

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de
Recherche Scientifique



المدرسة الوطنية المتعددة الفتيات
Ecole Nationale Polytechnique

Ecole Nationale Polytechnique
Département de Génie Industriel

Mémoire du projet de fin d'études d'ingénieur

Thème

**Etude comparative des modèles de prévision ARMA et
des réseaux de neurones artificiels**

Application : prévision à moyen terme de la charge électrique

Présenté par :

M. H.A.RAHAL

Dirigé par :

Mlle. N.ABOUN

Promotion : Juin 2006

DEDICACES

A ma mère Nadia et à mon père Amokrane,

A mes sœurs : Lina et Hanane,

A mes amis Anisse, Moh, Willy,

A Faicall pour sa précieuse aide,

A tous ceux que j'ai connu durant ma formation,

A tous ceux pour qui je compte et qui comptent pour moi,

A ma très chère Rym,

Je dédie ce travail.

REMERCIEMENTS

Le travail présenté dans ce mémoire a été dirigé par Mme ABOUN, Directrice du Département Génie Industriel. Je lui adresse toute ma gratitude pour la confiance qu'elle m'a toujours accordé, pour son soutien permanent, sa disponibilité, son engagement, sa patience, et qui m'a encouragé et aidé, sans défaillances, pendant toutes ces années.

Je remercie Mme MOUHAMMEDI, directrice de la stratégie à la Société Algérienne de l'Electricité et du Gaz (SONELGAZ) pour son aide et sa disponibilité tout au long de ce projet.

Je remercie M. HOUARI pour son aide inestimable, son soutien continu et sa disponibilité.

Je remercie M. BERMAD pour toute l'aide qu'il m'a apporté à la réalisation de ce projet, et pour toutes les fois où il a répondu patiemment à mes questions.

Je remercie tous les professeurs du département de Génie industriel de m'avoir formé, de m'avoir appris la rigueur et le goût du travail bien fait.

Mes remerciements envers tous ceux qui m'ont accordé leur soutien, tant par leur gentillesse que par leur dévouement et à toutes les personnes qui de près ou de loin m'ont aidé et encouragé.

Je remercie mes parents qui m'ont toujours témoigné tant d'affection, qui m'ont toujours soutenu et aidé et que je ne remercierais jamais assez.

ملخص:

الهدف من هذه الدراسة هو تأسيس تنبؤ لكمية الكهرباء و هذا باستعمال طريقتين مختلفتين: في الطريقة الاولى، سوف نستعمل وسيلة تنبؤ تقليدية مستعملة عادة لهذا النوع من المعطيات: طريقة بوكس و جنكس. خلال الطريقة الثانية سوف نستعمل شبكة العصبونيات الاصطناعية. في المرحلة الاولى من هذا البحث سوف نعرف مختلف طرق التنبؤ الكلاسيكية و بالأخص طريقة بوكس و جنكس بعد ذلك سوف نتعرف على المبادئ و الأسس لشبكة العصبونيات. في المرحلة الثانية من هذا العمل، سوف نقوم بتطبيق و ذلك بإنشاء مختلف تمثيلي التنبؤ. أخيرا سنختم هذا البحث بمقارنة نتائج التمثيلات.

الكلمات المفتاحية: تنبؤ، بوكس و جنكس، أرما، الشبكات العصبونيات الاصطناعية، التمثيل، كمية الكهرباء.

Résumé :

L'objectif poursuivi à travers cette étude est d'établir une prévision de la charge électrique et ce en utilisant deux méthodes différentes. Dans la première, nous ferons usage d'une méthode de prévision traditionnellement utilisée pour ce type de séries temporelles : la méthode de Box et Jenkins. Dans la seconde, nous utiliserons les réseaux de neurones artificiels.

Ainsi, notre travail consistera dans un premier temps à prendre connaissance des différentes méthodes de prévision classiques plus particulièrement celle de Box et Jenkins. Puis, nous prendrons connaissance des fondements et principes des réseaux de neurones. Dans un second temps, nous effectuerons notre application en concevant nos différents modèles de prévision. Nous conclurons par une étude comparative des résultats.

Mots clés : Prévision, Box et Jenkins, ARMA, réseau de neurones artificiels, modèle, charge électrique.

