

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

COMME EXIGENCE PARTIELLE
À L'OBTENTION DE LA

MAÎTRISE EN GÉNIE ÉLECTRIQUE

M.Ing.

PAR
OMAR CHADID

ANALYSE DES SÉRIES HYDROMÉTÉOROLOGIQUES
PAR LA TRANSFORMÉE EN ONDELETTES

MONTRÉAL, LE 10 SEPTEMBRE 2003

© droits réservés de Omar Chadid

CE MÉMOIRE A ÉTÉ ÉVALUÉ
PAR UN JURY COMPOSÉ DE :

M. Maarouf Saad, professeur et directeur de mémoire
Département de génie électrique à l'École de technologie supérieure

M. Christian Gargour, professeur et codirecteur
Département de génie électrique à l'École de technologie supérieure

M. Marcel Gabrea, professeur et président du jury
Département de génie électrique à l'École de technologie supérieure

M. Saad Bennis, professeur
Département de génie de la construction à l'École de technologie supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 11 juillet 2003

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ANALYSE DES SÉRIES HYDROMÉTÉOROLOGIQUES PAR LA TRANSFORMÉE EN ONDELETTES

Omar Chadid

SOMMAIRE

Dans ce travail, nous avons principalement étudié quelques caractéristiques statistiques annuelles de quelques grands lacs NBS (Net Bassin Supply) situés en Amérique du Nord. Les niveaux de ces grands lacs sont considérés comme des séries chronologiques. Nous avons utilisé trois méthodes de détection de ruptures pour ces séries. La première est basée fondamentalement sur la procédure de segmentation des séries hydrométéorologiques. La deuxième fait appel de la méthode bayésienne de Lee-Heghinian. Nous avons rencontré des difficultés à interpréter et à analyser le vrai comportement de ces séries chronologiques à l'aide de ces deux méthodes. Nous avons donc proposé dans le cadre de ce travail une troisième méthode basée sur l'analyse par ondelettes et nous l'avons nommée méthode de l'ondelette de Haar modifiée (OHM). Elle nous a permis de mieux analyser la présence des ruptures dans ces séries. Les résultats des simulations des trois méthodes sont présentées et comparées entre elles.

ANALYSE DES SÉRIES HYDROMÉTÉOROLOGIQUES PAR LA TRANSFORMÉE EN ONDELETTES

Omar Chadid

ABSTRACT

In this study the statistical characteristics of the annual NBS (Net Bassin Supply) which are located in different regions of North America have been studied. These characteristics series are considered as chronological series. We have used three techniques for the detection of the sudden shifting patterns in these NBS series. The first technique proposed by Hubert is based on segmentation procedure of hydrometeorological series, the second is based on the Bayesian method of Lee-Heghinian. Both of these two techniques show difficulties in interpreting and reproducing the analyzed behavior of these NBS series. We have proposed in this work, a technique which we have named Modified Haar wavelet approach (OHM), this technique gives a better interpretation and analysis for the detection of the sudden shifting patterns in the NBS series. Simulation results based on the use of the three methods are given and compared.

REMERCIEMENTS

Je tiens à remercier mon directeur de recherche, monsieur Maarouf Saad, professeur à l'École de technologie supérieure, pour l'assistance technique et morale qu'il m'a accordée durant ma recherche. Je le remercie également pour ses conseils judicieux, et son aide lors des travaux expérimentaux.

J'aimerais aussi remercier mon codirecteur de recherche, monsieur Christian Gargour, professeur à l'École de technologie supérieure, pour ses directives pertinentes et son aide lors de la recherche et de la révision du mémoire.

Je tiens à remercier profondément l'IJC (International Joint Commission), Madame Laura Fagherazzi, Monsieur Luc Perreault, et les membres du département de la gestion des ressources hydriques d'Hydro-Québec qui ont fourni les bases de données et ont exprimé leur fort intérêt envers les résultats obtenus.

Enfin, j'aimerais remercier mes parents, mes frères, mes sœurs, mes oncles, mes tantes et tous mes amis de près et de loin.

TABLE DES MATIÈRES

	Page
SOMMAIRE	iii
ABSTRACT	iv
REMERCIEMENTS	v
TABLE DES MATIÈRES	vi
LISTE DES FIGURES	ix
LISTE DES ABRÉVIATIONS ET SIGLES	xi
INTRODUCTION	1
CHAPITRE 1 REVUE DE LA LITTERATURE	4
1. 1 Introduction.....	4
1. 2 Problématique.....	4
1.3 Objectifs.....	5
1.4 Revue Bibliographique	7
1.4.1 Définition de la sécheresse.....	8
1.4.2 Identification de la sécheresse.....	11
1.4.3 Quelques classes d'événements hydrologiques.....	13
1.5 Conclusion	15
CHAPITRE 2 MÉTHODES DE DETECTION DE RUPTURES DANS LES SÉRIES CHRONOLOGIQUES	16
2.1 Introduction.....	16
2.2 Méthodes statistiques.....	17
2.3 Test de Pettitt	17
2.4 Méthode bayésienne de Lee et Heghinian	19
2.5 Procédure de segmentation de Pierre Hubert.....	22
2.6 Statistique de Buishand.....	25
2.7 Test de corrélation sur le rang	27
2.8 Conclusion	28

CHAPITRE 3 TRANSFORMÉE EN ONDELETTES.....	29
3.1 Introduction.....	29
3.2 Transformée de Fourier	29
3.3 Transformée de Fourier à fenêtre glissante	30
3.4 Transformée en ondelettes.....	30
3.5 Représentation temps-fréquence et temps-échelle.....	32
3.6 Analyse multirésolution[27].....	35
3.7 Interpolation en bancs de filtres.....	38
3.7.1 Algorithme de Mallat	39
3.7.2 Propriétés des filtres h et g	43
3.8 Les ondelettes de Haar.....	44
3.9 Conclusion	47
CHAPITRE 4 LA MÉTHODE DE L'ONDELETTE DE HAAR MODIFIÉE	48
4.1 Introduction.....	48
4.2 Décomposition en ondelettes de Haar du signal initial traité et de ses translatés...	49
4.2.1 Calcul de W_1	53
4.2.2 Calcul de W_2	54
4.2.3 Calcul de W_3	54
4.2.4 Calcul de W_4	55
4.3 Calcul “ des entropies apparentes” des coefficients d’approximation et de détails de l’ondelette de Haar approchée	57
4.4 Application numérique de la méthode proposée	62
4.4.1 Calcul numérique de W_1	63
4.4.2 Calcul numérique de W_2	63
4.4.3 Calcul numérique de W_3	63
4.4.4 Calcul numérique de W_4	64
4.4.5 Calcul numérique des entropies apparentes.....	64

4.5	Comparaison de l'utilisation de l'ondelette et de celle de l'ondelette de Haar	
	modifiée (OHM) pour la détection d'une rupture	66
4.6	Conclusion	69
	CHAPITRE 5 EXPERIMENTATIONS ET RESULTATS	71
5.1	Introduction.....	71
5.2	Bases de données	71
5.3	Application de quelques méthodes de détection de rupture à quelques séries chronologiques	72
5.3.1	Lac Ontario.....	72
5.3.2	Lac Erie	74
5.3.3	Lac Michigan.....	75
5.3.4	Lac St-Clair	76
5.3.5	Lac Supérieur	77
5.4	Conclusion	78
	CONCLUSION.....	80
	BIBLIOGRAPHIE.....	81

LISTE DES FIGURES

	Page	
Figure 1	Illustration des grands lacs situés en Amérique du Nord [35]	2
Figure 2	Représentation des variables aléatoires caractérisant la sécheresse.....	12
Figure 3	Représentation de la série S_t en fonction de l'offre et de la demande	12
Figure 4	Identification des sécheresses par la méthode OHM	13
Figure 5	Quatre classes d'événements hydrologiques définis par Dracup [15]	14
Figure 6	La statistique de Mann-Whitney.....	19
Figure 7	Quelques formes d'ondelettes.....	31
Figure 8	Pavage temps-fréquence pour la transformée de Gabor.	34
Figure 9	Représentation temps échelle de la transformée en ondelettes.....	35
Figure 10	Décimation du signal x par facteur N	40
Figure 11	Échantillons du signal original $x(n)$	41
Figure 12	Échantillons du signal $y(n) = x(2.n)$	41
Figure 13	Schéma équivalent de l'algorithme de Mallat.....	42
Figure 14	Structure de banc de filtres à 3 niveaux d'échelle.	42
Figure 15	Fonction d'échelle de l'ondelette de Haar	45
Figure 16	Ondelette de Haar.....	46
Figure 17	Décomposition en ondelettes du signal S_1 en 4 niveaux de résolution.....	49
Figure 18	Les premières huit fonctions de Haar [23, 20] , $a = 1/\sqrt{2}$	51
Figure 19	La décomposition en ondelette de Haar.....	66
Figure 20	La décomposition en ondelette de Haar pour un signal aléatoire	67
Figure 21	a) Signal aléatoire à 3 états b) Entropie apparente du détail D4 de l'OHM c) Entropie apparente normalisée du détail D4 de l'OHM	68
Figure 22	Entropie apparente de l'approximation A3 de l'OHM	69
Figure 23	Entropie apparente du détail D3 de l'OHM	69
Figure 24	Méthodes de Hubert et de Lee-Heghinian appliquées au Lac Ontario	72

Figure 25	Méthode de l'ondelette de Haar modifiée appliquée au Lac Ontario	73
Figure 26	Méthodes de Hubert et de Lee-Heghinian appliquées au Lac Erie.....	74
Figure 27	Méthode de l'ondelette de Haar modifiée appliquée au Lac Erie.....	74
Figure 28	Méthodes de Hubert et de Lee-Heghinian appliquées au Lac Michigan	75
Figure 29	Méthode de l'ondelette de Haar modifiée appliquée au Lac Michigan	76
Figure 30	Méthodes de Hubert et de Lee-Heghinian appliquées au Lac St-Clair.....	76
Figure 31	Méthode de l'ondelette de Haar modifiée appliquée au Lac St-Clair.....	77
Figure 32	Méthode de Hubert et de Lee-Heghinian appliquées au Lac Supérieur	77
Figure 33	Méthode de l'ondelette de Haar modifiée appliquée au Lac Supérieur.....	78

LISTE DES ABRÉVIATIONS ET SIGLES

a	Facteur d'échelle.
b	Facteur de translation.
AMR	Analyse multirésolution
G_a	Transformée de Gabor.
NBS	Net Bassin Supply
OHM	Ondelette de Haar modifiée
R_t	Résolution temporelle
R_f	Résolution fréquentielle
TF	Transformée de Fourier.
TO	Transformée en ondelettes
a_j	Coefficients d'approximation à l'échelle j
d_j	Coefficients du détail à l'échelle j
\hat{EW}_{aj}	Entropie apparente de l'approximation à l'échelle j .
\hat{EW}_{dj}	Entropie apparente du détail à l'échelle j .
V_j	Espace d'approximation à l'échelle j
W_j	Espace de détail à l'échelle j
Δt	Résolution temporelle
$\Delta \omega$	Résolution fréquentielle
φ	Fonction d'échelle
ψ	Fonction d'ondelette
Φ	Transformée de Fourier de φ
Ψ	Transformée de Fourier de ψ
φ_j	Fonction d'échelle au niveau d'échelle j
ψ_j	Fonction d'ondelette au niveau d'échelle j

INTRODUCTION

La croissance continue de la population mondiale et l'augmentation du niveau de vie dans certaines parties de la planète constituent des contraintes lourdes et pesantes sur la gestion des ressources hydriques. En effet, ces deux phénomènes influencent la demande qualitative et quantitative des ressources hydriques, nécessitant ainsi une gestion adéquate basée fondamentalement sur des techniques de modélisation faisant appel à des méthodes d'analyses stochastiques et statistiques. La connaissance a priori des propriétés statistiques des sécheresses est essentielle pour la conception et la réglementation des politiques de gestion et de contrôle des ressources hydriques.

Afin d'assurer une gestion efficace de ces ressources hydriques, il est essentiel d'avoir en main des outils de modélisation statistiques des sécheresses. Ces outils devraient être en mesure de s'adapter aux séries hydrologiques de caractères non stationnaires. D'où la nécessité d'utiliser de nouveaux outils mathématiques basés sur l'analyse multirésolution, tel que la transformée en ondelettes.

Plusieurs approches d'analyses spectrale et statistique paramétrique et nonparamétrique sont utilisées pour analyser les séries hydrologiques représentant les grands lacs NBS Net Bassin Supply définis tel que:

$$\text{NBS} = P + R - E$$

Le paramètre P désigne la précipitation directement sur le lac, R représente le flux des autres bassins vers le lac et E caractérise l'évaporation de la surface du lac. La figure 1 ci-dessous représente une illustration des grands lacs situés en Amérique du Nord.

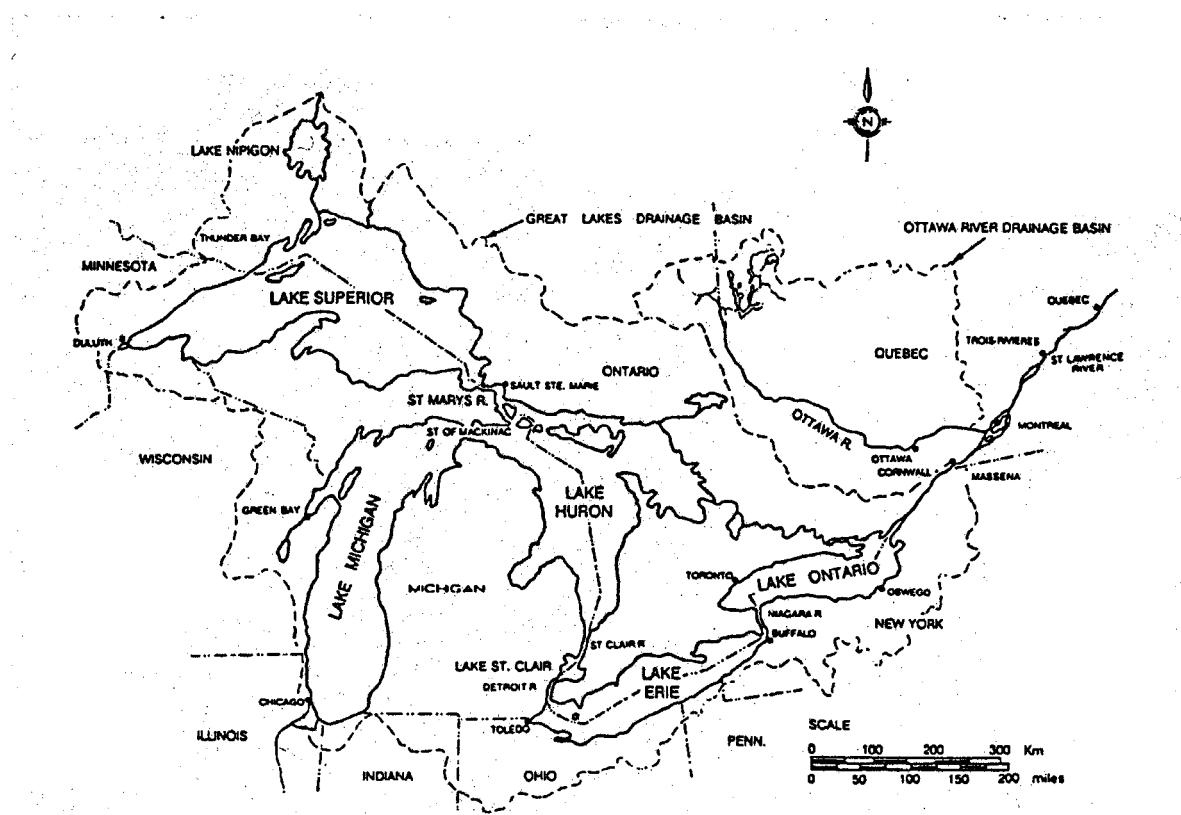


Figure 1 Illustration des grands lacs situés en Amérique du Nord [35]

Ces méthodes permettent de détecter les ruptures dans les séries chronologiques, elles sont très intéressantes du fait qu'elles nous donnent la possibilité de détecter et d'analyser les changements des moyennes dans les séries chronologiques. Dans le cas où l'on a plusieurs changements qui se présentent dans la série, il est nécessaire d'adopter des techniques qui permettent d'analyser séparément différentes sous séries du signal original.

Dans ce travail, nous étudierons trois méthodes de détection de rupture dans quelques séries chronologiques NBS. La première est basée sur la méthode de segmentation de Hubert [21], la deuxième est la méthode bayésienne de Lee-Heghinian [22]. Nous avons développé la troisième, et que nous avons nommée méthode de Haar modifiée (OHM).

Quoique la méthode de Lee-Heghinian ait été largement utilisée dans ce domaine, elle ne permet d'identifier qu'un seul parmi plusieurs changements de moyennes dans les séries chronologiques, d'où la nécessité d'utiliser une autre technique qui permet de détecter plusieurs changements dans une même série. La méthode de Hubert présente elle aussi un inconvénient du fait qu'elle découpe parfois la série chronologique en un nombre très important de segments. Ceci entraîne une difficulté considérable pour identifier le vrai changement de moyennes dans la série. Nous verrons que notre méthode de Haar modifiée permet de déterminer exactement la rupture réelle dans la série chronologique, de plus, elle donne une information supplémentaire sur la rupture. En effet elle présente un avantage remarquable du fait qu'elle permet de classifier les ruptures dans la série chronologique en fonction de leurs ordre d'occurrence, et qu'elle permet de détecter les ruptures à des niveaux de résolution plus fins.

Dans le premier chapitre nous présenterons une brève revue de la littérature et donnerons les définitions les plus importantes de la sécheresse que l'on y retrouve. Le second chapitre présente quelques méthodes de détection de ruptures dans les séries chronologiques. Le troisième chapitre décrit quelques notions de la transformée en ondelettes et de l'analyse multiresolution qui lui est associée. Le quatrième chapitre présente d'une façon détaillée notre méthode de Haar modifiée (OHM). Finalement le chapitre 5 correspond aux expérimentations, résultats et comparaisons.

CHAPITRE 1

REVUE DE LA LITTERATURE

1. 1 Introduction

Nous commencerons dans ce chapitre par présenter les principales problématiques qui ont été soulevées dans le domaine de la gestion des ressources hydriques. Les objectifs essentiels à atteindre dans le cadre de ce sujet seront exposés en détails. Ils portent fondamentalement sur le développement et la mise en œuvre d'un nouveau modèle d'identification et de modélisation des changements du niveau d'eau. Par la suite nous présenterons une brève revue bibliographique. Nous y énumérerons les références les plus pertinentes dans ce domaine d'étude.

1. 2 Problématique

Parmi les problèmes les plus souvent rencontrés dans le domaine de la gestion des ressources hydriques on trouve la croissance permanente de la population mondiale et l'augmentation du niveau de vie. Ces deux phénomènes constituent une contrainte considérable sur la demande qualitative et quantitative de la ressource hydrique (eau), nécessitant par conséquent une gestion basée principalement sur des études de modélisation faisant usage d'outils mathématiques (stochastiques et statistiques).

Une bonne gestion des ressources hydriques nécessite souvent des méthodes de modélisation statistique et de simulation des sécheresses et des étiages (faibles débits). Ces méthodes permettent de s'adapter avec les séries chronologiques malgré leurs caractères non stationnaires. L'étude a porté sur les propriétés statistiques des sécheresses mène à la réglementation des politiques de gestion et de contrôle des

ressources hydriques. L'information utile et pertinente concernant les faibles débits est souvent utilisée pour :

- déterminer les débits réservés.
- planifier les débits de régulation.
- analyser les impacts environnementaux et économiques.
- modéliser la qualité.
- améliorer le niveau général de compréhension des systèmes naturels et régularisés.

1.3 Objectifs

Les principales caractéristiques statistiques des séries temporelles annuelles et mensuelles de variables climatiques des grands lacs (température et précipitation) ont été largement étudiées dans les travaux de recherches antérieurs. Quelques modèles stochastiques multivariables ont été préparés synthétiquement pour produire la série de flux NBS (Net Bassin Supply). La plupart des études ont montré, par un certain nombre de méthodes et de techniques différentes que quelques séries exhibent certaines formes de décalages dans les moyennes qui sont souvent difficiles à interpréter et à reproduire en utilisant le modèle stochastique.

L'objectif à atteindre dans le cadre de ce sujet de recherche est de concevoir, de développer et de mettre en œuvre un nouveau modèle de modélisation permettant d'identifier, de modéliser, et de mettre à jour des changements du niveau d'eau. Les principaux objectifs à atteindre sont énumérés comme ce qui suit :

- l'extension des changements du modèle de niveau de décalage multivariable stochastique développé en 1991 par une équipe d'Hydro-Québec pour les grands

lacs afin d'inclure la période de bas niveau (sécheresse) dans la partie temporelle du modèle et d'inclure la rivière Ottawa, et autres affluents.

- l'analyse du niveau des rivières en tenant compte de la température.
- la modélisation du système en fonction des variables qui sont liées aux inondations et aux sécheresses.
- un autre objectif supplémentaire et collatéral de cette étude consiste à passer en revue la littérature liée aux étiages et à la caractérisation de la sécheresse. Ces caractéristiques pour des flux hauts et bas seront utilisés plus tard comme un outil de comparaison pour choisir le modèle de génération des séries chronologiques de données hydrométéorologiques le plus adéquat.

Le principal but à atteindre dans ce travail de recherche est l'évaluation des processus pour modéliser les durées des étiages et des surplus tout en tenant compte de la non-stationnarité du phénomène, ceci dans le but de caractériser et de quantifier les risques qui en découlent. Dans le cadre de ce travail, un étiage est interprété comme étant un déficit défini par un seuil selon El-Jabi et al. [16]. Un cours d'eau peut par conséquent être considéré comme étant un système à deux états ; le premier état est celui qui correspond à l'état satisfaisant (c'est le cas de surplus), et le deuxième représente l'état non-satisfaisant (c'est le cas du déficit et d'étiage). L'incident indésirable qui correspond au passage à l'état de déficit se produit généralement lorsque le débit descend ou décroît en dessous d'un seuil critique donné et prédéterminé.

L'étude des étiages est divisée en deux parties principales. La première peut être abordée comme étant un événement aléatoire dont on cherche à connaître la loi de probabilité. La deuxième est caractérisée par un événement naturel dont on cherche à connaître le processus physique. Dans ce qui suit trois étapes principales seront identifiées et énumérées comme suit :

- la modélisation, des séries de données hydrométéorologiques des grands lacs (données de débits, précipitations, et températures) et des durées des surplus et étiages.
- la construction et l'évaluation de ces modèles sur différentes rivières.
- la construction d'une nouvelle méthode statistique basée sur la transformée en ondelettes (Haar).

1.4 Revue Bibliographique

Dans la littérature, il existe une multitude d'études très importantes faisant appel à des méthodes qui traitent le comportement des divers phénomènes naturels. Cependant il y a peu d'études qui traitent du cas des étiages et des sécheresses selon Abi-zeid [1]. Dans ce qui suit, nous présenterons la différence qui existe entre la sécheresse et l'étiage, car il est primordial et essentiel de faire une distinction entre les deux termes. Par la suite nous donnerons la définition de chacun d'entre eux.

La sécheresse est considérée comme étant un phénomène général, elle peut être caractérisée par plusieurs types de facteurs. Dans certains cas, les sécheresses incluent des périodes de débit faible. Toutefois, il existe des cas où par exemple un événement saisonnier (continu) de débit faible ne constitue pas obligatoirement ou nécessairement une période de sécheresse. Notons de même que selon quelques chercheurs en l'occurrence Zelenhasic et Salvai. [41], Clauser et Pearson. [11], et Tallaksen et al. [38], ces derniers se réfèrent à une période d'une année continue de débit faible comme étant une période de sécheresse annuelle.

Un étiage peut être considéré comme étant une conséquence d'une sécheresse hydrologique. En générale un étiage peut avoir une durée pouvant varier d'une période

de quelques journées ou bien de quelques semaines. Selon Abi-zeid et Bobée [2], un étiage est considéré comme étant le plus bas niveau d'un cours d'eau.

