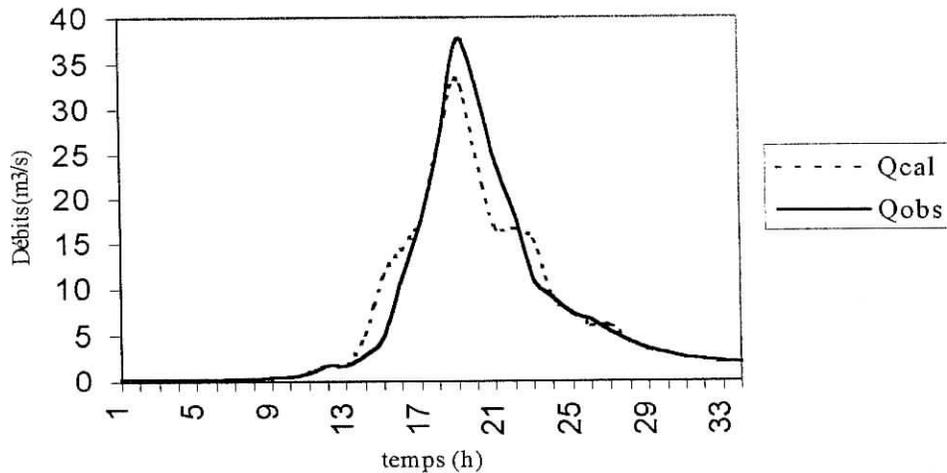


Figure 5-9 : Reconstitution de la crue n° 5

Pour la crue n°1, la figure 5-8 montre que les valeurs estimées par le réseau de neurones suivent parfaitement les valeurs observées, ce qui nous donne un très bon coefficient de

corrélation atteignant une valeur de 0.99, de même que pour le coefficient de Nash, $R^2 = 99\%$, (tableau 5-4-a), cela nous permet de dire que la crue n°1 a été fidèlement reproduite.

Tableau 5-4-a : Paramètres statistiques de la période de calage.

Crue	Crue n°1	Crue n°2	Crue n°3	Crue n°4	Crue n°5
Coeff Corr.	0.99	0.94	0.90	0.93	0.97
Coeff Nash %	99	88	80	84	93

Tableau 5-4-b : paramètres statistiques de la période de calage.

Crue	Crue n°6	Crue n°7	Crue n°8	Crue n°9	Crue n°10
Coeff Corr.	0.94	0.94	0.88	0.97	0.97
Coeff Nash %	88	89	76	94	94

Le test du modèle réseau de neurones a été effectuée sur les cinq crues restantes. Nous avons constaté la capacité du modèle à prévoir les débits, l'allure des hydrogrammes était bien repris par le modèle comme le montrent les figures 5-10 et 5-11. Les coefficients de corrélations propres à ces crues sont compris entre 0.95 et 0.97.

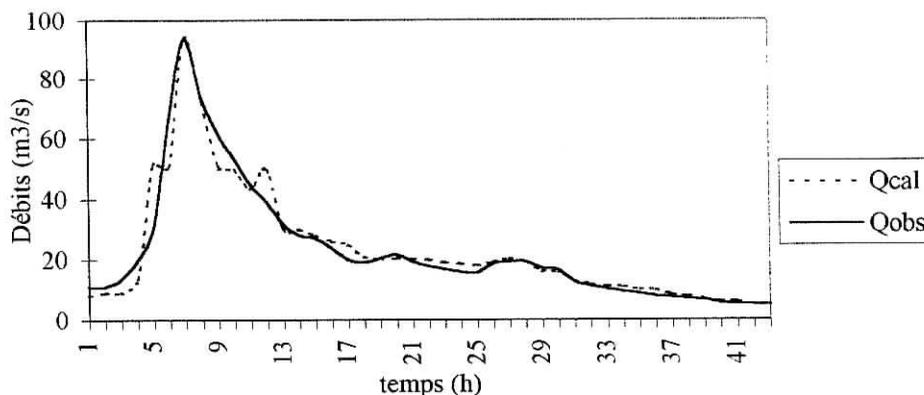


Figure 5-10 : Reconstitution des crues test (crue n°13)

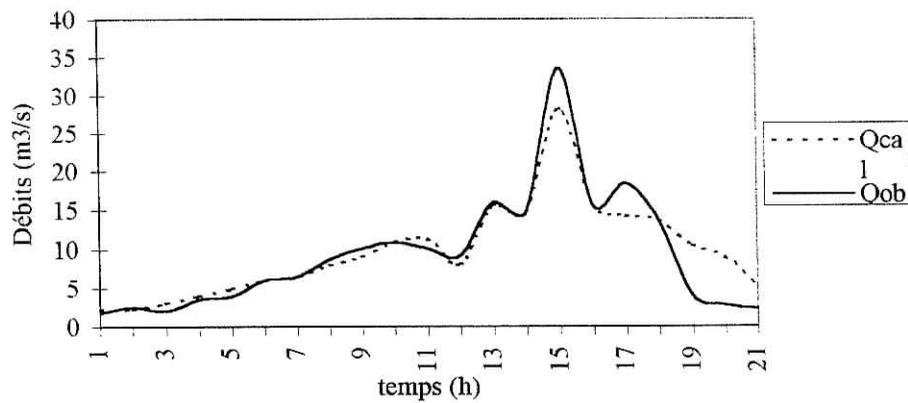


Figure 5-11 : Reconstitution des crues test (crue n° 15)