Dans les prochains paragraphes, nous présenterons un grand éventail d'informations concernant la définition de la sécheresse et son identification. Nous présenterons ainsi quelques différences entre la définition de la sécheresse et quelques phénomènes météorologiques naturels. Par la suite cette définition sera présentée en fonction des domaines d'étude, en l'occurrence le domaine de la météorologie, de l'hydrologie, et d'autres.

1.4.1 Définition de la sécheresse

La définition de la sécheresse diffère dans la plupart des cas selon les phénomènes météorologiques et naturels considérés. Dépendamment de plusieurs points de vue, nous pouvons citer quelques définitions de la sécheresse, telles qu'énumérées ci-dessous :

- la sécheresse frappe lentement et progressivement une région, par conséquent ses principaux effets peuvent s'accumuler pendant une longue période de temps, et nous pouvons imaginer que pour les cas les plus extrêmes, elle peut éventuellement durer plusieurs années.
- souvent, les impacts de la sécheresse sont moins spectaculaires que les dommages en résultants, et ceci comparativement aux autres sources de risques naturels ou météorologiques, nous pouvons citer l'exemple des inondations d'après Bootsma et al. [6].
- la définition de la sécheresse est généralement relative à la région considérée. Ce qui fait que cette définition dépend d'un ou de plusieurs aspects régionaux. Ceci peut être expliqué par le fait qu'elle est basée sur ce que l'on appelle la

pluviométrie moyenne d'une région à un moment donné de l'année. Donc, selon Abi-zeid [1] Abi-zeid et Bobée [2], il n'existe pas de définition précise de la sécheresse.

Géographiquement parlons, les définitions précises de la sécheresse varient et diffèrent généralement d'un pays à l'autre. Nous pouvons citer quelques exemples ; en l'occurrence le terme sécheresse aux États-Unis est utilisé lorsqu'une zone étendue reçoit un pourcentage de 30% ou moins de précipitation, qu'en temps normal sur un minimum de vingt et un jours. En Inde, la sécheresse est déclarée ou annoncée quand les précipitations annuelles sont inférieures à un pourcentage de 75% aux normales saisonnières. Le dernier exemple que nous citerons est celui de l'Australie. Pour ce pays, on parle de sécheresse lorsqu'une région reçoit un pourcentage de 10% ou moins de précipitation par rapport à la moyenne annuelle, et ceci selon Bootsma et al. [6].

- le mot sécheresse est généralement lié à un concept qui touche plusieurs domaines, et par conséquent on estime que la définition de la sécheresse varie selon le secteur de l'économie concerné.
- dans le domaine de la météorologie, la sécheresse se caractérise le plus souvent par plusieurs phénomènes météorologiques, on peut citer le cas d'une absence prolongée, ou d'un déficit marqué ou encore le cas d'une faible distribution des précipitations par rapport à la normale.
- dans le secteur d'hydrologie, la sécheresse survient quand les précipitations sont en dessous de la moyenne requise. Ce phénomène peut se traduire soit par le fait que le niveau d'approvisionnement des cours d'eau soit bas par rapport au seuil requis, soit par des réservoirs de surface ou encore soit par les souterrains.

- en agriculture, la sécheresse est définie comme un déficit marqué et soutenu des précipitations qui réduit considérablement les productions agricoles par rapport à la normale [5].
- la sécheresse socio-économique [30] quant à elle apparaît lorsque les précipitations insuffisantes ont un impact observable et significatif sur les communautés et leur économie.

Les définitions correspondantes à des événements de sécheresse sont souvent reliées à la climatologie ainsi qu'à la géologie de l'endroit considéré. La sévérité de la sécheresse est difficile à déterminer, car elle dépend de la durée, de l'intensité, de l'étendue géographique ainsi que de la dépendance des activités humaines des ressources disponibles de la région.

En général, on peut distinguer les sécheresses entre elles selon quatre paramètres que nous citons ci-dessous :

- l'origine
- la durée
- l'intensité
- la zone géographique.

Selon ces différents paramètres, plusieurs définitions spécifiques ont été suggérées dans le but d'identifier et de classifier les sécheresses. On peut donner quelques exemples de définitions basées sur :

- les conditions hydrologiques et climatologiques [4].
- l'activité économique [1].

En résumé, on parle de sécheresse météorologique lorsque les précipitations sont sous la normale pour une période prolongée. La sécheresse hydrologique est reliée à une période prolongée de bas débits, de bas niveaux des eaux de surface ou des eaux souterraines. Pour les cas sociaux et économiques on tient compte de l'offre et de la demande et de l'impact de la sécheresse sur l'activité économique.

1.4.2 Identification de la sécheresse

Dans cette section, nous présenterons deux approches souvent utilisées dans la littérature pour identifier une sécheresse. La première a été introduite par Yevjevich [40], son approche est reliée fondamentalement à la théorie des séquences, qui se base essentiellement et particulièrement sur l'analyse statistique des séries chronologiques. Pour être plus explicite, nous désignerons par S_t une série temporelle d'une variable aléatoire qui peut définir éventuellement la sécheresse, tel que par exemple la précipitation, l'humidité du sol, etc.

Notons de même que les principales caractéristiques mesurables des étiages et des sécheresses sont le volume et la durée des déficits.

Dans la première approche mentionnée précédemment, on étudie souvent les variables aléatoires décrivant la sécheresse, par un seuil que l'on désigne par X_0 , nous désignerons aussi par n la durée du déficit et par Vn le volume de celui-ci. Selon El-jabi [16], ce seuil X_0 est généralement choisi en dépit des besoins imposés et en fonction de la réalité physique des disciplines comme l'exemple d'agriculture, d'hydroélectricité, etc. Par exemple, pour un intervalle de temps borné entre $t=0$ et $t=T$, on cherche à sélectionner les débits inférieurs au seuil X_0 . La durée du déficit n - en prenant comme unité de temps le jour - représente le nombre de jours consécutifs pour lesquels le débit déficitaire est en dessous du seuil X_0 . La figure 2 ci-dessous illustre un exemple qui permet de représenter la définition des variables aléatoires caractérisants la sécheresse.

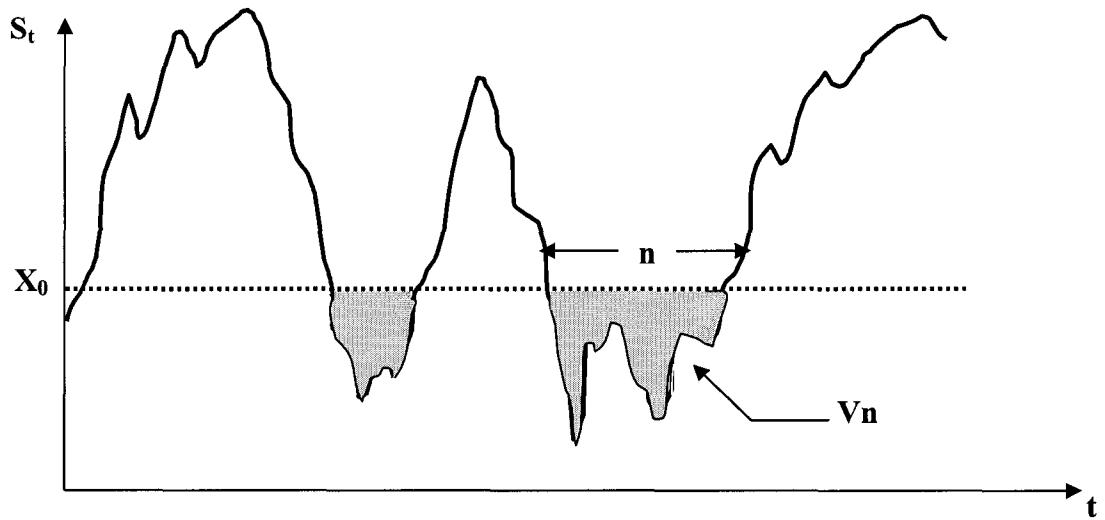


Figure 2 Représentation des variables aléatoires caractérisant la sécheresse

Dans la deuxième approche considérée, Yevjevich [40] suggère qu'au lieu d'utiliser le seuil X_0 prédéterminé comme étant un repère permettant de définir les déficits et les surplus, on considère cette fois ci la série S_t comme la différence entre la série temporelle de l'offre et celle de la demande tel qu'indiqué dans la figure 3. selon que S_t est positive on est en présence d'un surplus, par contre lorsque S_t est négative on a dans ce cas un déficit.

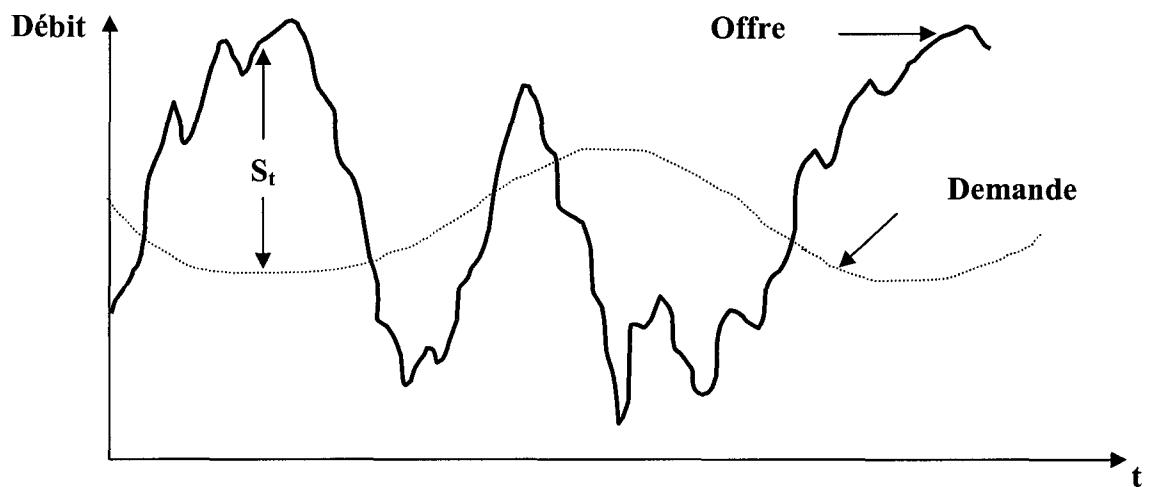


Figure 3 Représentation de la série S_t en fonction de l'offre et de la demande

Dans ce travail, nous identifions les périodes de sécheresses dans les séries NBS par la méthode OHM en déterminant les intervalles de temps qui dépassent un an dans laquelle l'entropie apparente $\hat{E}W_D$ est suffisamment négative dépendamment au seuil fixé par le gestionnaire. Nous présentons sur la figure 4 un exemple d'identification des périodes de sécheresse que nous pouvons extraire de la série chronologique qui correspond au débit du lac Ontario par la méthode OHM. Nous pouvons ainsi localiser pour cet exemple la première sécheresse qui se situe entre l'année 1928 et 1934 et la deuxième entre 1951 et 1960.

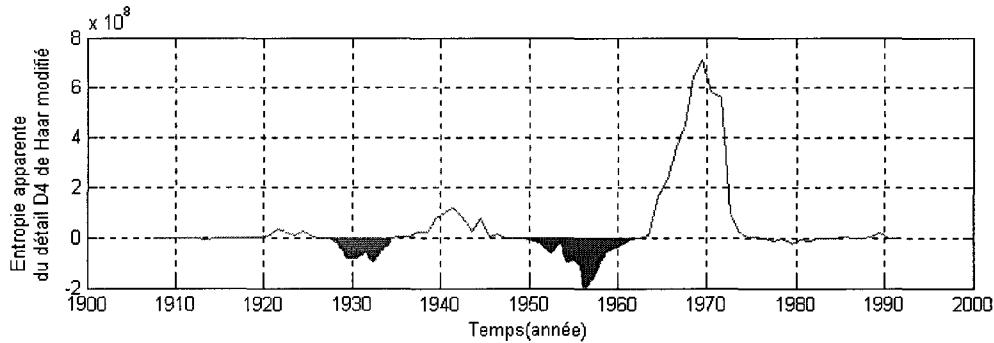


Figure 4 Identification des sécheresses par la méthode OHM

1.4.3 Quelques classes d'événements hydrologiques

Dans le but de déterminer le type de sécheresse analysé, Dracup [15] a proposé des suggestions visant à considérer quelques réponses sur la corrélation, la durée, et la sévérité, etc. Ses études ont été accentuées pour répondre aux questions suivantes :

- L'intérêt principal réside-t-il dans l'analyse des débits (sécheresse hydrologique) ou dans les précipitations (sécheresse météorologique) ou encore dans l'humidité du sol (sécheresse agricole) ?
- Quelle unité de temps doit-on adopter ?
- Quelle technique faudrait-il adopter pour faire la distinction entre les sécheresses et les autres événements de la série chronologique ?

d) Comment peut-on localiser la sécheresse dans son aspect régional ?

Un déficit est souvent lié à la précipitation, à l'humidité du sol, au débit, et à leurs effets combinés. En ce qui concerne l'unité de temps, son choix est déterminé dans la plupart des cas en fonctions des objectifs de l'étude abordée, et de la taille des données historiques disponibles. L'unité de temps annuelle est utilisée dans la majorité des études de cas. Le seuil de dépassement est un paramètre essentiel pour la détermination de la différence qui existe entre les sécheresses et les autres événements de la série chronologique. Son choix dépend généralement de la tendance centrale de la série étudiée. Il est souvent défini comme étant une moyenne ou encore une médiane. Dans certains cas, ce seuil représente le maximum des minima annuels ou selon Ozga-Zielinska [32], un critère économique d'offre et de demande. Dracup [15] a défini quatre classes d'événements hydrologiques en combinant le seuil de dépassement et l'unité de temps, comme illustré sur la figure 5. Finalement, l'aspect régional constitue un point important dans la localisation des sécheresses. Cela est dû au fait que ces dernières sont fondamentalement un phénomène régional.

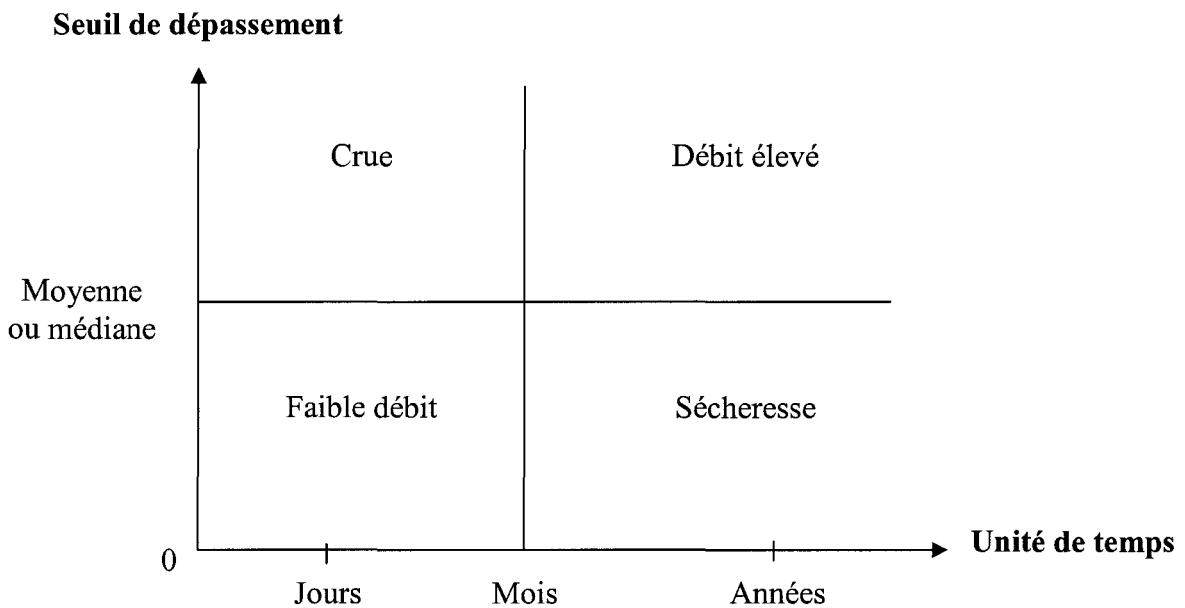


Figure 5 Quatre classes d'événements hydrologiques définis par Dracup [15]

1.5 Conclusion

Nous avons exposé au début de ce chapitre les principales problématiques qui ont été soulevées dans le domaine de la gestion des ressources hydriques. Par la suite nous avons présenté en détails les objectifs fondamentaux et essentiels à atteindre dans le cadre de ce sujet. Ces objectifs ont été principalement focalisés vers : a) la modélisation, des séries de données hydrométéorologiques des grands lacs (données de débits, précipitations, et températures) et des durées des surplus et étiages, b) la construction et l'évaluation de ces modèles sur différentes rivières, et c) la construction d'une nouvelle méthode statistique basée sur la transformée en ondelettes. Cette méthode devrait être en mesure de permettre d'effectuer l'identification et la modélisation des changements du niveau d'eau. De même nous avons présenté un aperçu de la revue bibliographique dans laquelle nous avons cité et énuméré les références bibliographiques les plus pertinentes dans ce domaine d'études. Nous y avons aussi présenté les notions de base de la définition et de l'identification de la sécheresse. A la fin de ce chapitre nous avons présenté Quatre types de classes d'événements hydrologiques définis par Dracup [15].

Dans le prochain chapitre, nous présenterons quelques méthodes statistiques et stochastiques souvent utilisées dans le domaine d'hydrologie. Ces méthodes ont pour but d'identifier, de modéliser, et de mettre à jour les changements du niveau d'eau.

CHAPITRE 2

MÉTHODES DE DETECTION DE RUPTURES DANS LES SÉRIES CHRONOLOGIQUES

2.1 Introduction

Dans ce chapitre, nous présenterons quelques méthodes de détection de rupture souvent utilisées dans les séries chronologiques. Ces méthodes permettent principalement d'effectuer l'identification et la modélisation des changements des paramètres statistiques et stochastiques liés à la série chronologique. Ces paramètres statistiques sont éventuellement caractérisés dans notre cas par le changement des moyennes d'une série temporelle. En premier lieu, nous commencerons par présenter la méthode de test de Pettitt [33]. En second lieu nous présenterons un aperçu sur la méthode bayésienne de Lee et Heghinian [22], qui elle, n'est pas considérée comme étant une méthode de test statistique classique, mais néanmoins demeure une méthode qui est basée sur une approche paramétrique permettant ainsi de confirmer ou d'infirmer l'éventualité d'un changement de moyenne dans la série chronologique. La troisième méthode que nous exposerons dans ce chapitre est celle qui correspond à la méthode basée sur la procédure de segmentation des séries hydrométéorologique [21]. Elle ne constitue absolument pas un test statistique, cependant elle reste adaptée à la recherche des multiples changements de moyenne dans la série. Par la suite nous présenterons la quatrième méthode qui est en l'occurrence la méthode de statistique de Buishand [8], cette méthode est généralement utilisée pour effectuer un test bilatéral de rupture en moyenne à un instant inconnu. Finalement, la cinquième méthode que nous présenterons est celle qui correspond à l'approche basée sur le test de corrélation sur le rang [31].

2.2 Méthodes statistiques

Dans une série temporelle non-stationnaire, la modification brusque est caractérisée par un changement de la distribution de probabilité de la variable aléatoire. Les méthodes présentées ci-dessous permettent d'identifier la modification des paramètres statistiques et particulièrement le changement de moyennes d'une série temporelle.

À l'exception de l'approche de Pettitt [33] et de la segmentation de Hubert [21], ces méthodes supposent une absence de modification de la variance de la série étudiée. En outre, elles ne sont généralement pas adaptées à la recherche de plusieurs ruptures dans une même série [24].

Dans la suite de ce texte $x_i (i = 1, \dots, N)$ désigne la série chronologique étudiée.

2.3 Test de Pettitt

Nous nous basons ici sur l'article de Pettitt [33] pour décrire le test de celui-ci. Ce dernier a considéré les séries x_1, x_2, \dots, x_N comme étant une séquence de variables aléatoires indépendantes.

La séquence x_1, x_2, \dots, x_N est supposée contenir un point de rupture à l'instant τ et ceci en ayant la condition que les séries $x_{t=1, \dots, \tau}$ aillent une distribution commune $F_1(x)$, et aussi que les séries $x_{t=\tau+1, \dots, N}$ aillent une distribution commune $F_2(x)$ tel que :

$$F_1(x) \neq F_2(x).$$

L'hypothèse nulle qui correspond au cas de non-rupture est désignée par H_0 telle que $\tau = N$, contrairement à cette hypothèse, c'est à dire celle qui correspond à l'alternative de rupture est désignée par H_1 telle que $1 \leq \tau < N$.

Le test de Pettitt est considérée comme un test statistique non paramétrique. Pour lui, on estime qu'aucune condition particulière n'est obligatoire pour les formes fonctionnelles de F_1 et F_2 .

Pettitt a prouvé efficacement comment une formulation appropriée du test de Mann-Whitney [24,33] peut être utilisée pour effectuer le test des hypothèses de non-rupture H_0 et de rupture H_1 .

Nous pouvons expliquer ce traitement par l'organigramme représenté par la figure 6, qui est explicatif et qui décrit la statistique de Mann-Whitney pour tester l'appartenance des deux échantillons x_1, \dots, x_t et x_{t+1}, \dots, x_N à la même population.

La statistique $U_{t,N}$ est considérée pour les valeurs de t bornée entre 1 et N . Pour le test de l'hypothèse H_0 contre celle H_1 , Pettitt a proposé d'utiliser la variable définie par :

$$K_N = \max_t |U_{t,N}| \quad (2.1)$$

En utilisant la théorie des rangs, Pettitt a donné la probabilité de dépassement approximative d'une valeur k par :

$$\Pr ob(K_N > k) \approx 2 \exp\left(\frac{-6k^2}{(N^3 + N^2)}\right) \quad (2.2)$$

Pour un risque α de première espèce donnée, H_0 est rejetée si cette probabilité est inférieur α . Dans ce cas, la série présente une rupture du temps $t = \tau$ définissant.

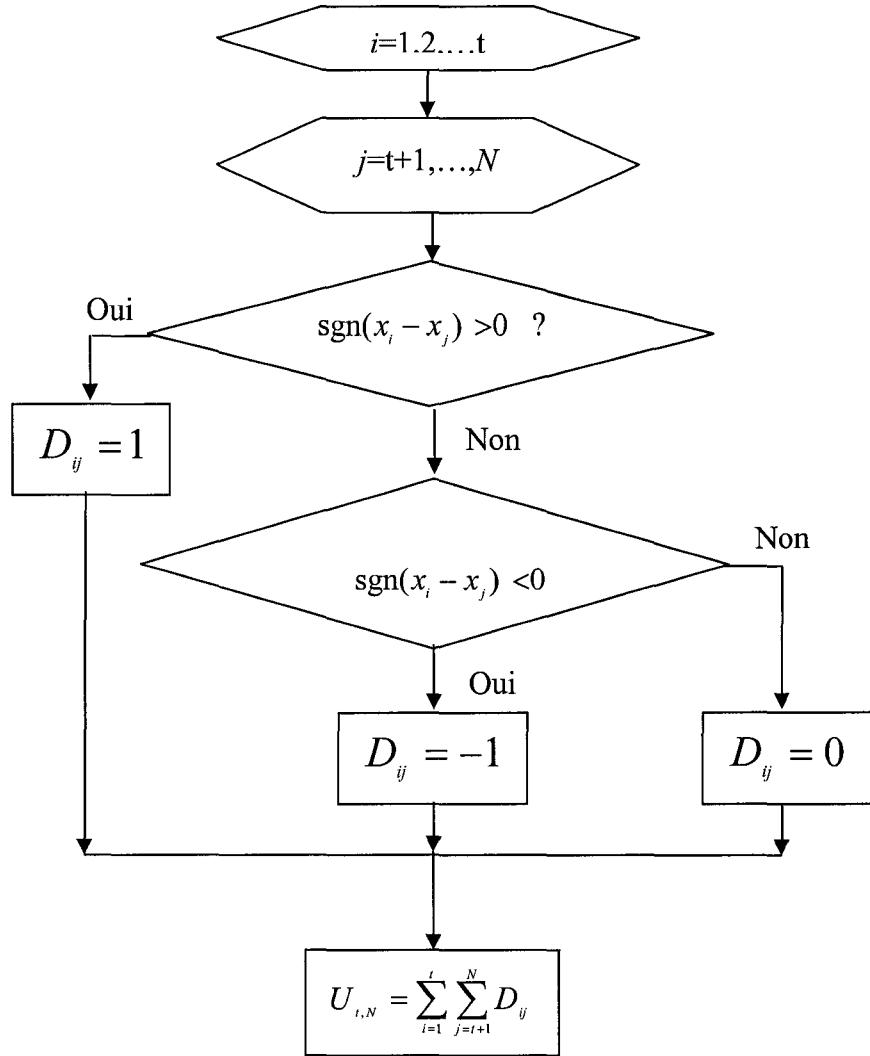


Figure 6 La statistique de Mann-Whitney

2.4 Méthode bayésienne de Lee et Heghinian

La méthode bayésienne de Lee et Heghinian [22] ne s'exprime pas comme un test statistique classique. Il s'agit plutôt d'une approche paramétrique qui demande une

distribution normale des variables étudiées. L'interprétation de cette méthode vise à confirmer ou infirmer l'hypothèse d'un changement de moyenne dans la série chronologique. Le modèle de base de la procédure est le suivant [24, 22]:

$$X_j = \begin{cases} \mu + \varepsilon_j & j = 1, \dots, \tau \\ \mu + \delta + \varepsilon_j & j = \tau + 1, \dots, n \end{cases} \quad (2.3)$$

Où les ε_j sont des variables indépendants et normalement distribués, de moyenne nulle et de variance σ^2 . τ et δ représentent respectivement la position dans le temps et l'amplitude d'un changement éventuel de moyenne.

L'approche bayésienne est basée sur les distributions marginales a posteriori de τ et δ . Donc l'équation (2.4) peut être écrite sous sa forme matricielle telle que [22]

$$X = \mu I_n + \varepsilon + \delta c_\tau \quad (2.4)$$

Où

$$X = [X_1, X_2, \dots, X_n]^T$$

$$\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]^T$$

I_n est un vecteur à n dimensions dont les éléments sont des unités.

Notons de même que c_τ est un vecteur identité lequel les τ premières composantes sont nulles et les derniers $n - \tau$ composantes sont des identités.

Lee et Heghinian ont posé les distributions à priori des paramètres $\tau, \delta, \mu, \sigma$ comme suit [22]:

- $P(\tau) = \frac{1}{n-1}$, $\tau = 1, 2, \dots, n-1$;
- $P(\delta)$ est une probabilité qui suit la loi normale dont la variance est σ_δ^2 [22];
- $P(\mu)$ est une probabilité qui suit la loi normale dont la variance σ_μ^2 [22] ;
- $p(\sigma) \approx \sigma^{-1}$ est la probabilité de l'écart type inversement proportionnelle à celui-ci;

Ces paramètres sont statistiquement indépendants.

Le vecteur $x = [x_1, x_2, \dots, x_n]^T$ a la forme d'une probabilité normale avec une moyenne δc_τ et une matrice de covariance U de la forme suivante [22]:

$$U = \sigma^2 \left[I_n + \left(\frac{\sigma^2}{\sigma^2} \right) \frac{1_n 1^T}{n} \right] \quad (2.5)$$

Avec I_n est une matrice identité d'ordre n.

Et on a la matrice inverse qui est U^{-1}

$$U^{-1} = \sigma^{-2} \left[I_n - \left(1 + r \right)^{-1} \frac{1_n 1^T}{n} \right] \quad (2.6)$$

Avec $r = \left(\frac{\sigma^2}{n} \right) \frac{1}{\sigma^2}$

Et le déterminant qui est

$$\begin{aligned} |U| &= \sigma^{2n} \left| I_n + \left(\frac{\sigma^2}{\sigma^2} \right) \frac{1_n 1^T}{n} \right| \\ &= n \sigma^{2(n-1)} \sigma^2 \mu (1+r) \end{aligned}$$

Et qui peut s'écrire sous la forme

$$|U|^{-\frac{1}{2}} = \sigma^{-(n-1)} \frac{1}{n^{\frac{1}{2}}} \sigma_\mu \quad (2.7)$$

Avec cette condition $\sigma^2 \ll \sigma^2 \mu$

$$p(\tau, \delta, \sigma_X) \propto \sigma^{-n} \exp \left[- \frac{1}{(2\sigma^2) \left\{ Q(\tau, \delta) + \sigma^2 \delta^2 \frac{1}{\sigma^2} \right\}} \right] \quad (2.8)$$

$$Q(\tau, \delta) = (X - \delta c_\tau)^T \left[I_n - \frac{1_n 1^T}{n} \right] (X - \delta c_\tau) \quad (2.9)$$

Avec une transformation algébrique de Q

$$Q(\tau, \delta) = H(\tau) + \frac{\tau(n-\tau)(\delta - \hat{\delta}_\tau)^2}{n} \quad (2.10)$$

Où

$$\hat{\delta}_\tau = \bar{X}_{n-\tau} - \bar{X}_\tau \quad (2.11)$$

Et on trouve

$$H(\tau) = \sum_{i=1}^{\tau} (X_i - \bar{X}_{\tau})^2 + \sum_{i=\tau+1}^n (X_i - \bar{X}_{n-\tau})^2 \quad (2.12)$$

Et si on ajoute la condition $\sigma^2 \ll \sigma^2_s$ alors si l'étude est limitée à la distribution de τ seulement:

$$p(\tau, \delta, \sigma_X^2) \propto \frac{\exp[-Q(\tau, \delta)/2\sigma^2]}{\sigma^n} \quad (2.13)$$

Si p , la densité de probabilité de τ est uniforme, elle est alors définie par [24, 22]

$$p\left(\frac{\tau}{X}\right) \propto \left[\frac{n}{\tau(n-\tau)} \right]^{\frac{1}{2}} [R(\tau)]^{-\frac{n-2}{2}} \quad (2.14)$$

Et $0 \leq \tau \leq n-1$ avec

$$R(\tau) = \frac{\left[\sum_{i=1}^{\tau} (x_i - \bar{x}_{\tau})^2 + \sum_{i=\tau+1}^n (x_i - \bar{x}_{n-\tau})^2 \right]}{\sum_{i=1}^N (x_i - \bar{x}_n)^2} \quad (2.15)$$

avec

$$\bar{x}_n = \frac{\sum_{i=1}^n x_i}{n}, \bar{x}_{\tau} = \frac{\sum_{i=1}^{\tau} x_i}{\tau}, \bar{x}_{n-\tau} = \frac{\sum_{i=\tau+1}^N x_i}{(N-\tau)}$$

Donc la méthode établit la distribution de probabilité a posteriori de la position dans le temps d'un changement. Le mode de la distribution à posteriori de τ sert à estimer la date de la rupture, lorsque la distribution est unimodale. Plus la dispersion de la distribution est faible, plus l'estimation est précise. La méthode bayésienne de Lee et Heghinian impose normalité, constance de la variance et non-autocorrélation [22].

2.5 Procédure de segmentation de Pierre Hubert

La méthode de Pierre Hubert [21] est basée sur une procédure originale de segmentation des séries hydrométéorologiques qui représentent le plus souvent la température, les précipitations, et le débit des grands lacs. Cette méthode permet de mettre en évidence le

problème de la stationnarité et de la modélisation des séries hydrométéorologiques. Pierre Hubert a testé sa méthode de segmentation en modélisant de courtes séries empiriques [21].

Il a utilisé une procédure de segmentation lui permettant de partitionner une série hydrométéorologique donnée en m segments de façon à ce que la moyenne de chaque segment soit nettement différente de celle du segment adjacent.

La segmentation de Pierre Hubert est acceptée lorsque l'écart quadratique entre celle-ci et de la série est minimale. Quoique cette condition a été jugée nécessaire, néanmoins elle a fait preuve de non suffisance pour la détermination de la segmentation initiale. D'où l'utilisation du test de Scheffé [36] qui représente une partie complémentaire et auxiliaire de la méthode de segmentation de Pierre Hubert.

La méthode de ce dernier augmente la capacité de rechercher les changements des moyennes des segments des séries hydrométéorologiques.

Nous présenterons dans ce qui suit les principales définitions de la procédure de segmentation des séries. Ces définitions sont extraites de l'ouvrage de pierre Hubert [21].

Supposons que l'on ait une série chronologique de n valeurs numériques $x_{i=1,2,\dots,n}$. Un segment donnée extrait de la série initiale x_i est désigné par $x_{i=i_1,i_2}$ avec $1 \leq i_1 < i_2 \leq n$.

La partition de la série initiale en m segments représente une segmentation d'ordre m . Pierre Hubert a défini $i_{k=1,2,\dots,m}$ comme étant le rang de la série initiale de l'extrémité du k ième segment avec $i_0 = 0 < i_1 < \dots < i_k < \dots < i_{m-1} < i_m < n$.

On désigne par n_k la longueur du k ième segment défini comme suit :

$$n_k = i_k - i_{k-1}$$

sa moyenne locale est exprimée par :

$$\bar{x}_N = \frac{\sum_{i=i_{k-1}+1}^{i_k} x_i}{n_k}$$

on définit les deux quantités suivantes :

$$d_k = \sum_{i=i_{k-1}+1}^{i_k} (x_i - \bar{x}_k)^2$$

et

$$D_m = \sum_{k=1}^{k=m} d_k$$

avec D_m représentant l'écart quadratique entre la série et la segmentation considérée.

Cette quantité permet de calculer l'approximité de la série et de la segmentation qui lui est appliquée. Cet écart ne dépend pour une série initiale donnée que de la segmentation considérée.

Le test de Scheffé [36] s'applique pour vérifier si les moyennes de deux segments successifs sont nettement différentes. Il a défini ψ_s comme étant le contraste du segment initial tel que:

$$\psi_s = \sum_{k=1}^{k=m} c_k \mu_k \quad (2.16)$$

Il a démontré que tout contraste estimé est exprimée par :

$$\psi_{s_e} = \sum_{k=1}^{k=m} c_k \bar{x} \quad (2.17)$$

Pierre Hubert estime qu'une segmentation d'ordre m n'est acceptable que si tous les contrastes définis par :

$$\psi_{s_{kk}} = \bar{x}_k - \bar{x}_{k-1}, \quad k = 1, 2, \dots, m-1 \quad (2.18)$$

sont différents de zéro, au niveau de signification du test de Scheffé.

L'efficacité de cette procédure a été confirmée par Lubés Niel [24], et ceci particulièrement au niveau de signification 0.01 du test de Scheffé. De plus elle s'est révélée robuste en confirmant les discontinuités détectées initialement après l'allongement de dix ans des séries étudiées en Afrique de l'ouest.

2.6 Statistique de Buishand

La statistique de Buishand est souvent utilisée pour tester l'homogénéité d'une série chronologique X_1, X_2, \dots, X_n . Cette méthode suppose deux hypothèses, l'une que nous désignerons par H_{B0} , et dans ce cas la série X_i a la même moyenne, et l'autre désignée par H_{B1} qui implique un changement de la moyenne de la série.

Les principales propriétés du test statistique de Buishand sont illustrées pour le cas où les séries X_i sont normalement distribuées avec une moyenne $E_B(X_i)$ définie telle que [8] :

$$E_B(X_i) = \begin{cases} \mu & i = 1, \dots, m \\ \mu + \Delta & i = m + 1 \dots n \end{cases} \quad (2.19)$$

avec

$$\text{var}(X_i) = \sigma_x^2 \quad (2.20)$$

$\text{var}(X_i)$ représente la variance de la série chronologique X_i .

Le modèle issu de ce test statistique indique suivant l'équation (2.19) un saut de la moyenne dont l'amplitude est désignée par Δ , et ceci après m observations.

La déviation cumulative est un paramètre qui permet de tester l'homogénéité de la série chronologique X_i . Elle est défini par [8, 9] :

$$\begin{cases} S_0^* = 0 \\ S_k^* = \sum_{i=1}^k (X_i - \bar{X}) \quad k = 1, \dots, n \end{cases} \quad (2.21)$$

Quand Δ est négative, cela implique que la plupart des valeurs de la déviation cumulative S_k^* sont positives, car la série X_i devient supérieure à sa moyenne \bar{X} quand $i \leq m$, et inférieure à celle-ci quand $i > m$. Par contre lorsque Δ est positive, S_k^* devient négative.

La statistique de Buishand [8, 9] a été obtenue en se basant sur la formulation originale de Gardner [18]. Ce dernier a utilisé pour un test bilatéral de rupture en moyenne à un instant inconnu, et est définie par :

$$G = \sum_{k=1}^{N-1} P_k \left(\frac{S_k}{\sigma_x} \right)^2 \quad (2.22)$$

avec $S_k = \sum_{i=1}^k (X_i - \bar{X})$

P_k représente la probabilité à priori que la rupture survienne juste après la $k^{ième}$ observation. Le test de Gardner suppose que la variance σ_x^2 est connue. Par contre le test Buishand traite le cas où la variance est inconnue, et dans ce cas, il l'a remplacé par la variance de l'échantillon D_x^2 tel que :

$$D_x^2 = \sum_{i=1}^N \frac{(X_i - \bar{X})^2}{N} \quad (2.23)$$

P_k étant uniforme par hypothèse, et la statistique de Buishand est définie par :

$$U = \frac{\sum_{k=1}^{N-1} \left(\frac{S_k}{D_x} \right)^2}{N(N+1)} \quad (2.24)$$

2.7 Test de corrélation sur le rang

Dans la méthode de Mann-Kendall [31], le test de corrélation sur le rang est basé sur le calcul du nombre P de paires (X_i, X_j) pour lequel $X_j > X_i (j > i, i = 1, \dots, N-1)$. Le nombre de pair P est obtenu en comparant la valeur du premier terme de la série, X_1 avec les autres termes suivants jusqu'au dernier X_N . Et le nombre des termes dont la valeur dépasse X_1 est compté, et ainsi de suite pour les autres termes jusqu'à X_{N-1} .

Donc à partir du nombre de paire P et en posant l'hypothèse nulle H_0 de stationnarité de la série on définit la variable ω de Mann-Kendall comme suit:

$$\omega = \frac{4P}{N(N-1)} - 1 \quad (2.25)$$

qui suit une distribution normale de moyenne nulle et de variance égale à :

$$\sigma_\omega^2 = \frac{2(2N+5)}{9N(N-1)} \quad (2.26)$$

Pour un risque α de première espèce donnée, l'acceptation de H_0 est défini par l'appartenance de ω à l'intervalle $\left[-U_{\frac{1-\alpha}{2}} \sigma_\omega, U_{\frac{1-\alpha}{2}} \sigma_\omega \right]$, où $U_{\frac{1-\alpha}{2}}$ est la valeur de la variable normale réduite de probabilité de non dépassement $1 - \frac{\alpha}{2}$ [24].

l'hypothèse alternative H_1 de ce test est celle dans le cas où la série chronologique est non stationnaire.

2.8 Conclusion

Nous avons présenté dans ce chapitre cinq méthodes basées sur des tests statistiques. Ces derniers permettent de détecter les ruptures dans les séries chronologiques. Les principaux tests statistiques que nous avons exposés dans ce chapitre nous ont permis principalement de voir comment ces méthodes réalisent l'identification et la modélisation des changements des paramètres statistiques et stochastiques liés à plusieurs séries chronologiques. Nous avons constaté également que les paramètres statistiques sont généralement caractérisés par le changement des moyennes d'une série temporelle. Nous avons commencé par présenter la méthode de test de Pettitt. Nous avons ensuite donné un aperçu sur la méthode bayésienne de Lee-Heghinian, nous avons vu que cette méthode n'est pas considérée comme étant un test statistique classique. Elle est toutefois basée sur une approche paramétrique qui permet de confirmer ou d'infirmer l'éventualité d'un changement de moyenne dans la série chronologique. Par la suite nous avons exposé dans ce chapitre la méthode basée sur la procédure de segmentation des séries hydrométéorologique de Hubert. Nous avons vu qu'elle ne constitue pas un test statistique, cependant elle demeure adaptée à la recherche des multiples changements de moyenne dans la série. Nous avons présenté ensuite la méthode de Buishand, qui est généralement utilisée pour effectuer un test bilatéral de rupture en moyenne à un instant inconnu. Finalement, la dernière méthode que nous avons présenté est celle qui correspond à l'approche basée sur le test de corrélation sur le rang.

CHAPITRE 3

TRANSFORMÉE EN ONDELETTES

3.1 Introduction

Dans ce chapitre, nous commencerons par présenter brièvement la transformée de Fourier et la transformée en ondelettes. Nous montrerons la différence entre ces deux transformées au sens de la résolution temps fréquence. Nous présenterons ensuite un bref aperçu de l'analyse multiresolution et sa relation avec les bancs de filtres. À la fin de ce chapitre, nous aborderons les propriétés des ondelettes de Haar.

3.2 Transformée de Fourier

La transformée de Fourier permet de représenter le signal temporel dans le domaine fréquentiel. Cela s'effectue en projetant ce signal sur des bases exponentielles (cosinus et sinus). Elle s'écrit pour un signal $f(t)$ comme suit:

$$F(f) = \int_{-\infty}^{+\infty} f(t) e^{-j2\pi ft} dt \quad (3.1)$$

La transformée inverse de $F(f)$ est définie par:

$$f(t) = \int_{-\infty}^{+\infty} F(f) e^{j2\pi ft} df \quad (3.2)$$

La transformée de Fourier est bien adaptée à la représentation dans le domaine fréquentiel des signaux stationnaires (signaux ayant des propriétés statistiquement invariantes dans le temps). Cependant elle s'avère inefficace pour la description temps-fréquence des signaux non stationnaires. D'où l'utilisation de la transformée de Fourier à fenêtre glissante. Cette dernière permet de représenter les composantes fréquentielles du signal à chaque intervalle de temps.

3.3 Transformée de Fourier à fenêtre glissante

La transformée de Fourier à fenêtre glissante permet d'analyser le signal en le découplant arbitrairement en segments de longueur limités. Elle projette le signal sur une base de fonctions $g(t)$ appelées fenêtres glissantes selon la relation suivante:

$$F(\mu, f) = \int_{-\infty}^{+\infty} f(t) g(t - \mu) e^{-j2\pi t} dt \quad (3.3)$$

Où $g(t)$ est la fonction fenêtre glissante centrée autour de $t = \mu$. Gabor [17] a choisi la fonction gaussienne définie par:

$$g(t) = \pi^{-1/4} e^{-\frac{t^2}{2}} \quad (3.4)$$

comme étant la fonction à fenêtre glissante. La transformée de Gabor est définie par [17]:

$$G_a(\mu, f) = \pi^{-1/4} \int_{-\infty}^{+\infty} f(t) e^{-\frac{(t-\mu)^2}{2}} e^{-j2\pi ft} dt \quad (3.5)$$

La transformée de Fourier à fenêtre glissante permet une représentation temps-fréquence. Cependant la longueur de la fenêtre glissante étant fixe, cette représentation ne paraît pas bien adaptée aux signaux ayant des composantes de longueurs très différentes de celle de la fenêtre glissante. La transformée en ondelettes par contre, fait usage de fonctions qui en se dilatant et en se contractant, localisent mieux les différentes composantes temporelles et fréquentielles du signal étudié.

3.4 Transformée en ondelettes

La transformée en ondelettes d'un signal $f(t)$, s'obtient en projetant ce dernier sur une base de fonctions appelées ondelettes [14]. Ces fonctions sont générées à partir d'une fonction appelé ondelette mère indiquée par ψ . Chaque ondelette est caractérisée par deux paramètres. L'un représente le facteur d'échelle indicé par a , et l'autre représente le

paramètre de translation indicé par b. Une ondelette $\psi_{a,b}(t)$ est obtenue à partir de l'ondelette mère suivant l'expression de l'équation 3.6.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad b \in R, a > 0 \quad (3.6)$$

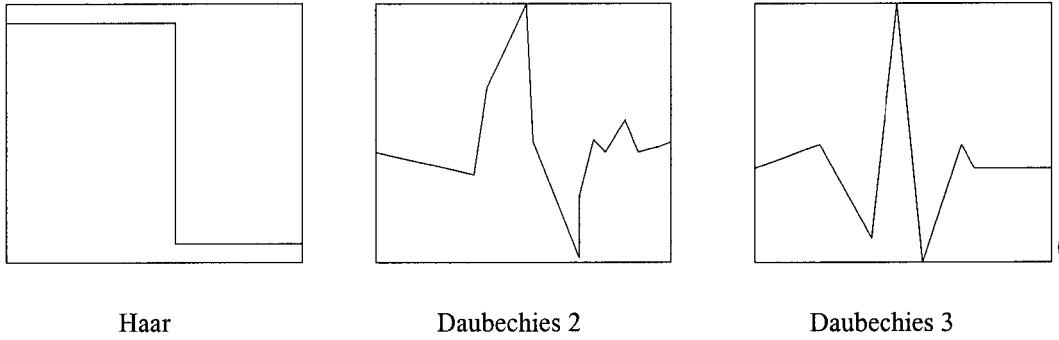


Figure 7 Quelques formes d'ondelettes

Ainsi, la transformée en ondelette continue d'un signal $f(t)$ est définie par:

$$W_f(a,b) = \int_{-\infty}^{+\infty} f(t) \psi_{a,b}(t) dt \quad (3.7)$$

La transformée en ondelettes continue est inversible si et seulement si les deux conditions d'admissibilité sont vérifiées [37]:

$$\int_{-\infty}^{+\infty} \frac{|\Psi(f)|}{|f|} df = k < \infty, \text{ ou } k = cte \quad (3.8)$$

$$\Psi(0) = \int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (3.9)$$

Où $\Psi(f)$ est la transformée de Fourier de $\psi(t)$.

L'expression (3.9) indique que la fonction d'ondelette mère doit être à moyenne nulle. Les deux conditions d'admissibilité étant satisfaites, la transformée en ondelette continue inverse existe et, nous avons son expression telle que [14]:

$$f(t) = \frac{C_\psi}{a^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_f(a, b) \psi_{a,b}(t) da db \quad (3.10)$$

Où C_ψ est une constante définie comme suit :

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(f)|^2}{f} df \quad (3.11)$$

La constante C_ψ est finie lorsque la deuxième condition d'admissibilité est satisfaite (c'est à dire lorsque l'intégrale de l'ondelette $\psi(t)$ est nulle).

La transformée en ondelette discrète dyadique est définie en discrétilisant le facteur d'échelle a et le paramètre de translation b , selon:

$$a = 2^m \text{ et } b = n 2^m \quad \text{avec } m, n \in \mathbb{Z} \quad (3.12)$$

En conséquence, la transformée en ondelette discrète d'un signal $f(t)$ est définie par [39] :

$$W_f(m, n) = \int_{-\infty}^{+\infty} f(t) \psi_{m,n}(t) dt \quad (3.13)$$

Où $\psi_{m,n}(t)$ est définie par:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{2^m}} \psi(2^{-m}t - n) \quad (3.14)$$

3.5 Représentation temps-fréquence et temps-échelle

La représentation temps-fréquence permet de localiser en temps et en fréquence les fonctions de base sur lesquelles le signal $f(t)$ est projeté. La résolution temporelle R_t et la résolution fréquentielle R_f de ces fonctions déterminent les pavés du plan temps-fréquence. Dans le cas de la transformée à fenêtre glissante, la fonction de base est déterminée par la fenêtre choisie $g(t)$, sa résolution temporelle et sa résolution fréquentielle peuvent être respectivement obtenues selon [17]:

$$R_t^2 = \int_{-\infty}^{+\infty} t^2 |g(t)|^2 dt \quad (3.15)$$

$$R_f^2 = \int_{-\infty}^{+\infty} f^2 |G(f)|^2 df \quad (3.16)$$

$$\int_{-\infty}^{+\infty} |g(t)|^2 dt = \int_{-\infty}^{+\infty} |G(f)|^2 df = 1$$

Compte tenu du principe d'incertitude de Heinsenberg, il est impossible d'avoir à la fois une localisation parfaite en temps et en fréquence (c'est-à-dire un pavé réduit à un point $[t_h, f_h]$) comme l'indique la figure 8. Selon le principe d'Heinsenberg le produit $R_t \times R_f$ est borné par:

$$R_t \times R_f \geq \frac{1}{2\pi} \quad (3.17)$$

Étant donnée l'expression (3.4) on peut démontrer [17] que pour la fenêtre de Gabor, nous avons les résolutions suivantes:

$$R_t = 1 \quad (3.18)$$

et

$$R_f = \frac{1}{2\pi} \quad (3.19)$$

d'où $R_t \times R_f = \frac{1}{2\pi}$

Nous pouvons constater à partir des expressions (3.18) et (3.19) que les résolutions temporelle R_t et fréquentielle R_f sont constantes. Par conséquent ces résolutions sont indépendantes, de sorte que le pavage de l'espace temps-fréquence (voir figure 8) est régulier.