Tableau 5-5 : paramètres statistiques des crues de test

Crue	Crue n°11	Crue n°12	Crue n°13	Crue n°14	Crue n°15
Coeff Corr.	0.98	0.92	0.97	0.79	0.95
Coeff Nash %	96	82	93	62	88

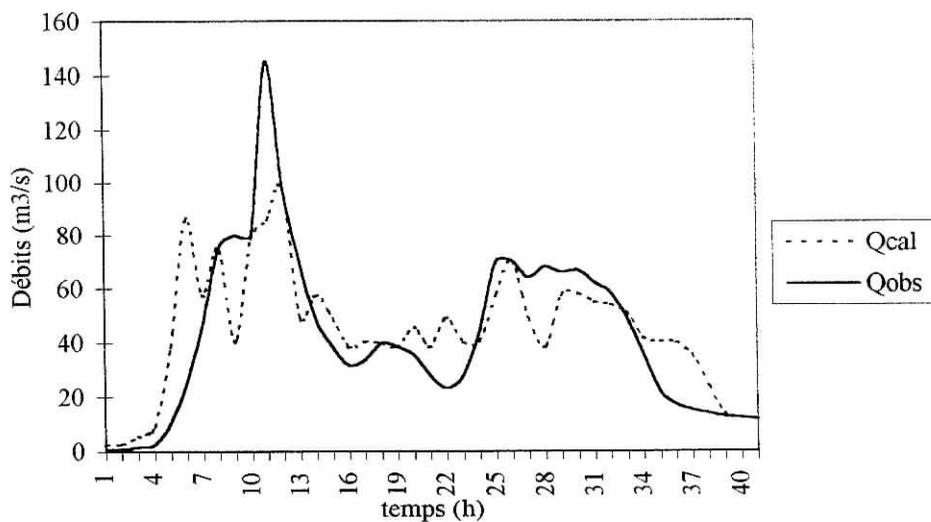


Figure 5-12 : Reconstitution des crues test (crue n°14)

5.4.4 Conclusion

Dans cette étude, un modèle de réseau de neurones à trois couches, avec un neurone à la couche d'entrée, cinq neurones à la couche cachée et un neurone à la couche de sortie a été utilisé afin de simuler un phénomène hydrologique complexe pluies-débits. L'estimation des différents paramètres du modèle ainsi que son entraînement s'est fait à l'aide du logiciel MATLAB dans sa version 5.3.

Le potentiel des réseaux de neurones pour la simulation en hydrologie présenté dans cette étude indique clairement qu'ils peuvent apporter des réponses satisfaisantes pour la modélisation pluies-débits et la prédiction des débits au pas de temps horaires avec un très haut degré de fidélité, en effet les résultats obtenus confirme notre idée de départ, quant à la capacité des réseaux de neurones à bien modéliser les phénomènes physiques non-linéaires.

Néanmoins, nous avons pu constater une certaine faiblesse quand il s'agit de reproduire des valeurs de débits extrêmes : le modèle à ce moment surestime ou sous-estime la valeur cible, ce phénomène a effectivement été constaté par d'autres auteurs (SHAMSELDIN, 1997), cela peut s'expliquer en partie par la fonction sigmoïde, qui est bornée entre [0,1], et de là, une relative difficulté à générer de grande valeurs de débits (FORTIN, 1997), (OUARDA, 1999).

5.5 Application de l'analyse en composantes principales :

5.5.1 ACP sur les données utilisées :

Les données qui seront utilisés pour l'application de l'analyse en composantes principales, sont les même utilisées pour la DPFT et les réseaux de neurones, c'est ainsi que les quinze épisode sont introduits dans cette application.

Une analyse en composantes principales appliquée à ces données selon la procédure suivante :

- 1- Constitution de la matrice de données brutes $m \cdot n$, avec :
 - m : Nombre de variables ;
 - n : Nombre d'observation.
- 2- Calcul des paramètres statistiques (moyenne et écart type).
- 3- Calcul des variables centrées réduites.

$$X_{CR} = \frac{X_i - \bar{X}}{\sigma} \quad (5.6)$$

Avec :

X_i : Variable brute ;

X : Moyenne des variables ;

σ : Ecart type.

4- Calcul des coefficients de corrélation à partir des données centrées réduites

$$COR(x, y) = \frac{COV(x, y)}{\sigma_x \sigma_y} \quad (5. 7)$$

COV(x, y) : Covariance des variable x et y ;

σ_x, σ_y : Ecart type.

5- Détermination des m valeurs propres et vecteurs propres (de la matrice de corrélation)

6- Calcul des composantes principales :

CP = Matrice variables centrées réduites \otimes Matrice des vecteurs propres.

$$(\mathbf{m}, \mathbf{m}) = (\mathbf{m}, \mathbf{n}) \cdot (\mathbf{m}, \mathbf{m})$$

7- Calcul de composantes principales

$$CP'_i = \frac{CP_i}{\sqrt{\lambda_i}} \quad (5. 8)$$

CP'_i : Composante principale normée ;

λ : Valeur propre.

8- Calcul de la matrice des coefficients de corrélation entre CP' et variables centrées réduites

$$COR(C', x_{cr}) = \frac{COV(C', x_j)}{\sigma_{c'} \cdot \sigma_{x_j}} \text{ avec } \sigma_{c'} = 1 \text{ d'où} \quad (5. 9)$$

$$COR(C', x_j) = \frac{\frac{1}{n} \sum_{j=1}^n C'_j(x_j, \bar{x})}{\sigma_x} \quad (5. 10)$$

9- Tracer les cercles de corrélation

10- Calcul des coefficients de régression

$$C_{REG} = \sigma_{x_j} \cdot (COR(CP' \cdot X_j)) \quad (5. 11)$$

C_{REG} : Coefficient de régression ;

COR(CP', X_j) : Coefficient de corrélation .

11- Reconstitution des variables :

$$X_j = \beta_0 + \sum \beta_j \cdot CP'_j + \varepsilon_j \quad (5. 12)$$

avec :

β_0 : Moyenne variable ;

β_j : Coefficients de régression ;

CP'_j : Composantes principales ;

ε_j : Résidu.

12- Calcul des résidus

$$\varepsilon_j = X_j - (\beta_0 + \sum_{j=1}^m \beta_j \cdot CP'_j) \tag{5.13}$$

L'analyse en composantes principales utilisée sur l'ensemble des épisodes pluie - débit est une ACP normée.

Les vecteurs propres sont obtenus après diagonalisation de la matrices des correiations (tableau 5. 6)

Tableau 5.6 : Les valeurs propres correspondants à chacune des 10 CP

CP	1	2	3	4	5	6	7	8	9	10
Valeur P.	16.27	7.24	4.82	2.80	1.98	1.64	1.01	.095	0.77	0.72

La contribution de chaque CP est représentée dans le tableau 5. 7 :

Tableau 5.7 : Contribution de chacune des CP

CP	1	2	3	4	5	6	7	8	9	10
Valeur P.	38.73	17.23	11.48	6.68	4.71	3.91	2.41	2.26	1.84	1.71

Nous remarquons que les trois premières composantes expliquent plus de 67% plus de la variance totale.

5.5.2 Modèle élaboré :

Dans ce cas de l'étude de la relation pluie-débit, vu l'hétérogénéité des variables pris en considération, nous avons utilisé l'analyse en composantes principales normée, pour le développement du modèle de prévision.

$$C = \sum_{j=1}^p a_{ij} X_j \tag{5.14}$$

Avec :

C : Composante principale ;

a_{ij} : Cosinus directeur ;

X_j : Variable d'ordre j.

Soit la variable centrée réduite : $Y_j = \frac{X_j - \bar{X}_j}{\sigma_j}$

La composante principale nommée C'_1 s'écrit :

$$C'_1 = \lambda^{1/2} \sum_{j=1}^p a_{1j} \frac{X_j - \bar{X}_j}{\sigma_j} \quad (5.15)$$

Sous la forme matricielle, l'expression (5.21) s'écrit :

$$|C'| = [\lambda^{-1/2} |A| |Y| \quad (5.16)$$

$$\text{D'où :} \quad |Y| = |\lambda^{1/2} |A|^T |C'| \quad (5.17)$$

Y : Matrice des variables centrées réduites ;

λ : Vecteur des valeurs propres ;

C' : Matrice de composantes principales normées ;

A^T : Transposée de la matrice des vecteurs propres.

sachant que :

$$\lambda^{1/2} a_{1j} = COV(C'_1, Y_j)$$

$$\lambda^{1/2} a_{1j} = COV(C'_1, \frac{X_j - \bar{X}_j}{\sigma_{X_j}}) = COR(C'_1, X_j)$$

L'équation (5.17) devient :

$$Y_j = \sum_{l=1}^q COR(C'_l, X_j) C'_l + \varepsilon_j \quad (5.18)$$

D'où :

$$X_j = \bar{X}_j + \sigma_{X_j} \sum_{l=1}^m COR(C'_l, X_j) C'_l + \sigma_{X_j} \varepsilon_j \quad (5.19)$$

L'expression finale du modèle de prévision est donnée par :

$$X_j = \beta_{j0} + \sum_{l=1}^m \beta_{jl} C'_l + E_j \quad (5.20)$$

Avec :

$$\beta_{j0} = \bar{X}_j \quad (\text{la moyenne de la variable d'ordre } j).$$

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

En se basant sur la structure de ces deux 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 C P. Dans le cas d'une ACP normée, celle-ci est constituée d'un premier vecteur β_0 , dont les éléments sont les moyenne de chaque épisode.
- Les éléments de la matrice des C P : $[C']$.
- Les éléments de le matrice des résidus : $[E]$

5.5.3 Reconstitution

La reconstitution a été faite en tenant compte des trois premières composantes principales, les valeurs négatives reconstituées au début de la crue non aucune influence sur le modèle de prévision projeté, elles peuvent être facilement éliminer en appliquant une transformation logarithmique aux données brutes.