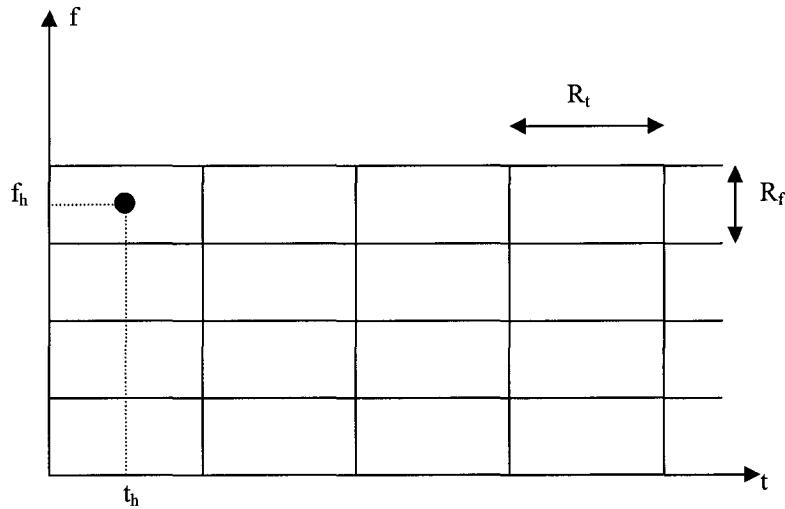


Figure 8 Pavage temps-fréquence pour la transformée de Gabor.

Dans le cas de la transformée en ondelette, nous parlons de la représentation temps échelle, où l'expression qui traduit la relation entre la fréquence et l'échelle s'écrit comme ci-dessous:

$$a = \frac{k}{f} \quad (3.20)$$

où k est une constante.

On montre que les relations qui existent entre les résolutions temporelle et fréquentielle d'une ondelette $\psi_{a,b}(t)$ et celles de l'ondelette mère $\psi(t)$ s'écrivent tel que [37]:

$$R_{t\psi_{a,b}} = a R_{t\psi} \quad (3.21)$$

$$R_{f\psi_{a,b}} = \frac{R_{t\psi}}{a} \quad (3.22)$$

où $R_{t\psi}$ et $R_{f\psi}$ représente respectivement la résolution temporelle et la résolution fréquentielle de l'ondelette mère $\psi(t)$. Il en est de même pour les résolutions $R_{t\psi_{a,b}}$ et $R_{f\psi_{a,b}}$ qui correspondent à l'ondelette $\psi_{a,b}(t)$.

Les expressions (3.21) et (3.22) montrent que lorsque le facteur d'échelle a augmenté la résolution temporelle de l'ondelette augmente aussi tandis que sa résolution fréquentielle

diminue et vice-versa. Nous illustrerons la représentation temps-échelle pour la transformée en ondelette discrète à la figure 9.

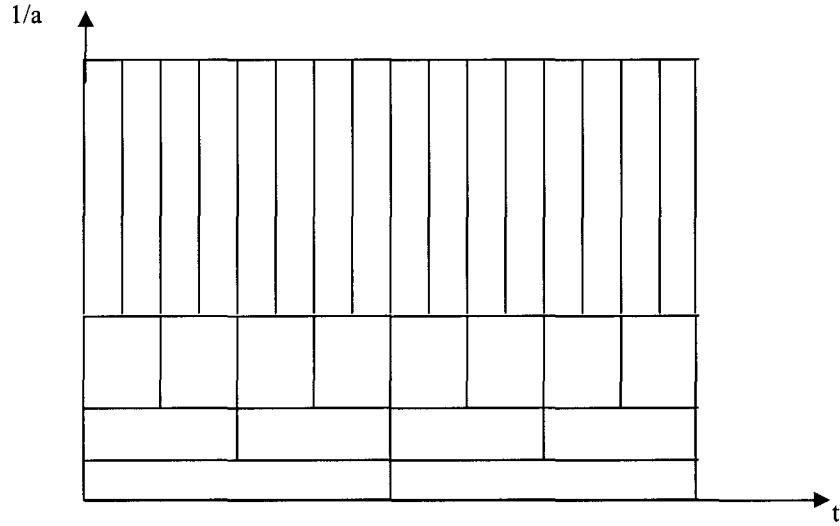


Figure 9 Représentation temps échelle de la transformée en ondelettes discrète

3.6 Analyse multirésolution[27]

Une analyse multirésolution (AMR) de l'espace des fonctions continues d'une variable réelle et de carré intégrable $L^2(\mathbb{R})$, est une suite de sous-espaces vectoriels (fermés) de $L^2(\mathbb{R})$ vérifiant les conditions suivantes [26]:

$$1. \forall j \in \mathbb{Z}, V_{j+1} \subset V_j \quad (3.23)$$

L'expression (3.23) traduit le fait que la projection du signal dans V_{j+1} constitue une approximation plus grossière de celui-ci que la projection dans V_j

$$2. \cap_{j \in \mathbb{Z}} V_j = \{0\} \quad (3.24)$$

$\cup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R})$, l'union des espaces V_j est dense dans $L^2(\mathbb{R})$

$$3. \forall j \in Z, f(x) \in V_j \Leftrightarrow f(2^{-j}x) \in V_{j+1} \quad (3.25)$$

L'équation (3.25) exprime l'invariance par rapport au changement d'échelle, dans l'espace d'approximation V_j .

$$4. \forall k \in Z, f(x) \in V_0 \Leftrightarrow f(x - k) \in V_0 \quad (3.26)$$

La relation (3.26) montre que chaque espace d'approximation est invariant par translation temporelle.

5. Il existe une fonction $\varphi(x)$, appelée fonction d'échelle telle que:

$$\varphi(x) \in V_0 \text{ et l'ensemble } \{\varphi(x - k)\}_{k \in Z} \text{ forme une base orthonormale de } V_0.$$

L'analyse multirésolution d'une fonction $f(x)$ de $L^2(\mathbb{R})$ consiste à calculer les projections orthogonales de $f(x)$ sur les espaces $V_j, j \in Z$. Nous donnerons des interprétations explicatives de la définition de l'AMR, ainsi que des conditions de 1 à 5 indiquées précédemment.

La condition 1 indique que V_{j+1} est un sous-espace de V_j . Elle montre aussi que l'approximation sur V_j contient toute l'information nécessaire pour calculer l'approximation V_{j+1} .

Les conditions 2 montrent que, quand j tend vers l'infini $+\infty$, on perd toutes l'information sur le signal traiter par contre, quand j tend vers 0 l'approximation tend asymptotiquement vers le signal d'origine.

La condition 3 permet de passer de l'espace V_j à V_{j+1} en multipliant par 2 le facteur d'échelle. j qui est appelé niveau de résolution. Cette condition permet de déduire que, quel que soit $j \in Z$, $f(x)$ appartient à V_0 si et seulement si $f(2^{-j}x)$ appartient à V_j .

La condition 4 implique que, pour tout $j \in Z$, l'espace V_j est invariant par translation d'un multiple de 2^j tel que:

$$\forall k \in Z, f(x) \in V_j \Rightarrow f(x - k 2^j) \in V_j \quad (3.27)$$

Les conditions 3 et 4 se traduisent par le fait que l'ensemble $\left\{2^{-j/2} \varphi(2^{-j}x - k), k \in \mathbb{Z}\right\}$ est une base orthonormale de V_j , car la base qui correspond à l'espace d'approximation centrale V_0 est orthonormale.

L'AMR d'un signal $f(x)$ permet de réaliser des projections successives de $f(x)$ sur les espaces V_j , cela conduit à des approximations de plus en plus grossières de $f(x)$ quand j augmente. La différence entre deux approximations successives A_j et A_{j+1} représente l'information caractérisant le détail du signal à l'échelle $j+1$. A_j (respectivement A_{j+1}) sont les projections du signal $f(x)$ sur les espaces V_j (respectivement V_{j+1}). Ce détail représente l'information contenue dans le sous-espace W_j orthogonal à V_j tel que [26]:

$$V_j = V_{j+1} \oplus W_{j+1} \quad (3.28)$$

Où le symbole \oplus indique qu'on calcule la somme directe des sous-espaces vectoriels. Le détail D_j désigne la projection du signal $f(x)$ sur les sous-espaces, et W_j représente l'espace des détails au niveau d'échelle j . Si nous désignons la projection du signal $f(x)$ sur l'espace V_j par $\text{Pro}_{V_j} f(x)$, on obtient en se basant sur l'expression (3.28) l'équation ci-dessous [26]:

$$\text{Pro}_{V_j} f(x) = \text{Pro}_{V_{j+1}} f(x) + \text{Pro}_{W_{j+1}} f(x) \quad (3.29)$$

Une conséquence supplémentaire de l'AMR est l'existence d'une fonction appelée ondelette mère $\psi(t)$ telle que l'ensemble $\left\{2^{-j/2} \psi(2^{-j}x - k), k \in \mathbb{Z}\right\}$ est une base orthonormale de W_j si et seulement si la condition 3 (condition qui caractérise les propriétés du changement d'échelle) ainsi que la condition 4 (condition d'invariance par translation) sont satisfaites.

3.7 Interpolation en bancs de filtres

On montre [19] que l'on peut écrire toute fonction de l'espace d'approximation V_1 comme une combinaison linéaire des fonctions $\varphi(x - n)$, $n \in Z$, qui constituent une base de V_0 , et on a :

$$\varphi\left(\frac{x}{2}\right) = \sqrt{2} \sum_{n \in Z} h(n) \varphi(x - n) \quad (3.30)$$

où $h(n)$ est une suite numérique qui peut être interprété comme étant la réponse impulsionnelle d'un filtre.

L'orthonormalité de la famille des fonctions $\{\varphi(x - n), n \in Z\}$ permet d'obtenir $h(n)$ par le produit scalaire comme indiqué ci-dessous

$$h(n) = \langle \varphi_{1,0}, \varphi_{0,n} \rangle \quad (3.31)$$

où $\varphi_{j,k} = 2^{\frac{j}{2}} \varphi(2^{-j}x - k)$ (3.32)

On montre que l'expression générale de l'équation (3.30) qui relie deux fonctions d'échelles de niveaux successifs j et $j+1$ s'écrit comme suit [7]:

$$\varphi_{j+1,k}(x) = \sum_{n=-\infty}^{+\infty} h(n - 2k) \varphi_{j,n}(x) \quad (3.33)$$

Si nous nous plaçons dans l'espace de détails W_j , on montre que la famille des ondelettes $\left\{ \frac{1}{\sqrt{2^{j+1}}} \psi\left(\frac{x}{2^{j+1}} - k\right), k \in Z \right\}$, forme une base orthonormale de $W_{j+1} \subset V_j$, et que les fonctions d'ondelette de cette base peuvent être décomposées sur la base orthonormale $\left\{ 2^{-\frac{j}{2}} \varphi\left(\frac{x}{2^j} - n\right), n \in Z \right\}$ de V_j , tel que:

$$\frac{1}{\sqrt{2^{j+1}}} \psi\left(\frac{x}{2^{j+1}} - k\right) = \sum_{n=-\infty}^{+\infty} g(n - 2k) \frac{1}{2^{\frac{j}{2}}} \varphi\left(\frac{x}{2^j} - n\right) \quad (3.34)$$

ou encore on a [7]

$$\psi_{j+1,k}(x) = \sum_{n=-\infty}^{+\infty} g(n - 2k) \varphi_{j,n}(x) \quad (3.35)$$

La suite $g(n)$ est une suite numérique qui peut être interprétée comme la réponse impulsionale d'un filtre numérique.

Les coefficients du filtre $g(n)$ s'obtiennent à partir de l'expression suivante:

$$g(n) = \langle \psi_{1,0}, \varphi_{0,n} \rangle \quad (3.36)$$

3.7.1 Algorithme de Mallat

La décomposition en ondelettes orthogonales d'un signal $f(x)$ s'obtient en déterminant, pour tout niveau d'échelle j sa projection sur la base $\left\{ 2^{-j/2} \varphi\left(\frac{x}{2^j} - k\right), k \in \mathbb{Z} \right\}$ de l'espace V_j , ainsi que sa projection sur la base $\left\{ 2^{-j/2} \psi\left(\frac{x}{2^j} - k\right), k \in \mathbb{Z} \right\}$ de l'espace W_j . Les coefficients $a_j(k)$ représentent le résultat de la projection de $f(x)$ dans V_j . Ces coefficients sont les coefficients d'approximation de $f(x)$ à l'échelle j . Ils sont définis par:

$$a_j(k) = \left\langle f(x), 2^{-j/2} \varphi\left(\frac{x}{2^j} - k\right) \right\rangle \quad (3.37)$$

Les coefficients de détails $d_j(k)$ sont les résultats de projections de $f(x)$ dans W_j . Ces coefficients sont obtenus par la relation ci-dessous:

$$d_j(k) = \left\langle f(x), 2^{j/2} \psi(2^{-j} x - k) \right\rangle \quad (3.38)$$

Grâce aux relations à 2 échelles (3.33) et (3.35), Mallat a décrit la relation qui relie les coefficients d'approximations à deux échelles successives $a_j(k)$ et $a_{j+1}(k)$, selon [27]:

$$a_{j+1}(k) = \sum_{l=-\infty}^{+\infty} a_j(l) h(l - 2k) \quad (3.39)$$

Il a démontré de la même manière la relation qui permet d'obtenir les coefficients de détail $d_{j+1}(k)$ à partir des coefficients d'approximations $a_j(k)$, est [27]:

$$d_{j+1}(k) = \sum_{l=-\infty}^{+\infty} a_j(l) g(l - 2k) \quad (3.40)$$

L'équation (3.39) montre que l'approximation du signal $f(x)$ au niveau d'échelle $j+1$ s'obtient par filtrage passe-bas (h) de l'approximation de $f(x)$ à l'échelle j , suivi par une décimation (sous échantillonnage) par facteur 2.

Par contre l'expression (3.40) permet d'obtenir le détail de $f(x)$ au niveau d'échelle j en passant l'approximation $a_j(k)$ à l'échelle j à travers le filtre passe-haut g , le résultat de filtrage est décimé par facteur 2.

La décimation est une réduction du taux d'échantillonnage par un facteur N . Supposons que l'on décime (sous échantillonne) un signal $x(n)$ par un facteur N , le résultat de la décimation s'exprime comme suit [3]:

$$y(n) = x(N \cdot n) \quad (3.41)$$

Nous schématisons cette expression par la figure suivante:

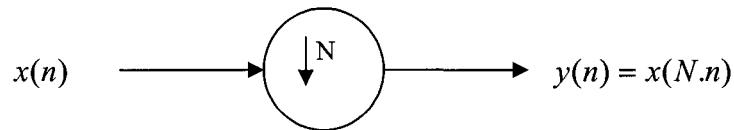


Figure 10 Décimation du signal x par facteur N

Où $\downarrow N$ symbolise l'opération de décimation d'un facteur N . Nous expliciterons l'opération de décimation d'un facteur $N=2$ en donnant un exemple illustrée par les figures 11 et 12 et par conséquent nous y constaterons que :

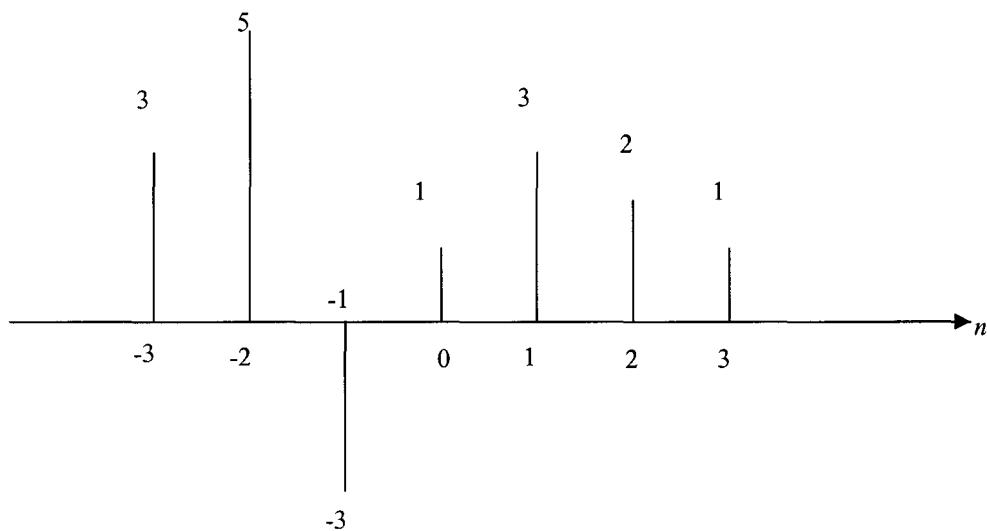
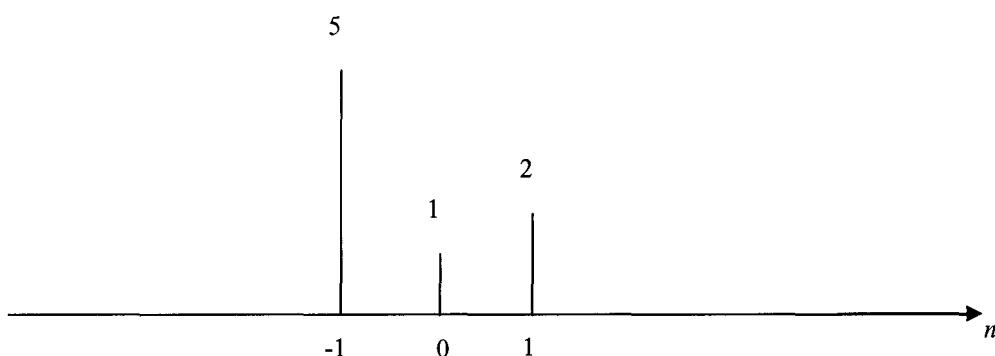
Pour $n=0$ $y(0) = x(0)=1$

Pour $n=1$ $y(1) = x(2)=2$

Pour $n = -1$ $y(-1) = x(-2)=5$

Donc pour un n quelconque on a :

$$y(n) = x(2 \cdot n)$$

Figure 11 Échantillons du signal original $x(n)$ Figure 12 Échantillons du signal $y(n) = x(2n)$

Les équations (3.39) et (3.40) décrivent l'algorithme de Mallat. Ces deux équations sont récursives. En effet elles permettent de générer itérativement les approximations et les détails du signal $f(x)$ à différents niveaux d'échelles. Le schéma équivalent aux expressions (3.39) et (3.40) est illustré sur la figure 13 ci-dessous [28]:

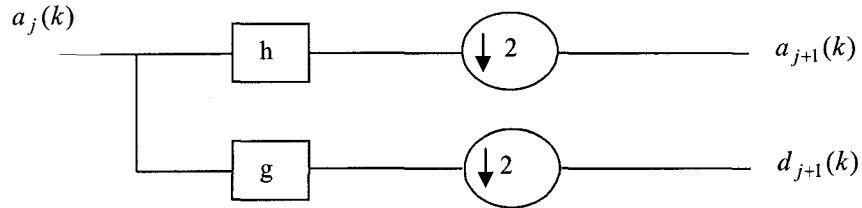


Figure 13 Schéma équivalent de l'algorithme de Mallat.

Le schéma équivalent de l'algorithme de Mallat montre que les coefficients d'approximation $a_j(k)$ peuvent être utilisés comme une entrée pour la paire de filtre (h : passe-bas, g : passe-haut), générant ainsi le détail $d_{j+1}(k)$ et l'approximation $a_{j+1}(k)$ au niveau supérieur $j+1$.

Supposons que l'on veut décomposer le signal $f(x)$ à trois niveaux d'échelles, nous obtenons ainsi par le biais de l'algorithme de Mallat la structure du banc de filtres indiqué par la figure 14.

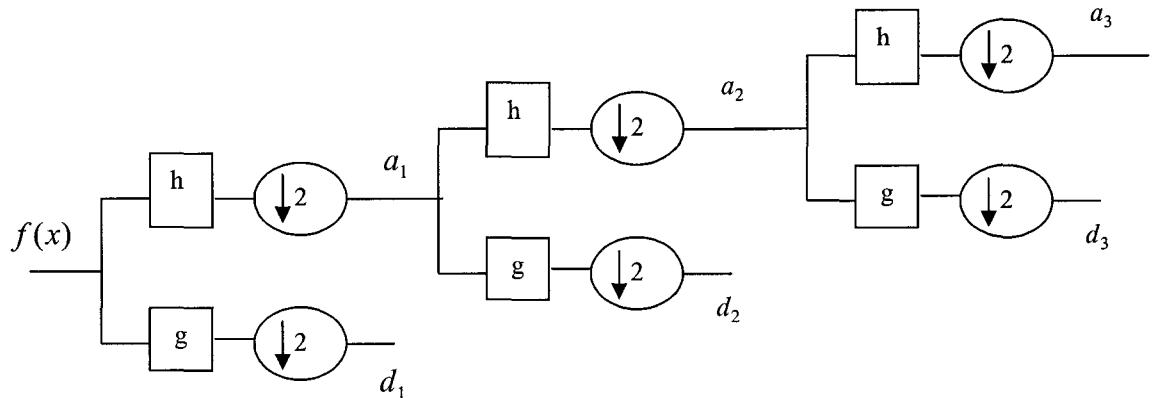


Figure 14 Structure de banc de filtres à 3 niveaux d'échelle.

3.7.2 Propriétés des filtres h et g

La réponse fréquentielle du filtre passe-bas $H(\omega)$ s'obtient à partir de la transformée de Fourier de $h(n)$ tel que:

$$H(\omega) = \sum_{n=-\infty}^{+\infty} h(n) e^{-jn\omega} \quad (3.42)$$

Si nous appliquons la transformée de Fourier à l'équation (3.30), en la combinant avec la relation (3.42) on obtient l'expression ci-dessous [29]:

$$\varphi(\omega) = \frac{1}{\sqrt{2}} H\left(\frac{\omega}{2}\right) \varphi\left(\frac{\omega}{2}\right) \quad (3.43)$$

Cette équation exprime la projection de la fonction d'échelle sur la résolution inférieure dans l'espace de Fourier. On montre que l'on peut étendre par récurrence l'équation (3.43), ainsi nous obtenons la relation qui relie la réponse fréquentielle de la fonction d'échelle et celle du filtre numérique $h(n)$ et on a [29]:

$$\varphi(\omega) = \prod_{j=1}^{+\infty} \frac{1}{\sqrt{2}} H(2^{-j}\omega) \quad (3.44)$$

On montre que la propriété d'orthonormalité de la fonction d'échelle φ , permet d'obtenir la relation ci-dessous [28]:

$$|H(\omega)|^2 + |H(\omega + \pi)|^2 = 2 \quad (3.45)$$

Nous savons que la fonction d'échelle n'est pas à moyenne nulle, en combinant cette propriété avec l'équation (3.45) et (3.43), on montre que l'on obtient les relations suivantes [28]:

$$H(0) = \sum_n h(n) = \sqrt{2} \quad (3.46)$$

$$H(\pi) = 0 \quad (3.47)$$

On montre que l'orthonormalité des fonctions d'ondelettes à travers les résolutions, implique l'expression (3.48) exprimée en fonction de $G(w)$. Cette dernière étant la réponse fréquentielle du filtre numérique $g(n)$.

$$|G(\omega)|^2 + |G(\omega + \pi)|^2 = 2 \quad (3.48)$$

Nous avons vu que la fonction d'ondelette satisfait la deuxième condition d'admissibilité qui s'exprime comme suit:

$$\Psi(0) = \int_{-\infty}^{+\infty} \psi(x) dx = 0 \quad (3.49)$$

De la même manière que l'on a trouvé l'équation (3.43) qui correspond à la fonction d'échelle, on a obtenu celle qui correspond à la fonction d'ondelette tel que [28]:

$$\Psi(\omega) = \frac{1}{\sqrt{2}} G\left(\frac{\omega}{2}\right) \Phi\left(\frac{\omega}{2}\right) \quad (3.50)$$

On montre qu'en combinant les expressions (3.48), (3.49) et (3.50) on trouve [28]:

$$\begin{cases} G(0) = \sum_n g(n) = 0 \\ G(\pi) = \sqrt{2} \end{cases} \quad (3.51)$$

Généralement les filtres h et g associés aux ondelettes orthogonales sont des filtres miroirs en quadrature (QMF: Quadrature Mirror Filter). L'équation qui régit ce genre de filtres s'exprime comme suit [34]:

$$g(n) = (-1)^n h(1-n) \quad (3.52)$$

Nous pouvons constater de l'équation (3.52) que la détermination des coefficients du filtre $h(n)$ permet de retrouver automatiquement les coefficients du filtre $g(n)$.