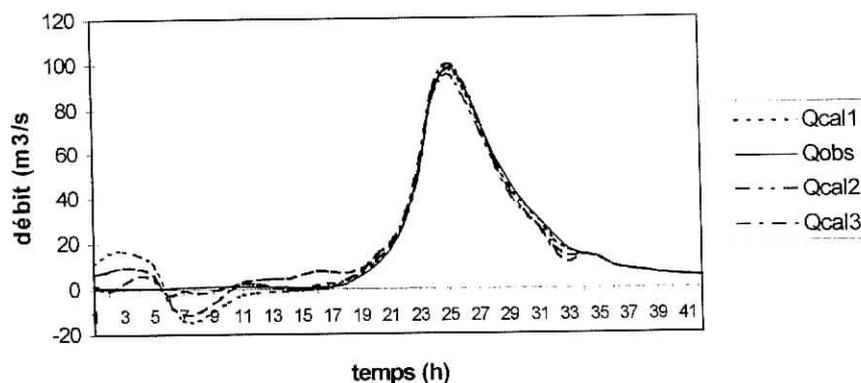


Figure 5-13 : Reconstitution de la crue n° 6 par le modèle

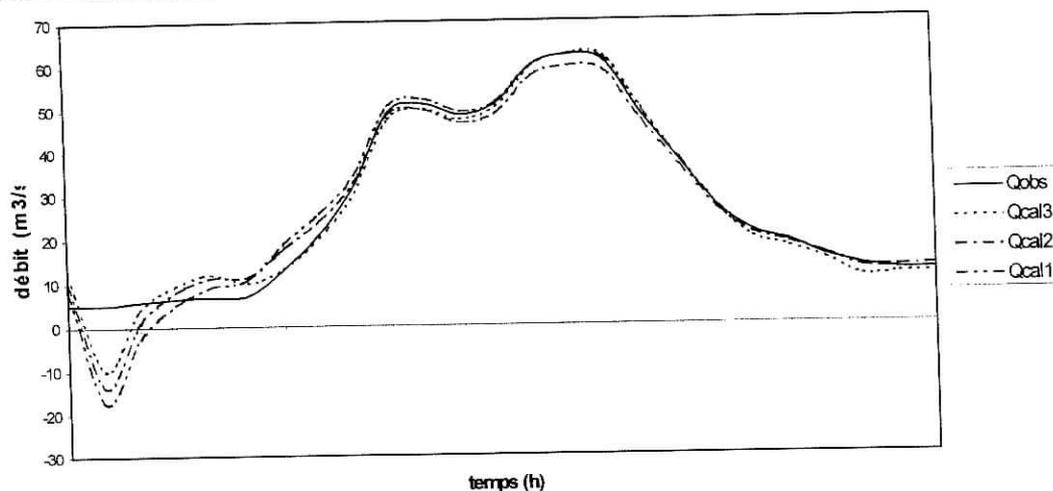


Figure 5-14 : Reconstitution de la crue n° 9 par le modèle

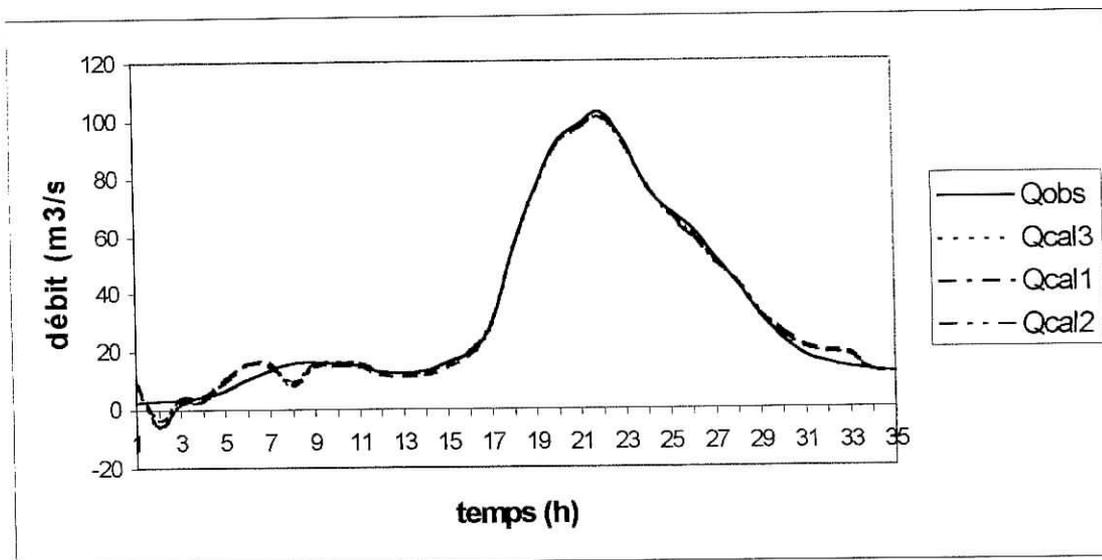


Figure 5-15 : Reconstitution de la crue n° 7 par le modèle.