3.8 Les ondelettes de Haar

Les ondelettes de Haar forment la plus ancienne des bases orthonormales d'ondelettes [20]. Ce sont des fonctions discontinues, utilisées souvent pour détecter les transitions dans un signal. Elles sont bâties à partir de la fonction d'échelle φ défini par [28]:

$$\varphi(x) = \begin{cases} 1, & 0 \leq x < 1 \\ 0, & \text{ailleurs} \end{cases} \quad (3.53)$$

La figure 15 illustre la forme d'onde de la fonction d'échelle de Haar.

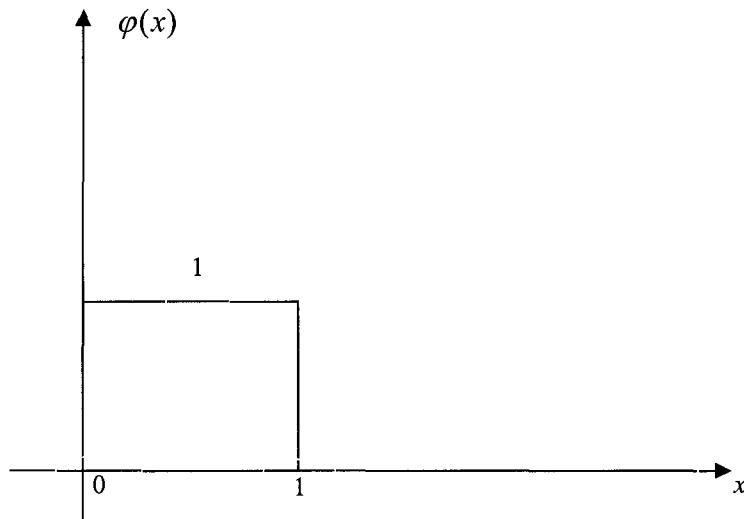


Figure 15 Fonction d'échelle de l'ondelette de Haar

La fonction d'ondelette ψ de Haar associée à cette fonction d'échelle est définie par [28, 23]

$$\psi(x) = \varphi(2x) - \varphi(2x - 1) \quad (3.54)$$

ou encore

$$\psi(t) = \begin{cases} 1, & 0 \leq x < \frac{1}{2} \\ -1, & \frac{1}{2} \leq x < 1 \\ 0, & \text{ailleurs} \end{cases} \quad (3.55)$$

La figure 16 montre la forme de l'ondelette de Haar.

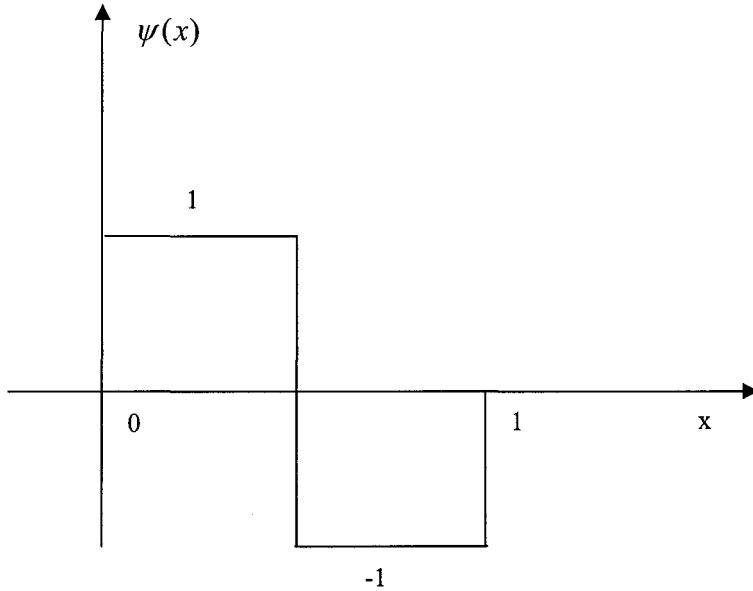


Figure 16 Ondelette de Haar

Les filtres de Haar associés à la fonction d'échelle et d'ondelette de Haar sont obtenus à partir de l'expression (3.54) et de l'équation (3.52) qui exprime la relation qui relie les filtres h et g qui sont des filtres miroirs en quadrature.

Sachant que l'on peut exprimer la fonction d'échelle $\varphi(x)$ comme suit :

$$\varphi(x) = \varphi(2x) + \varphi(2x - 1) \quad (3.56)$$

Nous pouvons combiner la relation (3.56) avec l'équation (3.30), et nous en déduisons les coefficients du filtre h correspondant selon [27, 23]

$$\left\{ \begin{array}{l} h(0) = \frac{1}{\sqrt{2}} \\ h(1) = \frac{1}{\sqrt{2}} \\ h(n) = 0 \quad \forall n \neq 0, 1 \end{array} \right. \quad (3.57)$$

À partir de l'équation (3.52) nous obtenons automatiquement les coefficients du filtre g tel que

$$\left\{ \begin{array}{l} g(0) = \frac{1}{\sqrt{2}} \\ g(1) = -\frac{1}{\sqrt{2}} \\ g(n) = 0 \quad \forall n \neq 0,1 \end{array} \right. \quad (3.58)$$

Les ondelettes de Haar sont simples et faciles à implémenter. De plus elles sont à support compact. Cette caractéristique permet le calcul de la transformée en ondelettes d'une façon exacte [20,25]. Elles forment une base orthonormale qui est le type de base utilisé pour l'analyse multirésolution. De plus pour détecter les frontières (ex : rupture), les ondelettes symétriques comme celles de Haar sont les plus adéquates [28,29].

3.9 Conclusion

Nous avons décrit la différence qui existe entre la transformée de Fourier T.F. et la transformée en ondelette T.O. particulièrement en terme de résolutions temps- fréquence. Nous avons aussi présenté un bref aperçu sur l'analyse multirésolution et son interprétation en banc de filtres. Cette analyse est souvent utilisé pour étudier le comportement du signal analysé en plusieurs résolutions (niveaux d'échelles).

Dans le cadre de ce travail nous utiliserons l'ondelette de Haar pour bénéficier de ses propriétés de symétrie, de compacité (à support compact), de facilité d'implémentation et finalement de sa capacité à détecter les frontières abruptes.

CHAPITRE 4

LA MÉTHODE DE L'ONDELETTE DE HAAR MODIFIÉE

4.1 Introduction

Nous avons présenté au chapitre 3 un aperçu sur la transformée en ondelettes et sa relation avec l'analyse multiresolution, nous y avons aussi abordé les propriétés de base de l'ondelette de Haar qui est simple et facile à implémenter et à support compact. Cette dernière caractéristique lui permet d'effectuer le calcul de la transformée en ondelettes d'une façon exacte et rapide. Elle forme aussi avec ses translatées une base orthonormale. Pour détecter les frontières (ex : rupture), on préfère souvent d'utiliser des ondelettes symétriques telle que l'ondelette de Haar. Dans ce chapitre nous décrirons et expliquerons la méthode que nous avons développée dans le cadre de ce travail. Elle est basée sur une forme modifiée de l'ondelette de Haar. La méthode de l'ondelette de Haar modifiée que nous proposons ici porte principalement sur la modification de l'ondelette de Haar dans le but d'améliorer sa capacité de détecter les périodes de sécheresse et d'inondation dans les séries chronologiques. Cette méthode consiste en la décomposition en premier lieu en ondelettes de Haar du signal traité selon un niveau de profondeur choisi. Le même traitement est ensuite effectué à nouveau avec le même signal mais cette fois-ci décalé d'un seul échantillon. Cette étape est réitérée un certain nombre de fois, puis les nouveaux coefficients d'approximations et de détails de l'ondelette de Haar approchée sont formés. L'étape suivante est le calcul des entropies correspondantes. Nous choisissons pour la normalisation les coefficients de détails du dernier niveau. Finalement le signal traité et ses coefficients d'approximation et de détails sont visualisés pour détecter les sécheresses et inondations qui existent dans les séries chronologiques analysées. Pour mieux expliquer et faciliter la compréhension de la

méthode suggérée, nous présenterons ci-dessous un exemple pour illustrer la façon d'obtenir les ruptures dans une série chronologique.

4.2 Décomposition en ondelettes de Haar du signal initial traité et de ses translatés

Étape 1.

Soit un signal S_1 de 16 échantillons, défini comme suit :

$$S_1 = [s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}, s_{13}, s_{14}, s_{15}, s_{16}].$$

Effectuons un décalage à gauche de ce signal, nous obtenons alors :

$$S_2 = [s_2, s_3, \dots, s_{16}, s_1]$$

Appliquons ensuite la même procédure au signal S_2 ceci donne :

$$S_3 = [s_3, s_4, \dots, s_{16}, s_1, s_2]$$

Le signal S_4 est à son tour obtenu par un décalage de S_3 :

$$S_4 = [s_4, s_5, \dots, s_{16}, s_1, s_2, s_3].$$

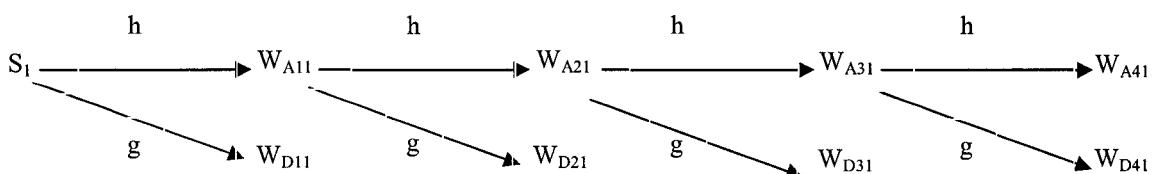


Figure 17 Décomposition en ondelettes du signal S_1 en 4 niveaux de résolution

Nous avons décomposé le signal jusqu'au niveau 4 par l'ondelette de Haar selon Mallat, S. et Daubechies, I. [28, 14, 23] tel que représenté par la figure 17. Où h et g représentent respectivement les filtres passe-bas et passe-haut. Les sorties des ces deux filtres sont décimées par un facteur 2. Les coefficients de l'approximation $W_{A_{jk}}$ peuvent par la suite être utilisés comme une autre donnée d'entrée pour la paire de filtres, générant un autre ensemble de coefficients de détails $W_{D_{j+1k}}$ et de coefficients d'approximation au niveau supérieur $W_{A_{j+1k}}$.

Nous nous référons aux équations (3.53), (3.54) et (3.55) du chapitre précédent concernant la formulation de l'ondelette de Haar qui nous permet d'obtenir les signaux suivants pour le cas où le signal S_1 contient uniquement 8 échantillons. On aura donc les premières huit fonctions de Haar [23, 20], avec $a = 1/\sqrt{2}$ tel que illustrés par la figure 18 ci-dessous.

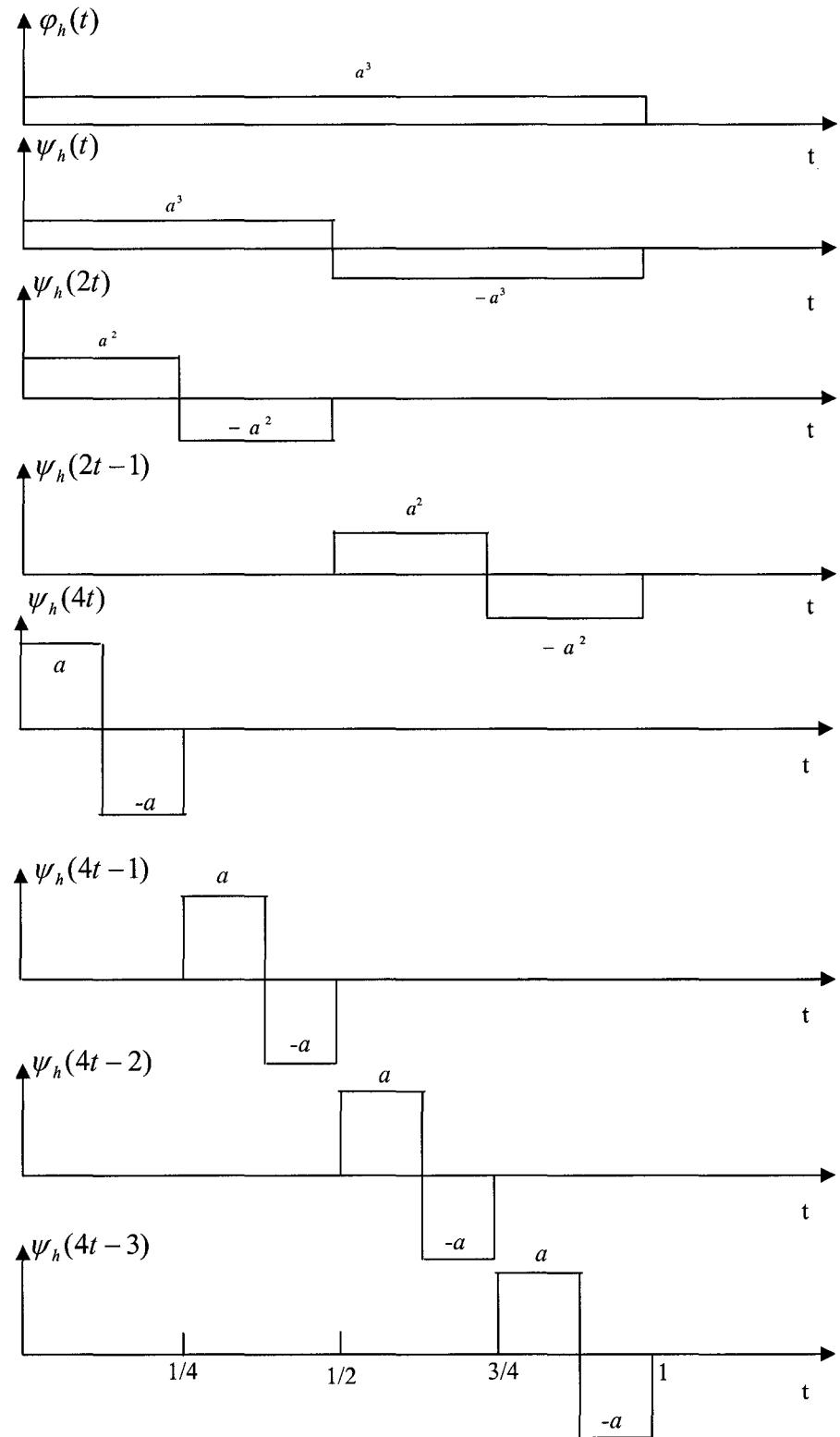


Figure 18 Les premières huit fonctions de Haar [23, 20] , $a = 1/\sqrt{2}$

La figure 18 nous permet d'obtenir les coefficients de la matrice de Haar [23, 20]

$$H_{8 \times 8} = \begin{bmatrix} a^3 & a^3 \\ a^3 & a^3 & a^3 & a^3 & -a^3 & -a^3 & -a^3 & -a^3 \\ a^2 & a^2 & -a^2 & -a^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a^2 & a^2 & -a^2 & -a^2 \\ a & -a & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a & -a & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a & -a & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a & -a \end{bmatrix}$$

De la même façon en généralisant la figure 18 pour 16 échantillons on obtient la matrice de Haar de taille 16x16 définie par :

$$H = \begin{bmatrix} a^4 & a^4 & a^4 & -a^4 & -a^4 & -a^4 & -a^4 & a^4 \\ a^4 & -a^4 \\ a^3 & a^3 & a^3 & a^3 & -a^3 & -a^3 & -a^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a^3 & a^3 & a^3 & a^3 & -a^3 & -a^3 & -a^3 & -a^3 & -a^3 \\ a^2 & a^2 & -a^2 & -a^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a^2 & a^2 & -a^2 & -a^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a^2 & a^2 & -a^2 & -a^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a^2 & a^2 & -a^2 & -a^2 & -a^2 \\ a & -a & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a & -a & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a & -a & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a & -a & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & -a & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & -a & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & -a & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & -a & 0 & 0 & 0 & 0 \end{bmatrix}$$

Les matrices $H_{8 \times 8}$ et H sont orthonormales et unitaires (leur déterminant est égal à l'unité). De plus nous posons $a = 1/\sqrt{2}$ pour normaliser l'énergie de chaque ligne de la matrice.

Multiplions la matrice de Haar (H) successivement par les signaux S_1 , S_2 , S_3 et S_4 nous obtenons alors :

$$W_1 = H \cdot S_1, W_2 = H \cdot S_2, W_3 = H \cdot S_3, \text{ et } W_4 = H \cdot S_4.$$

L'ensemble $\{W_k, k=1,2,3,4\}$ représente respectivement les coefficients des signaux $\{S_k, k=1,2,3,4\}$ résultants de ces opérations successives, et nous définissons cet ensemble W_k par :

$$W_k = \begin{bmatrix} W_{A4k}(1) \\ W_{D4k}(1) \\ W_{D3k}(l); l=1,2 \\ W_{D2k}(l); l=1,..,4 \\ W_{D1k}(l); l=1,..,8 \end{bmatrix}; k=1,..,4$$

4.2.1 Calcul de W_1

W_{A41} est le coefficient de l'approximation de l'ondelette de Haar au niveau 4

$$W_{A41} = \frac{1}{4} [s_1 + s_2 + s_3 + s_4 + s_5 + s_6 + s_7 + s_8 + s_9 + s_{10} + s_{11} + s_{12} + s_{13} + s_{14} + s_{15} + s_{16}]$$

tandis que W_{D11} , W_{D21} , W_{D31} , et W_{D41} représentent respectivement les coefficients des détails de l'ondelette de Haar aux niveaux 1, 2, 3 et 4, et nous avons :

$$W_{D11} = \frac{1}{\sqrt{2}} \begin{bmatrix} s_1 - s_2 \\ s_3 - s_4 \\ s_5 - s_6 \\ s_7 - s_8 \\ s_9 - s_{10} \\ s_{11} - s_{12} \\ s_{13} - s_{14} \\ s_{15} - s_{16} \end{bmatrix}, W_{D21} = \frac{1}{2} \begin{bmatrix} (s_1 + s_2) - (s_3 + s_4) \\ (s_5 + s_6) - (s_7 + s_8) \\ (s_9 + s_{10}) - (s_{11} + s_{12}) \\ (s_{13} + s_{14}) - (s_{15} + s_{16}) \end{bmatrix}$$

$$W_{D31} = \frac{1}{2\sqrt{2}} \left[\begin{array}{l} (s_1 + s_2 + s_3 + s_4) - (s_5 + s_6 + s_7 + s_8) \\ (s_9 + s_{10} + s_{11} + s_{12}) - (s_{13} + s_{14} + s_{15} + s_{16}) \end{array} \right]$$

$$W_{D41} = \frac{1}{4} \left[(s_1 + s_2 + s_3 + s_4 + s_5 + s_6 + s_7 + s_8) - (s_9 + s_{10} + s_{11} + s_{12} + s_{13} + s_{14} + s_{15} + s_{16}) \right]$$

4.2.2 Calcul de W_2

W_{A42} est le coefficient de l'approximation de l'ondelette de Haar au niveau 4

$$W_{A42} = \frac{1}{4} [s_2 + s_3 + s_4 + s_5 + s_6 + s_7 + s_8 + s_9 + s_{10} + s_{11} + s_{12} + s_{13} + s_{14} + s_{15} + s_{16} + s_1]$$

W_{D12} , W_{D22} , W_{D32} , et W_{D42} représentent respectivement les coefficients des détails de l'ondelette de Haar aux niveaux 1, 2, 3 et 4, et nous obtenons :

$$W_{D12} = \frac{1}{\sqrt{2}} \begin{bmatrix} s_2 - s_3 \\ s_4 - s_5 \\ s_6 - s_7 \\ s_8 - s_9 \\ s_{10} - s_{11} \\ s_{12} - s_{13} \\ s_{14} - s_{15} \\ s_{16} - s_1 \end{bmatrix}, W_{D22} = \frac{1}{2} \begin{bmatrix} (s_2 + s_3) - (s_4 + s_5) \\ (s_6 + s_7) - (s_8 + s_9) \\ (s_{10} + s_{11}) - (s_{12} + s_{13}) \\ (s_{14} + s_{15}) - (s_{16} + s_1) \end{bmatrix}$$

$$W_{D32} = \frac{1}{2\sqrt{2}} \left[\begin{array}{l} (s_2 + s_3 + s_4 + s_5) - (s_6 + s_7 + s_8 + s_9) \\ (s_{10} + s_{11} + s_{12} + s_{13}) - (s_{14} + s_{15} + s_{16} + s_1) \end{array} \right]$$

$$W_{D42} = \frac{1}{4} \left[(s_2 + s_3 + s_4 + s_5 + s_6 + s_7 + s_8 + s_9) - (s_{10} + s_{11} + s_{12} + s_{13} + s_{14} + s_{15} + s_{16} + s_1) \right]$$

4.2.3 Calcul de W_3

W_{A43} est le coefficient de l'approximation de l'ondelette de Haar au niveau 4

$$W_{A43} = \frac{1}{4} [s_3 + s_4 + s_5 + s_6 + s_7 + s_8 + s_9 + s_{10} + s_{11} + s_{12} + s_{13} + s_{14} + s_{15} + s_{16} + s_1 + s_2]$$

W_{D13} , W_{D23} , W_{D33} , et W_{D43} représentent respectivement les coefficients des détails de l'ondelette de Haar aux niveaux 1, 2, 3 et 4, nous obtenons ainsi :

$$W_{D13} = \frac{1}{\sqrt{2}} \begin{bmatrix} s_3 - s_4 \\ s_5 - s_6 \\ s_7 - s_8 \\ s_9 - s_{10} \\ s_{11} - s_{12} \\ s_{13} - s_{14} \\ s_{15} - s_{16} \\ s_1 - s_2 \end{bmatrix}, W_{D23} = \frac{1}{2} \begin{bmatrix} (s_3 + s_4) - (s_5 + s_6) \\ (s_7 + s_8) - (s_9 + s_{10}) \\ (s_{11} + s_{12}) - (s_{13} + s_{14}) \\ (s_{15} + s_{16}) - (s_1 + s_2) \end{bmatrix}$$

$$W_{D33} = \frac{1}{2\sqrt{2}} \begin{bmatrix} (s_3 + s_4 + s_5 + s_6) - (s_7 + s_8 + s_9 + s_{10}) \\ (s_{11} + s_{12} + s_{13} + s_{14}) - (s_{15} + s_{16} + s_1 + s_2) \end{bmatrix}$$

$$W_{D43} = \frac{1}{4} [(s_3 + s_4 + s_5 + s_6 + s_7 + s_8 + s_9 + s_{10}) - (s_{11} + s_{12} + s_{13} + s_{14} + s_{15} + s_{16} + s_1 + s_2)]$$

4.2.4 Calcul de W_4

W_{A44} est le coefficient de l'approximation de l'ondelette de Haar à la profondeur 4

$$W_{A44} = \frac{1}{4} [s_4 + s_5 + s_6 + s_7 + s_8 + s_9 + s_{10} + s_{11} + s_{12} + s_{13} + s_{14} + s_{15} + s_{16} + s_1 + s_2 + s_3]$$

W_{D14} , W_{D24} , W_{D34} , et W_{D44} représentent respectivement les coefficients des détails de l'ondelette de Haar aux niveaux 1, 2, 3 et 4.

$$\begin{aligned}
W_{D14} &= \frac{1}{\sqrt{2}} \begin{bmatrix} s_4 - s_5 \\ s_6 - s_7 \\ s_8 - s_9 \\ s_{10} - s_{11} \\ s_{12} - s_{13} \\ s_{14} - s_{15} \\ s_{16} - s_1 \\ s_2 - s_3 \end{bmatrix}, \quad W_{D24} = \frac{1}{2} \begin{bmatrix} (s_4 + s_5) - (s_6 + s_7) \\ (s_8 + s_9) - (s_{10} + s_{11}) \\ (s_{12} + s_{13}) - (s_{14} + s_{15}) \\ (s_{16} + s_1) - (s_2 + s_3) \end{bmatrix} \\
W_{D34} &= \frac{1}{2\sqrt{2}} \begin{bmatrix} (s_4 + s_5 + s_6 + s_7) - (s_8 + s_9 + s_{10} + s_{11}) \\ (s_{12} + s_{13} + s_{14} + s_{15}) - (s_{16} + s_1 + s_2 + s_3) \end{bmatrix} \\
W_{D44} &= \frac{1}{4} [(s_4 + s_5 + s_6 + s_7 + s_8 + s_9 + s_{10} + s_{11}) - (s_{12} + s_{13} + s_{14} + s_{15} + s_{16} + s_1 + s_2 + s_3)]
\end{aligned}$$

Pour mieux faciliter la compréhension de cet exemple, nous avons traité un signal de 16 échantillons, toutefois les approximations $W_{A41}, W_{A42}, W_{A43}, W_{A44}$ obtenus sont égaux, ceci n'est pas toujours vrai tel que nous verrons dans l'exemple que nous présenterons dans le paragraphe 4.4. Donc la formule compacte de W_k s'écrit comme suit:

$$W_k = \begin{bmatrix} W_{A4k}(l); l = 1, \dots, N/16 \\ W_{D4k}(l); l = 1, \dots, N/16 \\ W_{D3k}(l); l = 1, \dots, N/8 \\ W_{D2k}(l); l = 1, \dots, N/4 \\ W_{D1k}(l); l = 1, \dots, N/2 \end{bmatrix} \quad (4.1)$$

où N est le nombre des échantillons du signal S_1 .

Étape 2.

Selon l'analyse précédemment décrite il est possible d'écrire :

$$[W_1 \quad W_2 \quad W_3 \quad W_4] = H[S_1 \quad S_2 \quad S_3 \quad S_4]$$

Nous avons donc une équation de la forme :

$$W = H S \quad (4.2)$$

où $W = [W_1 \ W_2 \ W_3 \ W_4]$ et $S = [S_1 \ S_2 \ S_3 \ S_4]$ sont des matrices de 16x4. En utilisant $W = [W_1 \ W_2 \ W_3 \ W_4]$ nous pouvons obtenir la décomposition suivante du signal telle que $j=2$ et nous avons:

$$\hat{W}_{D2} = \left[W_{Djk}(L) = \frac{1}{2} \left(\sum_{i=k+2^j(L-1)}^{i=k+(j-1)+2^j(L-1)} s_i - \sum_{i=k+j+2^j(L-1)}^{i=k+(j+1)+2^j(L-1)} s_i \right) \right]_{\substack{L=1,2,3,(\frac{N}{2^j})=4 \\ j=2 \\ k=1,2,3,4}}$$

posons $s_i = 0$, pour $i \geq 17$.

$$\hat{W}_{D2} = [\hat{W}_{D2}(i)]_{i=1,\dots,N} \text{ alors } \hat{W}_{D2r} = [\hat{W}_{D2}(i)]_{i=1,\dots,N_c} \text{ avec } N_c = 13$$

avec N_c est le nombre des coefficients de \hat{W}_{D2r} défini par :

$$N_c = N - (2^j - 1)$$

De la même manière nous pouvons écrire cette équation sous la forme compacte comme Suit:

$$\hat{W}_{Dj} = [W_{Djk}(L)]_{\substack{L=1,\dots,\frac{N}{2^j} \\ k=1,\dots,2^j}} \quad (4.3)$$

Où

$$W_{Djk}(L) = \left(\frac{1}{\sqrt{2}} \right)^j \left(\sum_{i=k+2^j(L-1)}^{i=k+2^j(L-1/2)-1} s_i - \sum_{i=k+2^j(L-1/2)}^{i=k+2^jL-1} s_i \right) \quad (4.4)$$

4.3 Calcul “des entropies apparentes” des coefficients d’approximation et de détails de l’ondelette de Haar approchée

Dans ce qui suit nous renommons la quantité suivante $\hat{E}W_{D2r}$ “l’entropie apparente” dont la définition a été inspirée de celle de l’entropie classique. Nous y avons introduit toutefois quelques modifications pour l’adapter à nos besoins.

L’entropie de la norme d’un signal $x(n)$ est défini comme suit [12] :

$$\sum_n |x(n)|^p \text{ avec } p \geq 1 \quad (4.5)$$

Dans ce travail nous avons défini l'entropie apparente selon l'expression suivante :

$$\hat{E}W_{D2r} = - \left[\text{sign}(\hat{W}_{D2r}(i)) \left| \hat{W}_{D2r}(i) \right|^p \right]_{i=1, \dots, N_c}$$

d'une façon plus compacte, nous pouvons écrire:

$$\hat{E}W_{Djk} = - \left[\text{sign}(\hat{W}_{Djk}(i)) \left| \hat{W}_{Djk}(i) \right|^p \right]_{i=1, \dots, N_c} \quad (4.6)$$

Cette définition de l'entropie apparente a été obtenue en modifiant celle de la norme habituellement utilisée dans la littérature selon Coifman et al. [1,3]. En fait, nous l'avons défini de façon à mettre en relief les variations brusques du signal qui correspondent aux périodes des sécheresses et inondations. Cette propriété deviendra plus évidente plus loin.

Chaque élément de \hat{W}_{D2r} est un segment formé par une combinaison de 4 échantillons du signal. Si l'on désire utiliser un segment plus long formé d'un multiple de 2, tel que 8, échantillons, il faut alors effectuer 7 décalages du signal original et les multiplier par la matrice de Haar H tel qu'effectué précédemment, pour extraire les coefficients correspondants.

Nous avons utilisé l'entropie apparente définie par l'équation (4.5) du vecteur \hat{W}_{D2r} , ainsi le résultat est exploité pour des tâches de visualisation (détecter les changements de moyenne dans le signal (les séries N.B.S)).

Si nous multiplions EW_{D2r} par le signe du coefficient afin d'indiquer le sens de la rupture des moyens dans la série NB.S. Nous obtenons alors pour un p impair et égale à 3 l'expression de EW_{D2r} défini par:

$$\hat{E}W_{D2r} = - \left| \hat{W}_{D2r}(i) \right|^3 \quad (4.7)$$

En procédant de la même manière il est possible d'obtenir \hat{W}_{D3r} comme suit:

Dans ce cas il nous faut 7 décalages du signal original

$$[W_1 \quad W_2 \quad W_3 \quad W_4 \quad W_5 \quad W_6 \quad W_7 \quad W_8] = H [S_1 \quad S_2 \quad S_3 \quad S_4 \quad S_5 \quad S_6 \quad S_7 \quad S_8]$$

Nous avons donc une équation de la forme: $W = HS$

où

$$W = [W_1 \quad W_2 \quad W_3 \quad W_4 \quad W_5 \quad W_6 \quad W_7 \quad W_8]$$

et

$$S = [S_1 \quad S_2 \quad S_3 \quad S_4 \quad S_5 \quad S_6 \quad S_7 \quad S_8]$$

sont des matrices de 16x8.

Les $W_{D31}, W_{D32}, \dots, W_{D38}$ représentent respectivement les coefficients des détails de l'ondelette de Haar au niveau 3 pour chaque signal S_1, S_2, \dots, S_8 .

Nous posons $s_i = 0$, pour $i \geq 17$

$$\hat{W}_{D3} = \left[W_{D3k}(L) = \frac{1}{2\sqrt{2}} \left(\sum_{i=k+2^j(L-1)}^{i=k+j+2^j(L-1)} s_i - \sum_{i=k+j+1+2^j(L-1)}^{i=k+2j+1+2^j(L-1)} s_i \right) \right]_{\substack{j=3 \\ k=1, 2, \dots, 8}} \quad \hat{W}_{D3} = [\hat{W}_{D3}(i)]_{i=1, \dots, N} \text{ alors } \hat{W}_{D3r} = [\hat{W}_{D3}(i)]_{i=1, \dots, N_r} \text{ avec } N_r = 9$$

et nous avons :

$$\hat{W}_{D3r} = \left[\begin{array}{l} W_{D31}(1) = \frac{1}{2\sqrt{2}} [(s_1 + s_2 + s_3 + s_4) - (s_5 + s_6 + s_7 + s_8)] \\ W_{D32}(1) = \frac{1}{2\sqrt{2}} [(s_2 + s_3 + s_4 + s_5) - (s_6 + s_7 + s_8 + s_9)] \\ W_{D33}(1) = \frac{1}{2\sqrt{2}} [(s_3 + s_4 + s_5 + s_6) - (s_7 + s_8 + s_9 + s_{10})] \\ W_{D34}(1) = \frac{1}{2\sqrt{2}} [(s_4 + s_5 + s_6 + s_7) - (s_8 + s_9 + s_{10} + s_{11})] \\ W_{D35}(1) = \frac{1}{2\sqrt{2}} [(s_5 + s_6 + s_7 + s_8) - (s_9 + s_{10} + s_{11} + s_{12})] \\ W_{D36}(1) = \frac{1}{2\sqrt{2}} [(s_6 + s_7 + s_8 + s_9) - (s_{10} + s_{11} + s_{12} + s_{13})] \\ W_{D37}(1) = \frac{1}{2\sqrt{2}} [(s_7 + s_8 + s_9 + s_{10}) - (s_{11} + s_{12} + s_{13} + s_{14})] \\ W_{D38}(1) = \frac{1}{2\sqrt{2}} [(s_8 + s_9 + s_{10} + s_{11}) - (s_{12} + s_{13} + s_{14} + s_{15})] \\ W_{D31}(2) = \frac{1}{2\sqrt{2}} [(s_9 + s_{10} + s_{11} + s_{12}) - (s_{13} + s_{14} + s_{15} + s_{16})] \end{array} \right]$$

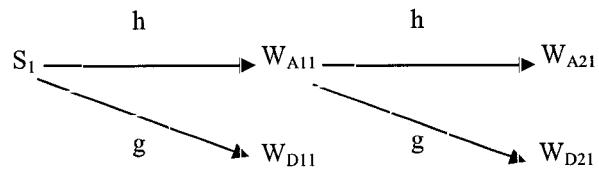
L'entropie apparente correspondante est donnée par :

$$\hat{E}W_{D3r} = -\left| \hat{W}_{D3r}(i) \right|_{i=1,..N_c}^3$$

La normalisation du vecteur $\hat{E}W_{D3r}$ mène à:

$$P\hat{E}W_{D3r} = \left[\frac{\left| \hat{E}W_{D3r}(i) \right|}{\sum_{k=1}^{N_c} \hat{E}W_{D3r}(k)} \right]_{i=1,2,\dots,N_c}$$

Décomposons le signal jusqu'au niveau 2 par l'ondelette de Haar illustrée comme suit:



La matrice de Haar H_2 correspondante pour le niveau 2 est définie par :

$$H_2 = \begin{bmatrix} a^2 & a^2 & a^2 & a^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a^2 & a^2 & a^2 & a^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a^2 & a^2 & a^2 & a^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a^2 & a^2 & a^2 & a^2 & a^2 \\ a^2 & a^2 & -a^2 & -a^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a^2 & a^2 & -a^2 & -a^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a^2 & a^2 & -a^2 & -a^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a^2 & a^2 & -a^2 & -a^2 & -a^2 \\ a & -a & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a & -a & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a & -a & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a & -a & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & -a & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & -a & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & -a & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & -a & 0 & 0 & 0 \end{bmatrix}$$

Multiplions la matrice de Haar (H_2) successivement par les signaux S_1 , S_2 , S_3 et S_4 nous obtenons alors :

$$W'_1 = H_2 S_1, W'_2 = H_2 S_2, W'_3 = H_2 S_3, \text{ et } W'_4 = H_2 S_4$$

Le vecteur $\{W_k, k = 1, 2, 3, 4\}$, représentent respectivement les coefficients des signaux $\{S_k, k = 1, 2, 3, 4\}$ résultants de ces opérations, et nous définissons :

$$W'_k = \begin{bmatrix} W_{A2k}(l); l=1,..,4 \\ W_{D2k}(l); l=1,..,4 \\ W_{D1k}(l); l=1,..,8 \end{bmatrix}; k=1,..,4,$$

avec

$$\begin{aligned} W_{D2k}(l) &= W_{D2k}(l); l=1,..,4 \\ W_{D1k}(l) &= W_{D1k}(l); l=1,..,8 \end{aligned}$$

donc nous avons :

$$W'_k = \begin{bmatrix} W_{A2k}(l); l=1,..,4 \\ W_{D2k}(l); l=1,..,4 \\ W_{D1k}(l); l=1,..,8 \end{bmatrix}; k=1,..,4$$

$$W_{A2k} = \begin{bmatrix} s_k + .. + s_{k+3} \\ s_{k+4} + .. + s_{k+7} \\ s_{k+8} + .. + s_{k+11} \\ s_{k+12} + .. + s_{k+15} \end{bmatrix}; k=1,..,4$$

Posons $s_i = 0$, pour $i \geq 17$

$$\begin{aligned} \hat{W}_{A2} &= \left[W_{Ajk}(L) = \frac{1}{2} \left(\sum_{i=k+2^j(L-1)}^{i=k+j+2^j(L-1)} s_i \right) \right]_{\substack{j=2 \\ k=1, 2, 3, 4 \\ L=1, 2, 3, 4 = (\frac{N_c}{2^j})}} \\ \hat{W}_{A2} &= [\hat{W}_{A2}(i)]_{i=1,..,N} \text{ alors } \hat{W}_{A2r} = [\hat{W}_{A2}(i)]_{i=1,..,N_c} \text{ avec } N_c = 13 \end{aligned}$$

L'entropie apparente correspondante est donnée par :

$$\hat{E}W_{A2r} = \left| \hat{W}_{A2r}(i) \right|_{i=1,..,N_c}^3$$

D'une façon plus compacte, nous pouvons écrire:

$$\hat{W}_{Aj} = \left[W_{Ajk}(L) \right]_{\substack{L=1,..,N \\ k=1,..,2^j}} \quad (4.8)$$

Où

$$W_{Ajk}(L) = \left(\frac{1}{\sqrt{2}} \right)^j \left(\sum_{i=k+2^j(L-1)}^{i=k+2^j(L-1/2)-1} s_i + \sum_{i=k+2^j(L-1/2)}^{i=k+2^jL-1} s_i \right) \quad (4.9)$$

$$\hat{E}W_{Ajk} = \left[sign(\hat{W}_{Ajk}(i)) \left| \hat{W}_{Ajk}(i) \right|^p \right]_{i=1, \dots, N_c} \quad (4.10)$$

Nous avons besoin des quatre vecteurs pour détecter les ruptures dans le signal S_1

- $\hat{E}W_{D3r} = -\left| \hat{W}_{D3r}(i) \right|_{i=1, \dots, N_c}^3$
- $P\hat{E}W_{D3r}$
- $\hat{E}W_{A2r} = \left| \hat{W}_{A2r}(i) \right|_{i=1, \dots, N_c}^3$
- $\hat{E}W_{D2r} = -\left| \hat{W}_{D2r}(i) \right|_{i=1, \dots, N_c}^3$

4.4 Application numérique de la méthode proposée

La même démarche décrite ci-dessus est effectuée à nouveau dans l'exemple numérique suivant:

Étape 1 :

$$S_1 = [95, 115, 105, 102, 108, 72, 29, 53, 61, 40, 55, 59, 85, 120, 115, 125, 96, 103, 86, 79]$$

$$S_2 = [115, 105, 102, 108, 72, 29, 53, 61, 40, 55, 59, 85, 120, 115, 125, 96, 103, 86, 79, 95]$$

$$S_3 = [105, 102, 108, 72, 29, 53, 61, 40, 55, 59, 85, 120, 115, 125, 96, 103, 86, 79, 95, 115]$$

$$S_4 = [102, 108, 72, 29, 53, 61, 40, 55, 59, 85, 120, 115, 125, 96, 103, 86, 79, 95, 115, 105].$$

4.4.1 Calcul numérique de W_1

$$W_{D11} = [-14.14, 2.12, 25.45, -16.97, 14.84, -2.82, -24.7, -7.07, -4.94, 4.94] ;$$

$$W_{D21} = [1.50, 49.00, -6.50, -17.50, 17.00] ;$$

$$W_{D31} = [54.80, -81.31] ;$$

$$W_{D41} = 4.75 ;$$

$$W_{A41} = 334.75 ;$$

4.4.2 Calcul numérique de W_2

$$W_{D12} = [7.07, -4.24, 30.40, -5.65, -10.60, -18.38, 3.53, 20.50, 12.02, -11.31] ;$$

$$W_{D22} = [5.00, -6.50, -24.50, 7.00, ; 7.50] ;$$

$$W_{D32} = [76.01, -76.72] ;$$

$$W_{D42} = -12.50 ;$$

$$W_{A42} = 335.00 ;$$

4.4.3 Calcul numérique de W_3

$$W_{D13} = [2.12, 25.45, -16.97, 14.84, -2.82, -24.74, -7.07, -4.94, 4.94, -14.14] ;$$

$$W_{D23} = [13.50, -9.50, -45.50, 20.50, -22.50] ;$$

$$W_{D33} = [72.12, -42.42] ;$$

$$W_{D43} = -47.00 ;$$

$$W_{A43} = 332.00 ;$$

4.4.4 Calcul numérique de W_4

$$W_{D_{14}} = [-4.24, 30.40, -5.65, -10.60, -18.38, 3.53, 20.50, 12.02, -11.31, 7.07];$$

$$W_{D_{24}} = [54.50, 9.50, -45.50, 16.00, -23.00];$$

$$W_{D_{34}} = [36.06, -10.96];$$

$$W_{D_{44}} = -67.25;$$

$$W_{A_{44}} = 327.25;$$

Etape 2 :

$$\hat{W}_{D_2} = [1.50, 5.00, 13.50, 54.50, 49.00, -6.50, -9.50, 9.50, -6.50, -24.50, -45.50, -45.50, \\ -17.50, 7.00, 20.50, 16.00, 17.00, 7.50, -22.50, -23.00]$$

$$\hat{W}_{D_{2r}} = [1.50, 5.00, 13.50, 54.50, 49.00, -6.50, -9.50, 9.50, -6.50, -24.50, -45.50, -45.50, \\ -17.50, 7.00, 20.50, 16.00, 17.00]$$

4.4.5 Calcul numérique des entropies apparentes

avec p=3; $\hat{E}W_{D_{2r}} = - \left[sign(\hat{W}_{D_{2r}}(i)) \left| \hat{W}_{D_{2r}}(i) \right|^p \right]_{i=1, \dots, N_c}$

L'entropie apparente est dans ce cas égale à :

$$\hat{E}W_{D_{2r}} = 1.0e+005 * [-0.0000, -0.0013, -0.0246, -1.6188, -1.1765, 0.0027, 0.0086, \\ -0.0086, 0.0027, 0.1471, 0.9420, -0.9420, 0.0536] ;$$

$$W_{D_{31}} = [54.80, -81.31]$$

$$W_{D_{32}} = [76.01, -76.72]$$

$$W_{D_{33}} = [72.12, -42.42]$$

$$W_{D_{34}} = [36.06, -10.96]$$

$$W_{D35} = [16.61, 28.63]$$

$$W_{D36} = [-8.48, 32.88]$$

$$W_{D37} = [-48.08, 22.62]$$

$$W_{D38} = [-60.10, 5.65]$$

$$\text{alors } \hat{W}_{D3} = [54.80, 76.01, 72.12, 36.06, 16.61, -8.48, -48.08, -60.10, -81.31, -76.72, \\ -42.4264, -10.96, 28.63, 32.88, 22.62, 5.65]$$

$$\hat{W}_{D3r} = [54.80, 76.01, 72.12, 36.06, 16.61, -8.48, -48.08, -60.10, -81.31, -76.72, \\ -42.4264, -10.96, 28.63]$$

$$\hat{E}W_{D3r} = 1.0e+005 * [-1.64, -4.39, -3.75, -0.46, -0.04, 0.00, 1.11, 2.17, 5.37, 4.51, 0.76, \\ 0.01, -0.2349]$$

$$P\hat{E}W_{D3r} = [0.4813, 1.2845, 1.0973, 0.1372, 0.0134, 0.0018, 0.3251, 0.6350, 1.5726, \\ 1.3207, 0.2233, 0.0039, 0.0687]$$

$$W_{A21} = [208.50, 131.00, 107.50, 222.50, 182.00]$$

$$W_{A22} = [215.00, 107.50, 119.50, 228.00, 181.50]$$

$$W_{A23} = [193.50, 91.50, 159.50, 219.50, 187.50]$$

$$W_{A24} = [155.50, 104.50, 189.50, 205.00, 197.00]$$

$$\hat{W}_{A2} = [208.50, 215.00, 193.50, 155.50, 131.00, 107.50, 91.50, 104.50, 107.50, 119.50, \\ 159.50, 189.50, 222.50, 228.00, 219.50, 205.00, 182.00, 181.50, 187.50, 197.00]$$

$$\hat{W}_{A2r} = [208.50, 215.00, 193.50, 155.50, 131.00, 107.50, 91.50, 104.50, 107.50, 119.50, \\ 159.50, 189.50, 222.50, 228.00, 219.50, 205.00, 182.00]$$

$$\hat{E}W_{A2r} = 1.0e+007 * [0.90, 0.99, 0.72, 0.37, 0.22, 0.12, 0.07, 0.11, 0.12, 0.17, 0.40, 0.68, \\ 1.10, 1.18, 1.05, 0.86, 0.60].$$

4.5 Comparaison de l'utilisation de l'ondelette et de celle de l'ondelette de Haar modifiée (OHM) pour la détection d'une rupture

Pour expliciter notre méthode de l'ondelette de Haar modifiée , nous donnons un exemple avec les séries synthétiques qui sont représentées aux figures 19, 20 et 21. La figure 19 illustre la décomposition au niveau 5 de l'ondelette de Haar d'un signal synthétique s . Les différents détails d_j ($j=1,\dots,5$) et approximation a_5 sont représentés dans la figure 19 ci-dessous, ils constituent le résultat de la décomposition du signal traité à différents niveaux de résolution. Nous observons que l'approximation a_5 du cinquième niveau d'échelle de l'ondelette de Haar donne une bonne représentation des trois états du signal s analysé.