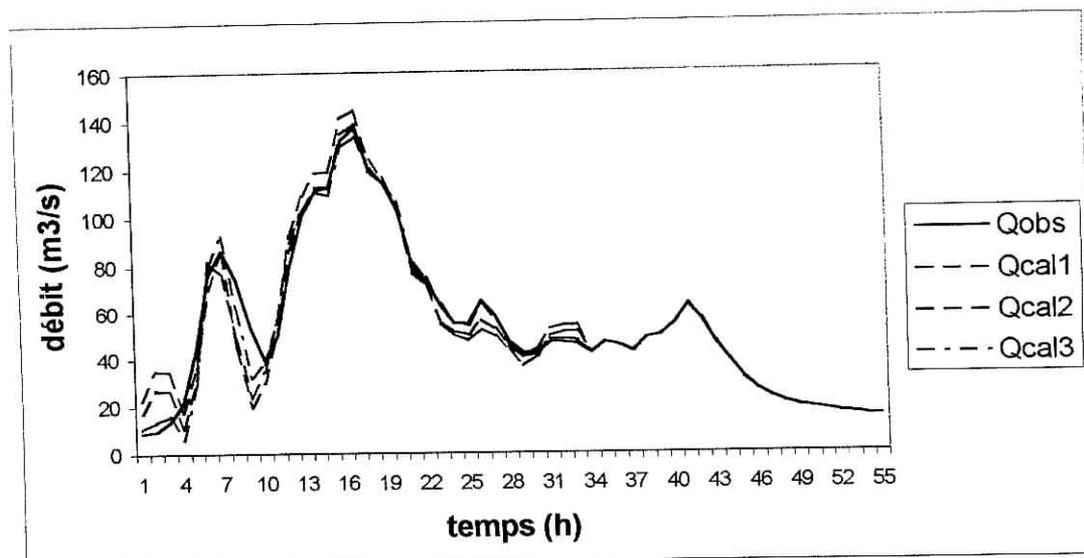


Figure5-16 : Reconstitution de la crue n° 8 par le modèle.

5.5.4 Prédiction :

Les moyennes des β_0 des débits sont déterminées, en connaissant la pluie de départ de l'épisode, nous avons choisi l'ACP correspondante à cette pluie. La même approche a été appliquée pour les résidus ε (figure 5-). Les résultats obtenus s'avèrent non représentatifs de la réalité, c'est pour cela qu'une autre voie de recherche a été choisie, nous avons changé d'approche.

Pour améliorer le modèle, le choix du β_0 est fait en fonction de la saison étudiée, dans ce cadre là, une moyenne saisonnière des β_0 est prise en considération. Le choix des β_i et des CP est fait en fonction de situations similaires dans l'historique, et dans ce cadre là, la pluie tombée juste où nous réalisons la prédiction sera le guide pour ce choix, pour tenir compte de l'état d'humidité du sol, et donc, sur son influence sur la valeur du débit sortant.

Une approche des résidus consécutifs a été faite. Donc, au lieu de prendre un résidu à une situation donnée, nous avons utilisé des résidus à $t - 1$ et $t - 2$. L'élaboration du modèle a permis une nette amélioration des prévisions, comme nous pouvons le constater sur les figures 5-17, 5-18 et 5-19 :