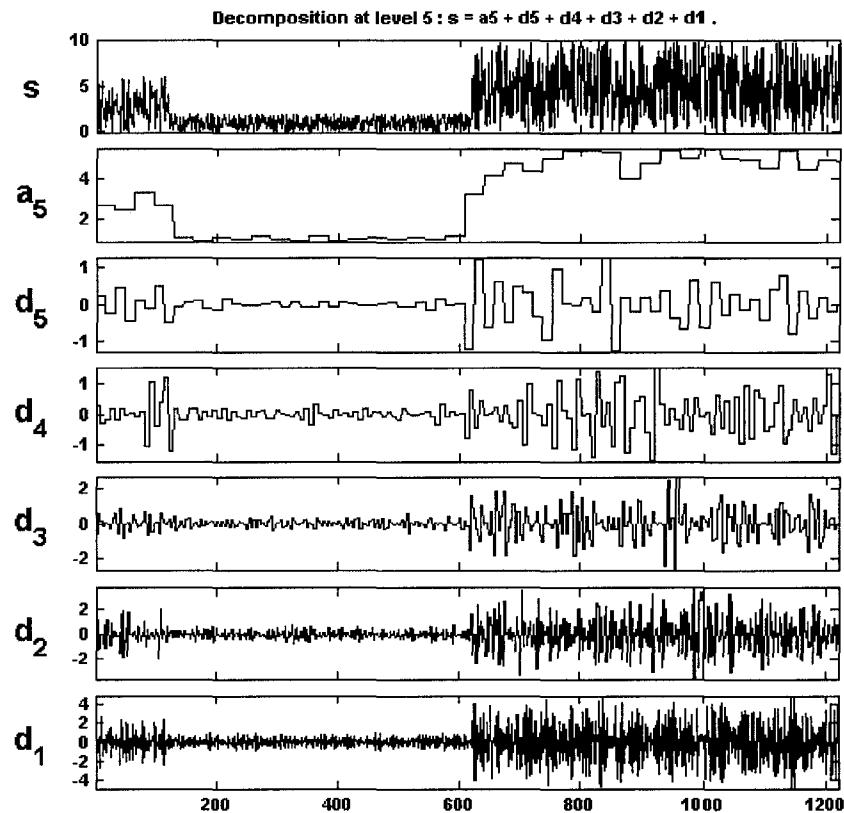


Figure 19 La décomposition en ondelette de Haar

Nous allons par contre démontrer l'inefficacité de l'ondelette de Haar pour détecter les brusques changements dans un signal aléatoire, et ceci en présentant un exemple que nous illustrons dans la figure 20. Celles-ci nous montre que la rupture n'a pas été détectée aux points 23 et 86 du signal aléatoire s sur tous les niveaux de détail et d'approximation.

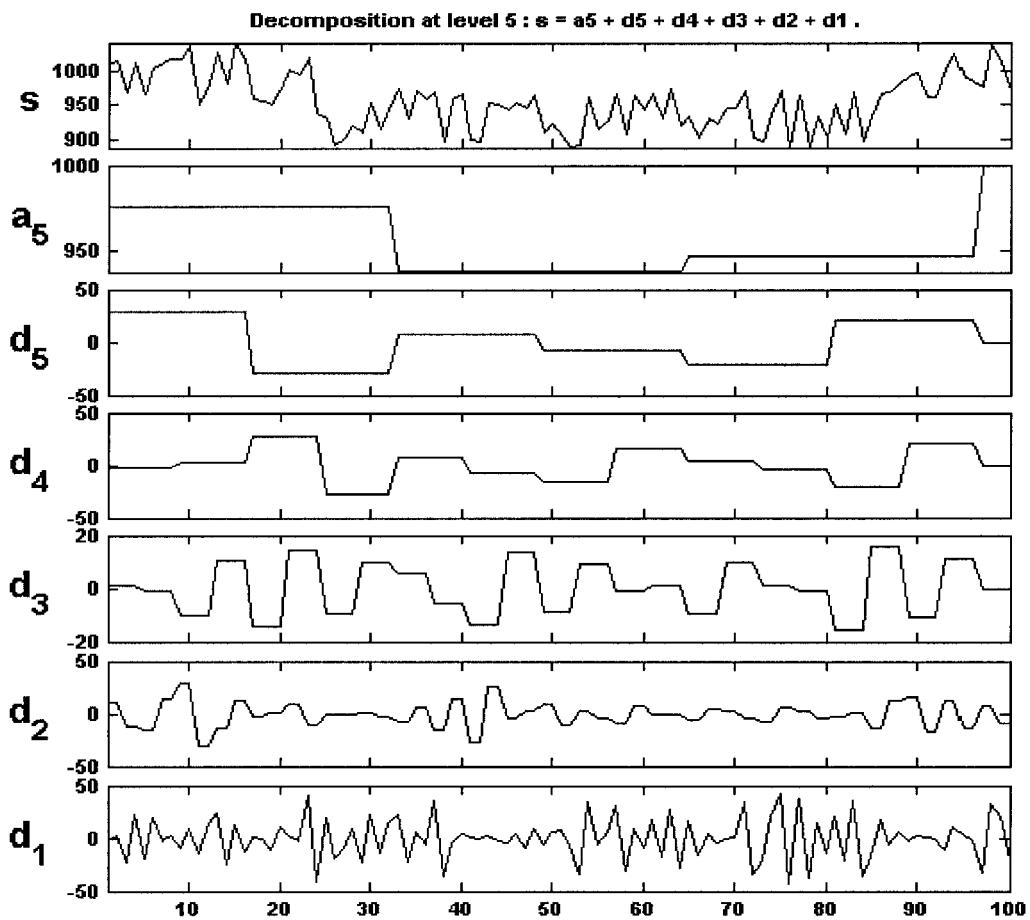


Figure 20 La décomposition en ondelette de Haar pour un signal aléatoire à 3 états

Avec le même signal s de la figure 20 notre méthode donne de meilleur résultat que celle de l'ondelette Haar, comme le montre les figures 21, 22, et 23.

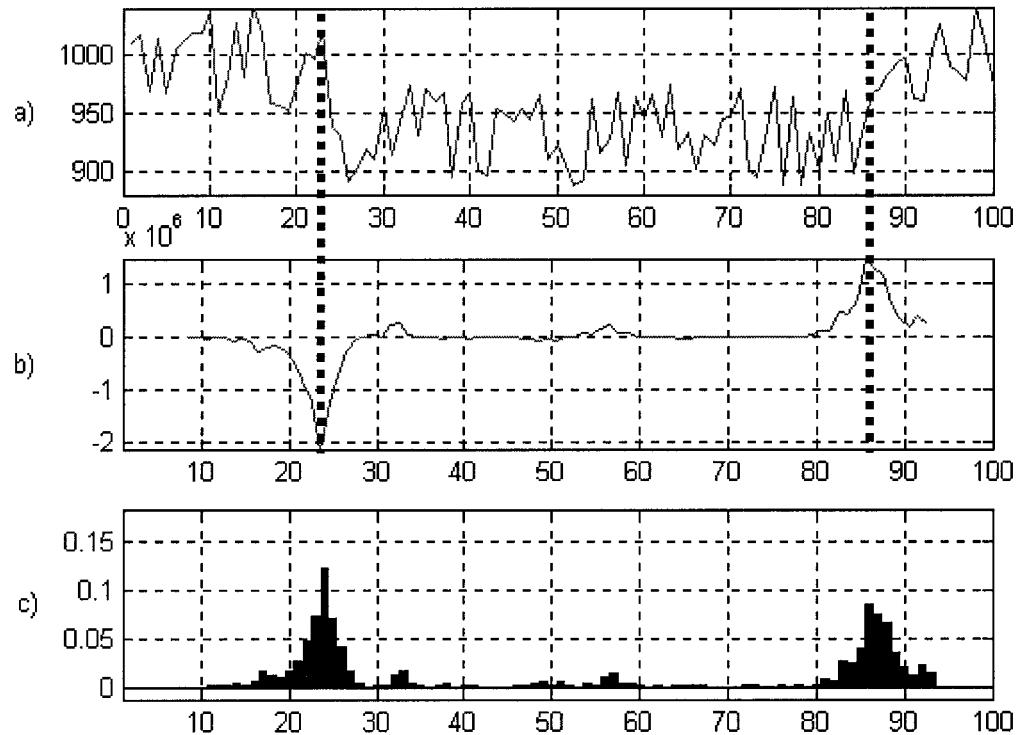


Figure 21 a) Signal aléatoire à 3 états b) Entropie apparente du détail D4 de l'OHM c) Entropie apparente normalisée du détail D4 de l'OHM

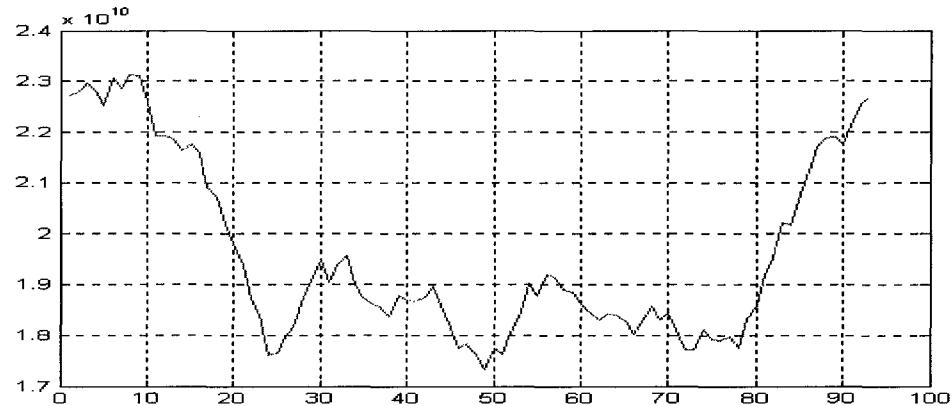


Figure 22 Entropie apparente de l'approximation A3 de l'OHM

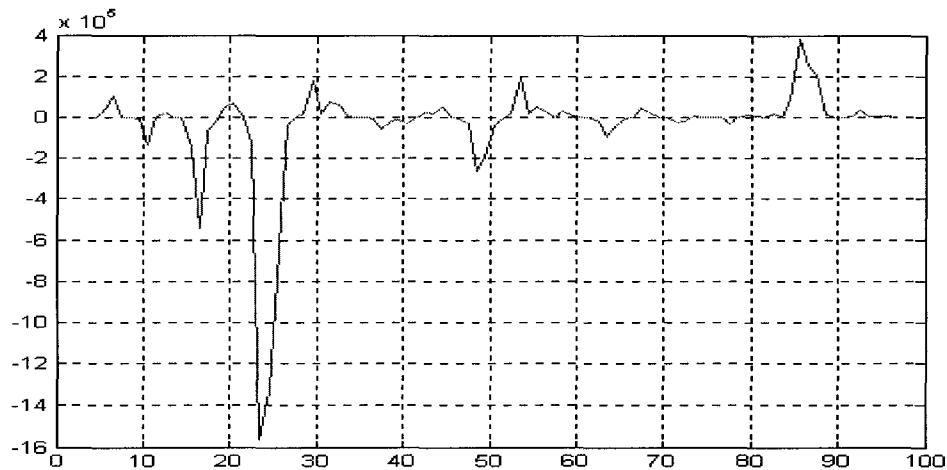


Figure 23 Entropie apparente du détail D3 de l'OHM

4.6 Conclusion

Nous avons décrit dans ce chapitre la méthode dite de l'ondelette de Haar modifiée (OHM) que nous avons développée dans le cadre de notre travail. Nous l'avons adoptée pour améliorer la capacité de détection des ruptures dans les séries chronologiques. Nous avons vu que la méthode proposée commence par la décomposition en ondelettes

de Haar du signal traité au niveau $j = 4$, car d'après l'historique des séries hydrométéorologique, on a constaté qu'un cycle d'occurrence autour de 8 ans mène à fixer le niveau de l'ondelette de Haar à $j = 4$. Nous avons ensuite refait le même traitement avec le même signal mais en le décalant à chaque fois d'un seul échantillon, nous avons ensuite réitéré ce traitement ($2^j - 1$). Nous avons ensuite formé les nouveaux coefficients d'approximations et de détails de l'ondelette de Haar modifiée (OHM), puis nous avons effectué le calcul des entropies correspondantes après avoir adéquatement défini une entropie permettant de mieux détecter les ruptures. Nous avons choisi pour la phase de normalisation les coefficients de détails du dernier niveau. Finalement nous avons visualisé le signal traité et ses coefficients d'approximation et de détails pour détecter les sécheresses et inondations qui existent dans les séries chronologiques utilisées dans ce travail.

CHAPITRE 5

EXPERIMENTATIONS ET RESULTATS

5.1 Introduction

Nous avons mis en évidence dans le chapitre précédent nos principales contributions dans le cadre de ce travail : la méthode de l'ondelette de Haar modifiée (OHM). Dans ce chapitre, nous commencerons par donner un aperçu sur la base de données utilisées dans le cadre de ce travail. Ces données représentent quelques séries chronologiques de quelques grands lacs NBS (Net Bassin Supply). Par la suite, nous présenterons l'application de quelques méthodes de détection de rupture utilisées dans nos essais expérimentaux. Nous exposerons les résultats obtenus par ces méthodes et nous donnerons nos interprétations concernant chacune d'elles. Finalement nous comparerons les différents résultats obtenus par les méthodes utilisées dans nos expérimentations.

5.2 Bases de données

La base de données sur laquelle nous avons effectué nos essais expérimentaux est composée de 5 séries chronologiques. Celles-ci représentent les débits de quelques grands lacs NBS (Net Bassin Supply) qui existent en Amérique du Nord. Chaque série est identifiée par son nom qui est celui du lac étudié, de même elle est composée de 4 principaux champs, le premier représente les données mesurées réelles du débit, le deuxième est l'année de début des données mesurées, l'année de fin de celles-ci constitue le troisième champ, et finalement le pas temporel en années de la série est donné par le quatrième champ. Dans le cadre de ce travail, nous avons analysé les séries chronologiques relatives aux grands lacs NBS qui sont : Lac Érie, Lac Michigan, Lac Ontario, Lac Supérieur, et Lac St-Clair. Toutes ces séries chronologiques ont la même année de début (1900) et la même année de fin (1999). Le débit de chaque lac est exprimé en mètre cube par seconde (m^3/s). Notons aussi que pour chaque point de la

série chronologique est en fait une donnée du débit. Celle-ci représente la moyenne de toutes les données journalières de l'année correspondante.

5.3 Application de quelques méthodes de détection de rupture à quelques séries chronologiques

Dans ce qui suit nous présenterons pour chaque lac étudié, les résultats de simulations qui correspondent à chacune des méthodes utilisées dans ce travail, en l'occurrence la méthode bayésienne de Lee-Heghinian, la méthode basée sur la procédure de segmentation de Pierre Hubert, et notre méthode de l'ondelette de Haar modifié.

5.3.1 Lac Ontario

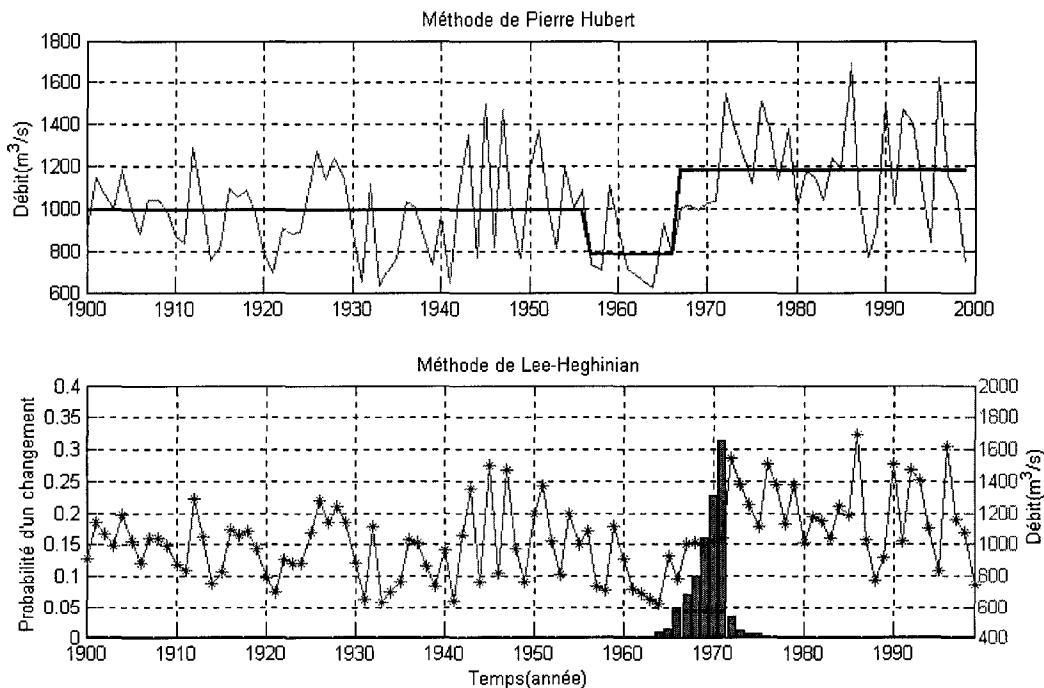


Figure 24 Méthodes de Hubert et de Lee-Heghinian appliquées au Lac Ontario

La figure 24 indique que la méthode de Hubert donne trois segments de moyennes différentes, selon son approche nous pouvons dire que sa méthode a détectée deux ruptures, une négative qui se situe entre le premier et le deuxième segment (1956-1957), et l'autre est positive qui se trouve entre le deuxième et le troisième segment (1966-1967). Le résultat de simulation de la méthode de Lee-Heghinian illustré sur la figure 24 nous montre clairement qu'elle a pu détecter une seule rupture (1970-1971) dans la série du Lac Ontario, cependant cette détection est accompagnée par une marge d'incertitude qui se traduit par l'existence de quelques probabilités de changement autour de la rupture détectée. En analysant les résultats obtenus par notre méthode de l'ondelette de Haar modifié (voir figure 25), nous remarquons que la plus grande valeur de l'entropie apparente du détail D4 normalisé correspond exactement à la rupture réelle (1970) dans la série chronologique. De plus, l'entropie apparente du détail D4 de Haar modifié considéré comme une entropie apparente ajoute une information supplémentaire sur la rupture, ainsi il permet d'indiquer le sens de celle-ci comme nous le montre la figure 25 qui nous permet de constater clairement que la rupture (1970) est positive, et ceci suivant le graphe qui correspond à la variation de l'entropie apparente du détail D4 de l'ondelette de Haar modifiée (OHM).

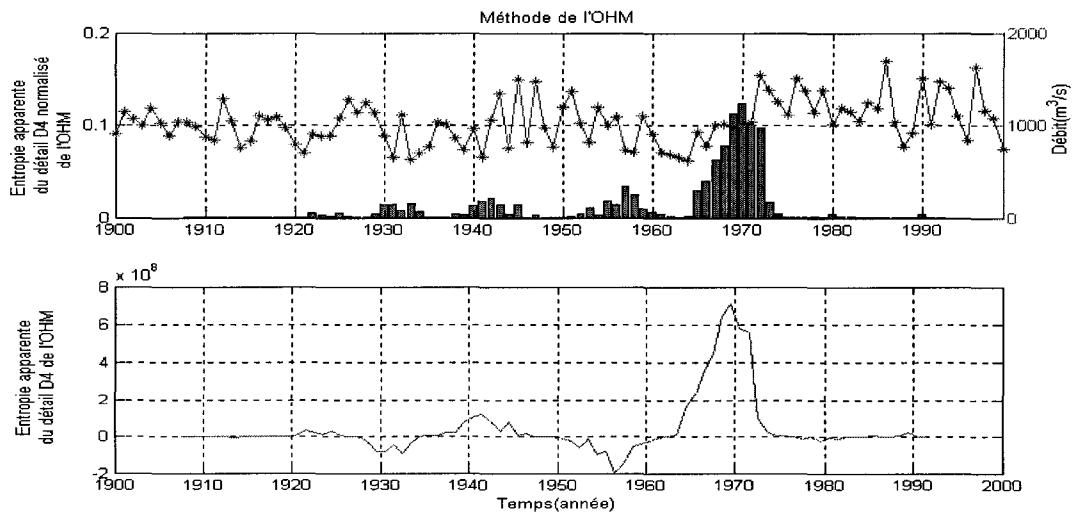


Figure 25 Méthode de l'ondelette de Haar modifiée appliquée au Lac Ontario

5.3.2 Lac Erie

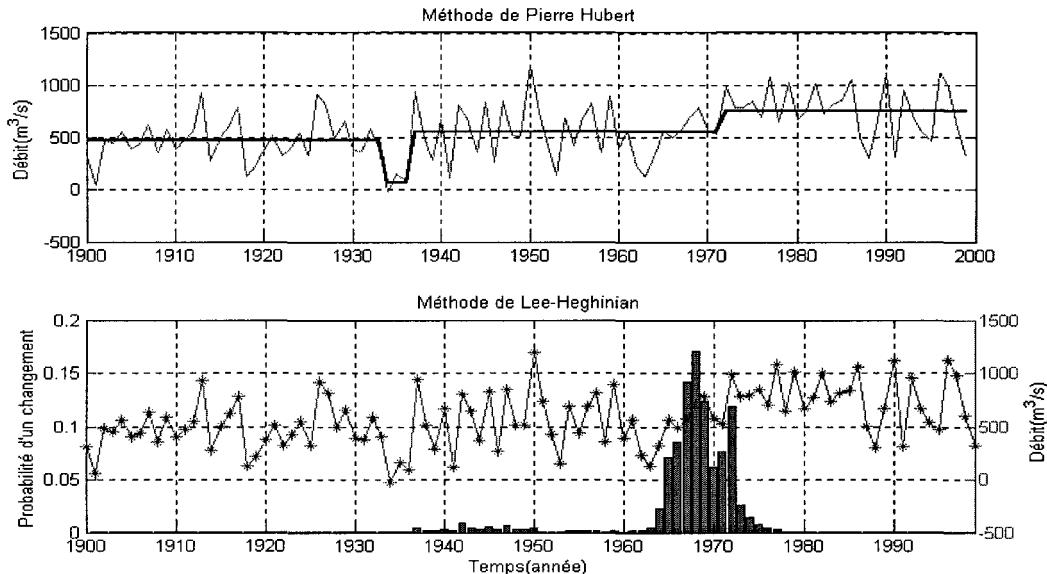


Figure 26 Méthodes de Hubert et de Lee-Heghinian appliquées au Lac Erie

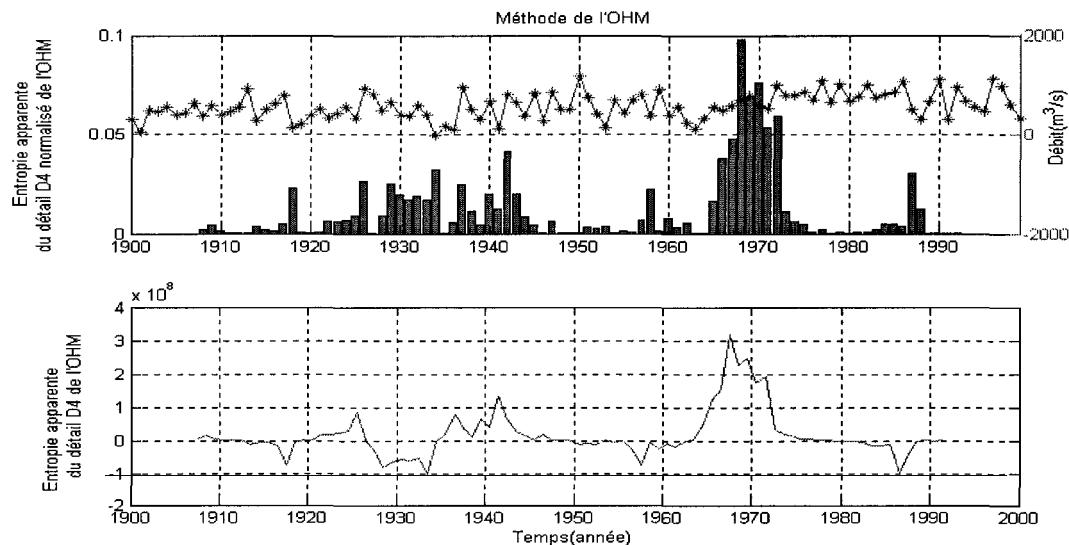


Figure 27 Méthode de l'ondelette de Haar modifiée appliquée au Lac Erie

Dans la figure 26, nous remarquons selon la méthode de Hubert l'existence de trois ruptures, la première est négative et se trouve entre 1933 et 1934, la seconde est positive et se situe entre 1936 et 1937, et finalement la troisième qui est positive aussi est localisée entre 1971 et 1972. La figure 26 présente le résultat de simulation de la méthode de Lee-Heghinian, ainsi nous pouvons voir que la rupture a eu lieu entre l'année 1968 et 1969. La figure 27 nous montre par le graphe de l'entropie apparente du détail D4 normalisé que la méthode de l'ondelette de Haar modifiée indique une rupture importante à l'année 1968. De même le graphe qui correspond à l'entropie apparente du détail D4 de Haar modifié indique que cette rupture est clairement positive.