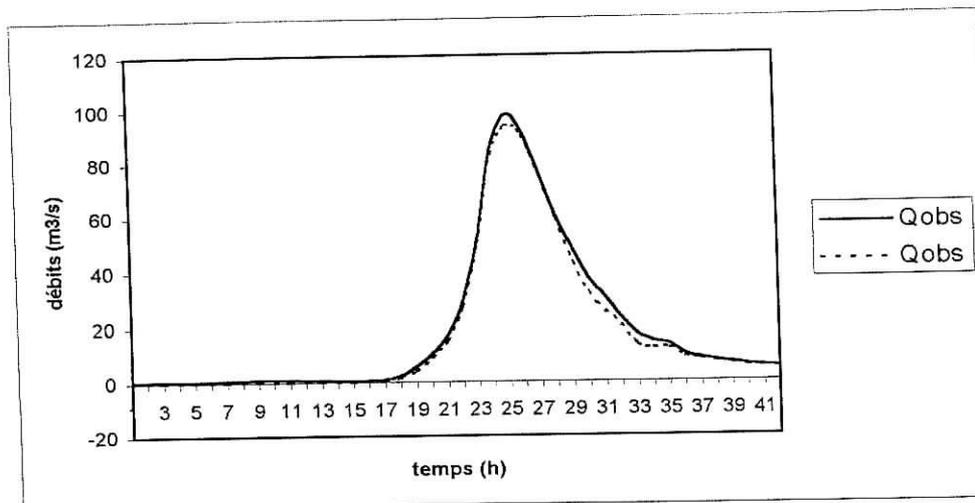


Figure 5-17 : Prévion de la crue n° 6

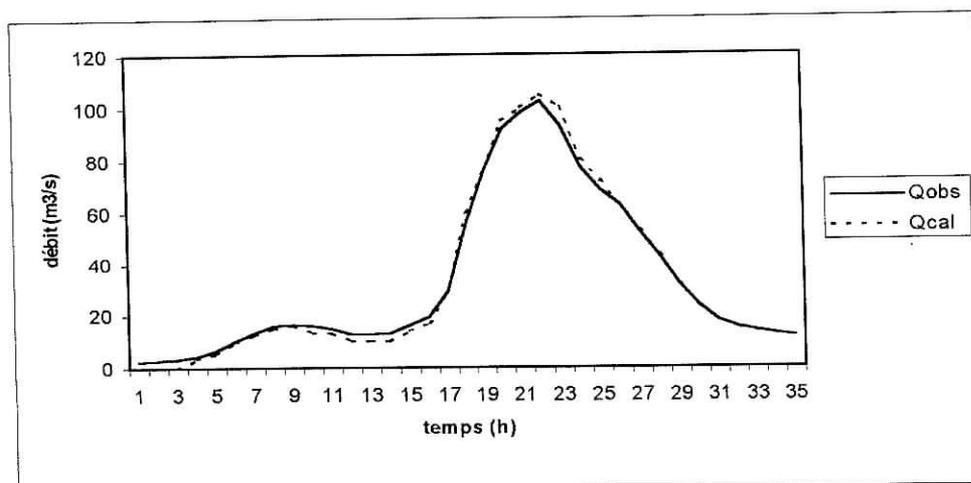


Figure 5-18 : Prévion de la crue n°8

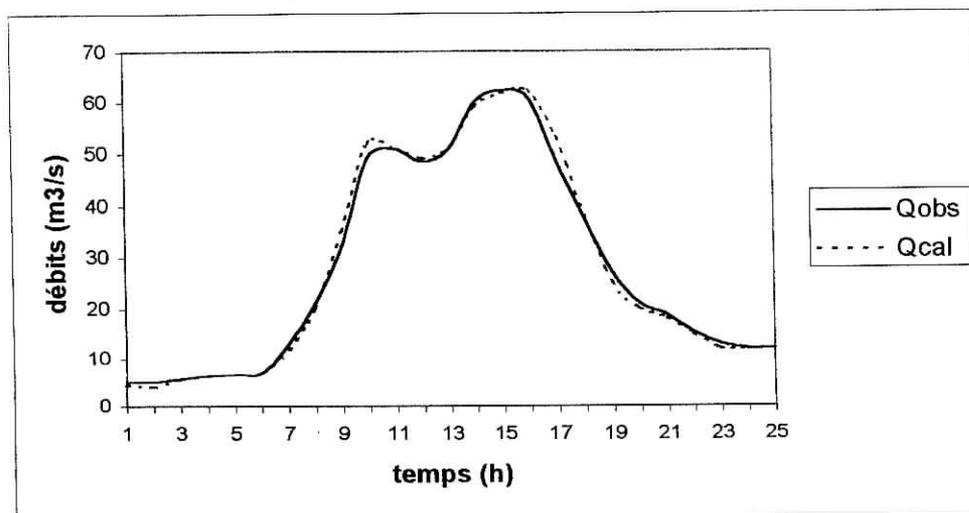


Figure 5-19 : Prévion de la crue n° 9

Le calcul des valeurs des coefficients de corrélation et de Nash permet quant à la précision de la précision réalisée par le modèle.

Tableau 5-8 : Valeurs des coefficients de corrélation et de Nash.

Crue	Crue n°6	Crue n°8	Crue n°9
Coeff. Corrélation	0.99	0.98	0.99
Coeff. NASH	98%	97%	98.75

Une autre approche basée sur les chaînes de MARKOV a été élaborée, du fait que les composantes principales et les résidus sont totalement aléatoires (figure 5-20 et 5-21) Les séries de données ont été subdivisées en classes en fonction de la moyenne et de l'écart type de l'échantillon, le choix de la CP et du résidu ce fait en fonction de la plus forte probabilité d'apparition de l'événement, cette approche s'est avérée non concluante, cela est dû à l'occultation des événement de très faibles probabilités d'apparition qui ont une primauté dans l'étude des événements pluie-débit à des pas de temps très petits (horaire)

Malgré les résultats très satisfaisants, le nombre d'épisodes pluie-débit demeure insuffisant (15), pour une représentativité de toute les situations possibles et réalisables, c'est pour cela qu'un développement futur est nécessaire, il sera tenu compte de la simulation des épisodes pluie-débit par le biais de l'ACP, le modèle de simulation devrait simuler des situations non observés durant l'historique, ce qui permet de apporter une information nouvelle qui sera le prélude d'une prévision remarquable.

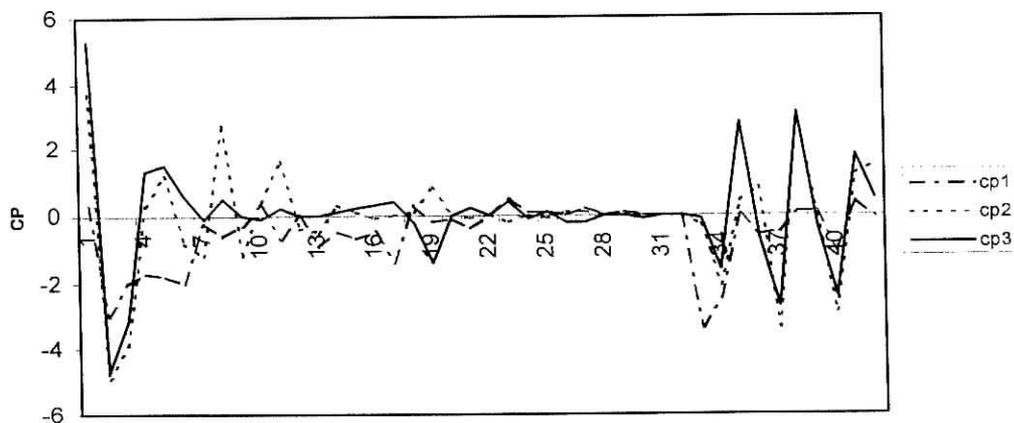


Figure 5-20 : Evolution des CP

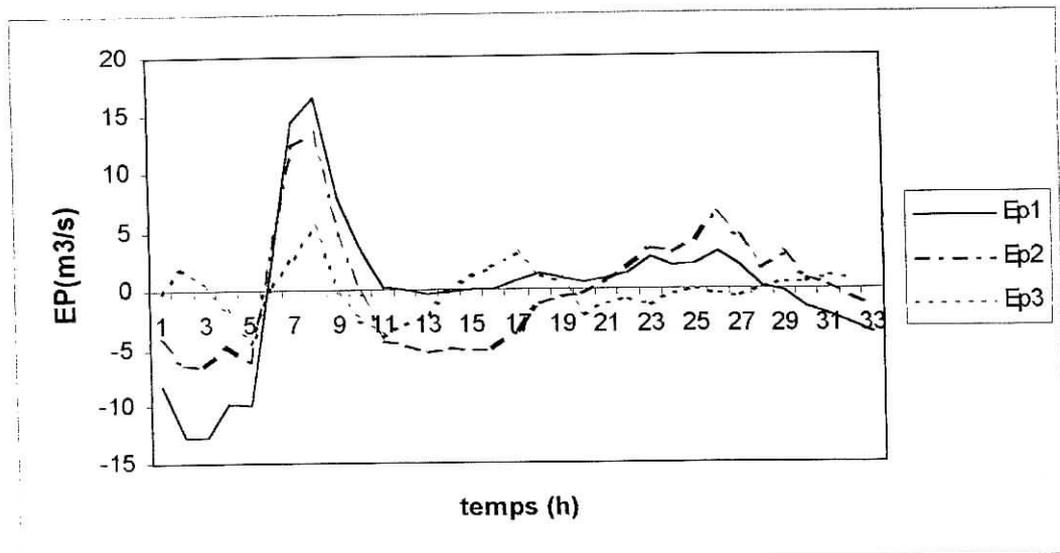


Figure 5-21 : Evolution des résidus

L'approche développée, se basant sur l'analyse en composantes principales, est une tentative purement personnelle qui, à ma connaissance, n'a pas été évoquée auparavant en ce qui concerne la modélisation de la relation pluie-débit au pas de temps horaire. Son utilisation est généralement limitée à la simulation des séries de données ou des prévisions à des pas de temps plus large.

Les résultats obtenus reflètent la capacité prévisionnelle du modèle, néanmoins des développements et des études plus approfondies mériteraient d'être entreprises.

CONCLUSION GENERALE

Dans cette étude, nous avons essayé de trouver une méthode de modélisation robuste afin de contribuer à réaliser un outil de prévision fiable, qui nous permettra de cerner les problèmes liés aux crues, et de minimiser les dégâts que peuvent engendrer.

C'est dans ce contexte que nous avons proposé et étudié trois techniques de modélisation issues de différents domaines.

La première méthode étudiée est la DPFT (différence première de la fonction de transfert) qui repose sur le principe de la transformation des pluies en débits via une fonction de transfert, mais en faisant intervenir les différences des débits et des coefficients de la fonction de transfert, et l'introduction de contraintes pour filtrer les données des pluies. L'application de la méthode sur des épisodes pluies-débits nous a montré qu'elle permet une parfaite reconstitution des crues lorsque celles-ci sont de type simple, c'est à dire les crues dont l'hydrogramme possède un seul pic, par contre les crues complexes résultantes de la succession de plusieurs crues sont très mal reconstituées, et ce sont celles qui engendrent les forts débits.

Quant à la seconde technique de modélisation, qui est issue du domaine neuromimétique, inspiré par le mode de fonctionnement des neurones du cerveau humains, un perceptron à trois couches a été utilisé. Cette configuration est choisie après plusieurs essais, incluant l'augmentation du nombre de neurones dans la couche cachée, et en prenant un neurone dans la couche d'entrée, et un autre pour la couche de sortie, afin d'éviter d'alourdir le modèle, et de rendre plus rapide la convergence du processus itératif, qui permet d'obtenir les poids des interconnexions entre les différentes couches, et avoir ainsi un assez bon calage du modèle.

L'application du modèle réseau de neurones sur les crues tests a démontré une capacité appréciable à reconstituer les débits soit pour les crues de type simple ou complexe.

En ce qui concerne le troisième modèle, il s'agit d'un modèle basé sur l'analyse en composantes principale (généralement utilisé dans le domaine du traitement des données), il permet de simplifier l'approche de la relation pluie-débit. L'interprétation des résultats obtenus nous a permis de conclure de la fiabilité du modèle et de son pouvoir prévisionnel, comparé aux deux premières méthodes, cependant des

améliorations peuvent être apporter pour garantir une meilleure prévision, parmi lesquelles nous citons :

- La simulation de nouvelles données pour augmenter la taille de l'échantillon et les injecter dans le modèle de prévision.
- La possibilité d'introduction de combinaisons avec ce modèle.
- Le travail avec des données brutes qui nécessite l'unification de l'unité de mesure prise en considération (toutes les données doivent être comptabilisées en millimètre), cela permettra d'avoir un modèle à trois paramètres qui devrait être facilement maîtrisable.
- Le développement d'une méthodologie tenant compte d'une prévision simultanée de la pluie et du débit engendré par cette dernière en tenant compte de l'état du sol