Pour les trois lacs qui restent en l'occurrence, Lac Michigan, Lac St-Clair, et Lac Supérieur, les mêmes interprétations et analyses que celles décrites précédemment pourraient être apportées à ces lacs suivant les trois méthodes utilisées dans cette partie.

5.3.3 Lac Michigan

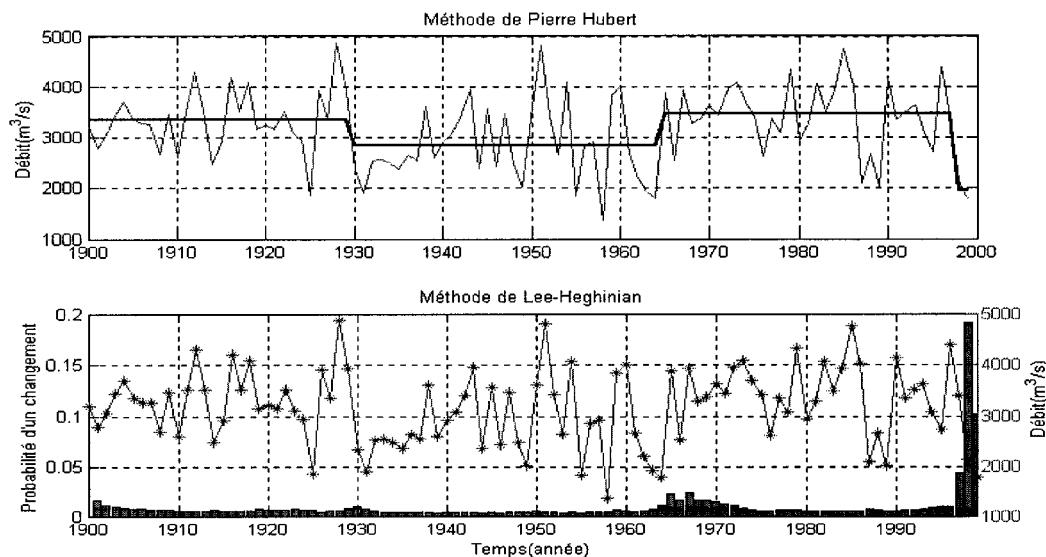


Figure 28 Méthodes de Hubert et de Lee-Heghinian appliquées au Lac Michigan

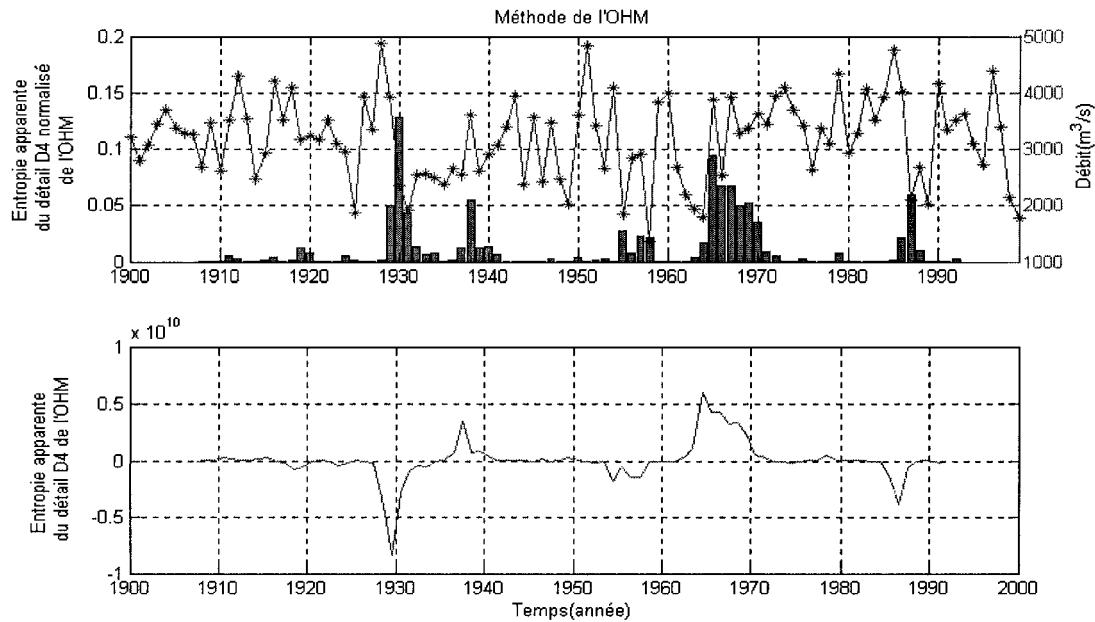


Figure 29 Méthode de l'ondelette de Haar modifiée appliquée au Lac Michigan

5.3.4 Lac St-Clair

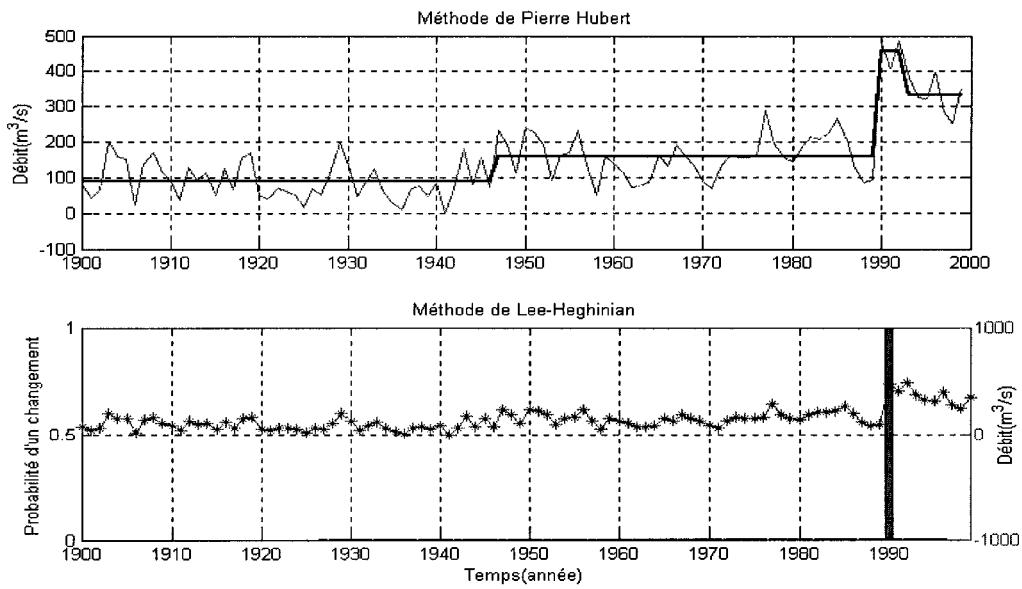


Figure 30 Méthodes de Hubert et de Lee-Heghinian appliquées au Lac St-Clair

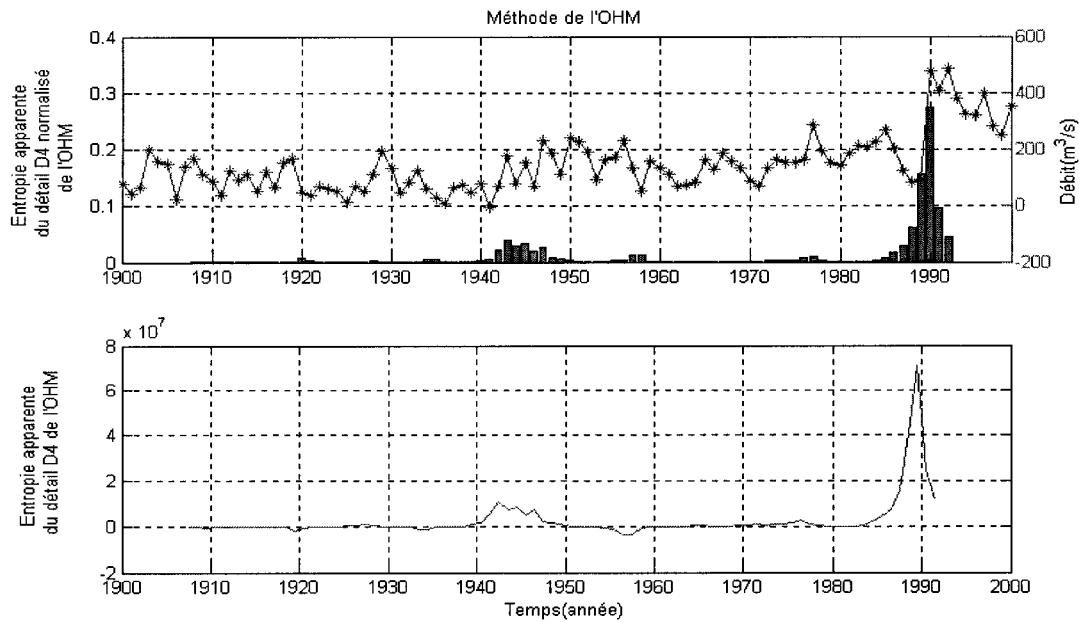


Figure 31 Méthode de l'ondelette de Haar modifiée appliquée au Lac St-Clair

5.3.5 Lac Supérieur

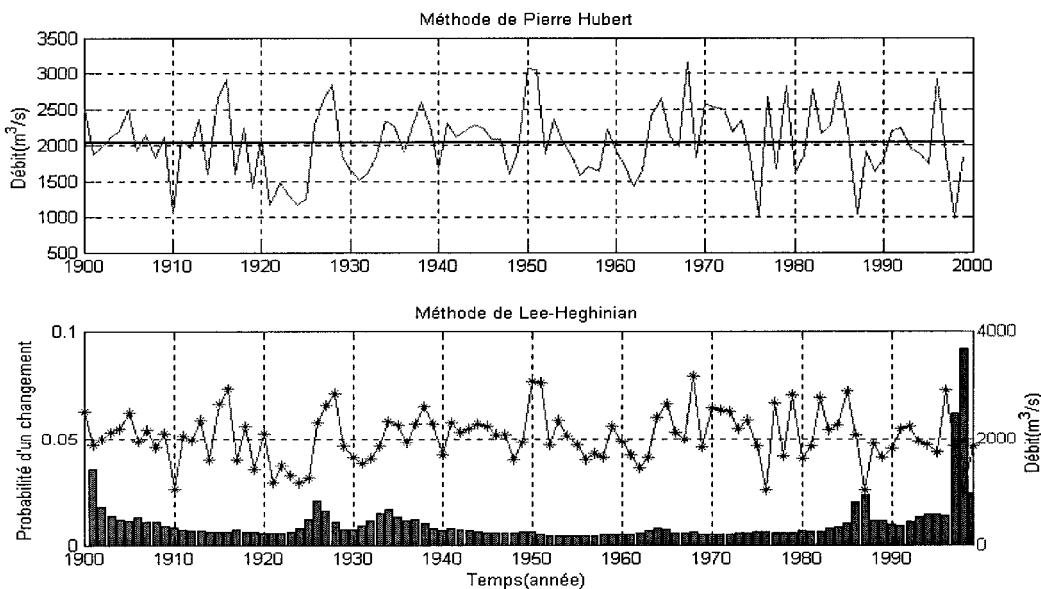


Figure 32 Méthode de Hubert et de Lee-Heghinian appliquées au Lac Supérieur

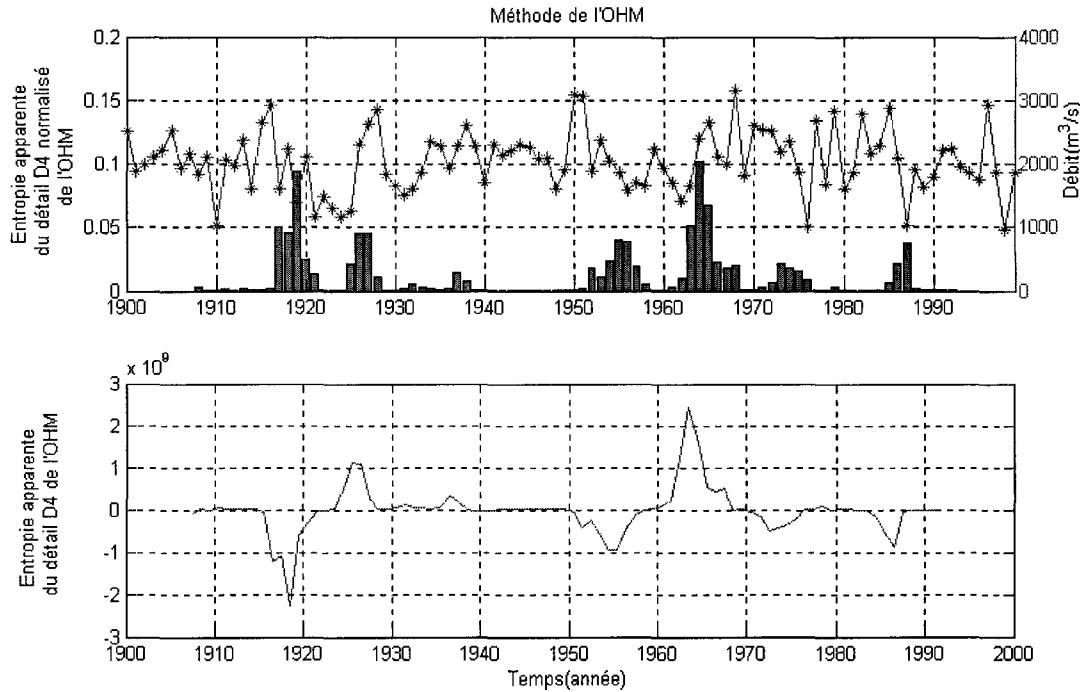


Figure 33 Méthode de l'ondelette de Haar modifiée appliquée au Lac Supérieur

5.4 Conclusion

Nous avons vu dans les résultats de simulations que la méthode de Hubert donne un nombre de segments de moyennes différentes. Entre chaque segment et le segment adjacent se situe une rupture négative ou positive. Cependant, la méthode de Hubert présente un inconvénient du fait qu'elle génère parfois un nombre important de segments qui créent par conséquent beaucoup de difficulté pour identifier le vrai changement de moyennes dans la série. La méthode de Lee-Heghinian détecte une seule rupture dans la série. Cependant, cette détection est souvent accompagnée par une marge d'incertitude qui se traduit par l'existence de quelques probabilités de changement autour de la rupture détectée. De plus, la méthode de Lee-Heghinian présente une difficulté à détecter l'existence de plusieurs ruptures dans la série. Ainsi si nous détectons par exemple une sécheresse, nous ne pouvons guère espérer détecter l'inondation et vice-versa. En

analysant les résultats obtenus par notre méthode de l'ondelette de Haar modifiée , nous remarquons que la plus grande valeur de l'entropie apparente du détail D4 normalisé correspond exactement à la rupture réelle dans la série chronologique. De plus, l'entropie apparente du détail D4 de l'ondelette de Haar modifiée ajoute une information supplémentaire sur la rupture (telle que nous l'avons définie). Elle permet d'indiquer clairement le sens de celle-ci qu'elle soit positive ou négative. Notre méthode de l'ondelette de Haar modifiée constitue un avantage remarquable du fait qu'elle permet de classifier les ruptures dans la série chronologique en fonction de leurs ordres d'occurrence. En conclusion nous constatons que notre méthode de l'ondelette de Haar modifié (OHM) permet de détecter les ruptures à des niveaux de résolution plus fines que les deux autres méthodes étudiées dans ce travail.

CONCLUSION

Dans ce travail, nous avons étudié trois méthodes de détection de rupture dans les séries chronologiques NBS. La première méthode est basée sur de segmentation de Hubert [21], la deuxième fait usage à la méthode bayésienne de Lee-Heghinian [22], et la troisième est celle que nous avons proposée dans le cadre de ce travail et que nous avons nommée méthode de l'ondelette de Haar modifiée.

Nous avons constaté que la méthode de Lee-Heghinian permet d'identifier un seul parmi plusieurs changements de moyennes dans les séries chronologiques. Par contre nous avons observé que la méthode de Hubert comporte aussi un inconvénient du fait qu'elle découpe parfois la série chronologique en un nombre très important de segments. Par contre nous avons constaté que notre méthode de l'ondelette de Haar modifiée a donné avec exactitude la rupture réelle dans la série chronologique. Nous avons aussi constaté qu'elle peut ajouter une information supplémentaire sur la rupture ce qui permet d'indiquer clairement le sens de celle-ci qu'elle soit positive ou qu'elle soit négative. De même nous avons remarqué que notre méthode comporte un avantage remarquable dans le fait qu'elle permet de classifier les ruptures dans la série chronologique en fonction de leurs ordres d'occurrence. Nous avons enfin observé que notre méthode de l'ondelette de Haar modifiée permet de détecter les ruptures à des niveaux de résolution plus fines que les deux autres méthodes étudiées.

Dans le cadre des travaux futurs, il est possible d'envisager l'utilisation d'autres types d'ondelettes symétriques comme l'est de Haar et de comparer les résultats donnés par les différents types d'ondelettes.

BIBLIOGRAPHIE

- [1] Abi-zeid, I. (1997). *La modélisation stochastique des étiages et de leurs durées en vue de l'analyse du risque*. Th.Doct.INRS-Eau/ENGREF, 278p.
- [2] Abi-zeid, I., Bobée, B. (1997). *La modélisation stochastique des étiages :Une revue bibliographique*. Rev.Sci.Eau, 459-484, 1999.
- [3] Akansu N., Ali A. Haddad Richard, (1992). *Multiresolution Signal Decompositionn, transforms – subbbands – wavelets*. Academic Press.
- [4] Beran, M.A., et J.A. Rodier (1985), *Hydrological aspects of drought*, Rapport UNESCO-WMO.
- [5] Bidwell, V. J., (1972), *Agricultural response to hydrologic drought*, Colorado State University Hydrology Paper, no. 53.,
- [6] Bootsma, A., Boisvert, J. B., Dejong, R., Baier, W., (1996). *La sécheresse et l'agriculture canadienne : une revue des moyens d'action*. Cahiers
- [7] Burrus ,C. S., R. A. Gopinath, H. Guo, (1998). *Introduction to wavelets and wavelet transforms, a primer*. Upper Saddle River, NJ (USA): Prentice Hall.
- [8] Buishand T.A. (1982), *Some mehods for testing the homogeneity of rainfall records*. Journal of hydrology, vol 58, pp. 11-27.
- [9] Buishand T.A. (1984), *Test for detecting a shift in the mean of hydrology time series*. Journal of hydrology, vol 73, pp. 51-69.
- [10] Chui, C. K. (1992).*An Introduction to Wavelets*, New York, Academic Press.
- [11] Clausen, B. , Pearson, C. P., 1995. *Regional frequency analysis of annual maximum streamflow drought*. J. Hydrol. 173, 111-130.
- [12] Coifman, R. R., and Wickerhauser, M. V., (1992). *Entropy-based algorithms for best basis selection*, IEEE Transactions on Information Theory, 38(2 pt II), pp. 713-718.
- [13] Daubechies, I., 1990. *The wavelet transform, time-frequency localization and signal analysis*, IEEE Trans. On Information Theory., 36(5), 961-1005.

- [14] Daubechies I., (1992). *Ten Lectures on Wavelets*, SIAM, Philadelphia.
- [15] Dracup, J.A, K.S. Lee, et E.G. Paulson Jr. (1980), *On the definitions of droughts*, water Resources Research, vol. 16, no. 2, pp. 297-302.
- [16] El-jabi, N., Ashkar, F., Issa, M., (1997). *Application stochastique du phénomène d'étiage*. Can. J. Civil Engng 24 (2), 191-200.
- [17] Gabor, D., (1946). *Theory of Communications*, Journal of the Institute for Electrical Engineers, pp. 429-457.
- [18] Gardner L.A. (1969). *On detecting changes in the mean of normal variates*. Ann. Math. Statist, 40, 116-126.
- [19] Gilbert Strang, Tuong Nguyen, (1996). *Wavelets and filters Banks*. Wellesley-Cambridge Press. Chapitre 1 –5.
- [20] Haar, A. (1910). *Zur Theorie der orthogonalen Funktionensysteme*, Math. Ann. 69, 331-371.
- [21] Hubert, P., Carbonnel, J.P. and Chaouche, A. (1989). *Segmentation des séries hydrométéorologiques: application à des séries de précipitations et des débits de l'Afrique de l'Ouest*, Journal of Hydrology, Vol.110, 349-367.
- [22] Lee, A.F.S., Heghinian, S.M. (1977). *A Shift Of The Mean Level In A Sequence Of Independent Normal Random Variables -A Bayesian Approach-*. Technometrics, Vol.19, n^o 4, 503-506.
- [23] Lifermann J. (1979). *Les méthodes rapides de transformation du signal*, Paris: Masson.
- [24] Lubès-niel, H., Masson, J.M., Paturel J.E., Servat É. (1998). *Variabilité climatique et statistiques. Étude par simulation de la puissance et de la robustesse de quelques tests utilisés pour vérifier l'homogénéité de chroniques*. Rev.Sci.Eau, 383-408.
- [25] Mallat S., (1989). *Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$* , Trans. Am. Math. Soc., vol. 315, N^o 1, pp. 69-87.
- [26] Mallat S., (1989). *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE, PAMI, vol. 11, N^o 7, pp. 674-693.

- [27] Mallat ,S. G., (1989). *A theory for multiresolution signal decomposition: The wavelet representation.* IEEE Transactions on pattern analysis and machine intelligence, Vol. 11 N°.7, pp.674-693.
- [28] Mallat S., (1996). *Wavelet signal Processing*, Academic Press.
- [29] Meyer Y.,(1994). *Les ondelettes et algorithmes et applications*. Armand Colin.
- [30] Millan, J. (1972), *Drought impact on regional economy*, Colorado State University Hydrology Paper, no. 55.,
- [31] Olaniran O. J. (1991). *Evidence of climatic change in Nigeria based on annual series of rainfall of different daily amounts, 1919- 1985*. Climatic change., 19, 319-341.
- [32] Ozga-Zielinska, M. (1989), *Droughts and floods – their definitions and modelling*, IAHS Publ., no. 181, pp. 313-322.
- [33] Pettit, A.N., (1979). *A nonparametric approach to the change point problem.* Applied Statistics, 28(2), 126-135.
- [34] Rioul O., , M., Vetterli, (1998). *Wavelets and Signal Processing*, IEEE SP magazine, 14-38.
- [35] Saad, B. and Gabriel, J, A. (1994). *Adaptive forecast of multi-month lake level elevations*. Canadian Journal of Civil Engineering, 21 (5), pp. 778-788.
- [36] Scheffé, M. (1959). *The Analysis of Variance*. Wiley, New York, N. Y., 477pp.
- [37] Sheng Y., (1996). *Wavelet Transform*, In: The transforms and applications handbook. Ed. By A. D. Poularikas, Boca Raton, FL (USA), The Electrical Engineering Handbook Series, pp.747-827.
- [38] Tallaksen, L.M., Madsen, H., Clausen, B., 1997. *On the definition and modeling of streamflow drought duration and deficit volume*. Hydrol. Sci. J. 42 (1), 15-33. « Sécheresse », Vol. 7, n° 4, 277-285.
- [39] Vaidyanathan P.P., (1993). *Multirate Systems and Filter Banks*, New Jersey, Prentice Hall.

- [40] Yevjevich, V. (1967), *An objective approach to definitions and investigations of continental hydrologic drought*, Colorado State University Hydrology Paper, no. 23.,
- [41] Zelenhasic, E. Salvai, A., (1987). *A method of streamflow drought analysis*. Water Resour. Res. 23 (1), 156- 168.