Bibliographie

- AFFANE. A , OUGLAOUENE. M ,(1997) : Contribution à l'étude et à la Prévision de la Pluviométrie sur le Littoral Algérien, Thèse d'ingénieur, ENP, Alger.
- BARDOSSY. A, (1990) : Note on fuzzy regression. Fuzzy sets and systems, 37, p 65-75.
- BASTIN. G , (1984) : "Identification de la Relation Pluie - Débit, Prévision en temps réel, Session de formation continue « crue précipitation intenses", Fascicule III, p. 1-30.
- BOX. G.E.P et JENKINS . G.M , (1970) : Time series analysis, forecasting and control, Holden Day, SAN FRANCISCO, p 578 .
- CEMAGREF, (1989) : Recherche sur la mise au point d'un Modèle Journalier Pluie - Débit .
- CHAMBAZ. H , (1990) : Contribution à l'étude de la Méthode DPFT, Thèse de Magister, ENP, Alger.
- DALRYMPLE. T ,(1960) : Flood frequency analysis, US Geol. Surv. Water Supply Pa p. 1543.
- DAVALO . E et NAÏM . P , (1993) : Des réseaux de Neurones, Edition Eyrolles, Paris.
- DONNA RIZZO. M, DAVID DOUGHERTY. E, (1994): Characterization of Aquifer proprieties using Artificial Neural Network, Water Ressources Research, Vol. 30(2), p. 483-497.
- FELLER . W , (1968) : An introduction to probability theory and its applications, Vol. 1, 3^e edition, Willy, 1968.
- FORTIN. V, OUARDA. T.B.M.J, RASMUNSSSEN. P .F et BOBEE. B, (1997) : Méthodes de prévisions des débits, Revue des Sciences de l'Eau, Vol. 4, p. 461- 487.
- 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.

GALEA. G , PRUDHOMME. C, (1997) : Notions de base et concepts utiles pour la compréhension de la modélisation synthétique des régimes de crue des bassins versant au sens des modèles QdF, Revue des Sciences de l'Eau, Vol. 1, p 83-101.

JATON. J. F, (1982) : Contribution à l'étude des Relations Pluie - Débit dans les petits Bassins Versant Ruraux, Thèse de Doctorat, Ecole Polytechnique Fédérale de Lausanne.

JODOUIN. J. F, (1994) : Les réseaux de neurones, Edition Hermès, Paris .

JUMBU. M, (2000) : Méthode de base de l'analyse des données, Edition Eyrolles, Paris p. 155-185.

KOSKO.B , (1992): Neural networks and fuzzy systems : a dynamical systems approach to machine intelligence, Prentis Hall, New Jersey, p 449.

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.

LETTMAIER D. P et WOOD. E. F, (1993) : Hydrologic Forecasting. In Handbook of Hydrology, Edition McGraw-Hill.

MASSON. J. M, (1971) : Comparaison expérimentale des Prévisions de débit obtenus, d'une part au moyen de méthodes de Régression, d'autre part en utilisant un modèle déterministe, la Houille Blanche, Vol. 2.

MASTER. T, (1993) : Practical Neural Network Recipes C++, Academic Press, Inc., USA.

NALBANTIS. I, (1987): Identification des Modèles Pluie - Débit du type Hydrogramme Unitaire, p. 67-100, Thèse de Doctorat, INPG, Grenoble.

NEWTON. D. W et VINYARD. J . W, (1967): Computer-Determined Unit Hydrograph from flows, Journal of Hydraulic Division, ASCE, n°93, Hy5, p. 219-235.

OUARDA . T.B.M.J, LANG . M, BOBEE . B, BERNIER . J et BOIS . P , (1999) : Synthèse de modèles régionaux d'estimation de crues utilisés en France et au Québec. Revue des Sciences de l'Eau, Vol. 7, p.155-182.

RODRIGUEZ. J.Y, (1989) : Modélisation Pluie - Débit par la méthode DPFT, p. 42-65, Thèse de Doctorat, INPG, Grenoble.

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.

SAKAKIMA. S, KOJIRI. T, ITOH. K , (1993) :Real-Time reservoir operation with neural network concept, Computational Mechanics Publications, Southampton, p 502-514.

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

VAZQUEZ. J, ZUG. M, BELLEFLEUR. D, GRANDJEAN. B, SCRIVENE. O (1999) : Utilisation d'un réseau de neurones pour appliquer le modèle de Muskingum aux réseaux d'assainissement, Revue des Sciences de l'eau, Vol.12(3), p 557-595.

VINE. P, GRESILLON. J.M, PUECH. C, (1997) : La télédétection, Proceeding of Rabat Symposium S3 Remote sensing and geographic information systemes for design and operation of water resources systeme, AHS, publ. n° 242, p. 205-217.

WERY. B, (1990) : Identification des Systèmes Hydrologiques : Application à la Prévision de crue, Thèse de Doctorat, Université Catholique de Louvain.

YEDDOU . Y, (1998) : Etude de Synthèse sur les Réseaux de Neurones, Thèse de Magister, ENP, Alger.

Annexes