Abstract :

In this paper we present an accurate model to forecast electricity load using two different methods. In the first one, we will use a classical method of forecasting which is Box and Jenkins method. In the second one, we will use Artificial Neural Networks.

Thus, our work will consist, in a first time, on acquainting the reader with the different classical methods for time series forecasting and particularly Box & Jenkins models. Then, we will familiarize with Artificial Neural Network theory.

In a second time, we will make an application by designing our different forecasting models. We will conclude by making a comparative study of our results.

Keywords: forecast, Box & Jenkins, ARMA, artificial neural network, model, electricity load.

TABLE DES MATIERES

INTRODUCTION GÉNÉRALE	1
PROBLEMATIQUE	3
PARTIE I : PARTIE THEORIQUE	
CHAPITRE I : LES METHODES DE PREVISION	6
I. LA PREVISION	6
II. METHODES DE PREVISION DANS LES CHRONIQUES	7
III. PREVISION PAR LA METHODE DE BOX & JENKINS	9
1. Définitions et concepts	10
a. Processus stationnaire au sens strict	10
b. Processus Bruit Blanc (White noise)	10
c. Fonction d'autocorrélation et fonction d'autocorrélation partielle	10
▪ Fonction d'autocorrélation	10
▪ Fonction d'autocorrélation partielle (FACP)	11
d. Le processus autorégressif AR (p)	12
▪ Définition	12
▪ Les équations de Yule-Walker	12
▪ Fonction d'autocorrélation partielle	13
e. Le processus moyenne mobile MA (q)	14
▪ Définition	14
▪ Fonction d'autocovariance	14
f. Le processus ARMA (p, q)	15
▪ Généralités et définitions	15
▪ Les autocovariances	16
▪ Caractérisation des ordres p et q	16
g. Le processus ARIMA (p,d,q)	18
h. Les modèles saisonniers	18
2. La méthode de box et Jenkins	20
a. Description de la méthode	20
b. Extension du modèle ARIMA	28
CHAPITRE II : LES RESEAUX DE NEURONES ARTIFICIELS	30
I. MODELE MATHEMATIQUE	31
1. Composant : le neurone formel (artificiel)	32
2. Comportement	33
3. Les variables descriptives	33
4. Structure de connexions	33
a. Réseau à couches	34
b. Réseau à connexions locales	35
c. Réseau à connexions récurrentes	35
d. Réseau à connexion complète	36
II. PRINCIPE DE BASE	38
III. PROPRIETES DES RESEAUX DE NEURONES	39
1. Approximation universelle (capacité de généralisation)	39
2. L'apprentissage des réseaux de neurones	39

a. L'apprentissage supervisé	40
b. L'apprentissage non supervisé	41
IV. LES DIFFERENTS TYPES DE RESEAUX DE NEURONES	42
1. Les réseaux « feed-forward »	42
a. Le Perceptron mono-couche	42
b. Adaline, Madaline	44
c. Le Perceptron multicouches (PMC)	45
d. Les réseaux à fonction radiale	46
2. Les réseaux « feed-back »	47
a. Les cartes auto-organisatrices de Kohonen	47
b. Les réseaux de Hopfield	47
c. Les ART	48
V. METHODOLOGIE DE CONCEPTION	48
 PARTIE II : PARTIE PRATIQUE	
 CHAPITRE III : COLLECTE ET TRAITEMENT DES DONNEES	
I. SYNTHESE DES TRAVAUX	51
II. SELECTION DES ENTREES	54
1. Principe de la méthode	58
2. Application de la méthode de régression pas à pas	59
 CHAPITRE IV : CONCEPTION DES MODELES DE PREVISION	
I. APPLICATION DE LA METHODE DE BOX ET JENKINS	63
1. Identification	64
2. Estimation des paramètres et vérification	66
3. Prévision	70
II. APPLICATION DES RESEAUX DE NEURONES ARTIFICIELS	73
1. conception du réseau	74
a. Traitement des variables	74
b. Echantillonnage des variables	74
c. Propriété du réseau	75
d. L'apprentissage	76
2. La détermination du réseau optimal	79
3. La prévision	81
III. COMPARAISON DES RESULTATS	83
 CONCLUSION GENERALE	
 ANNEXES	
Annexe 1. Processus stationnaire du second ordre	89
Annexe 2. Processus à accroissements indépendants	90
Annexe 3. Les opérateurs retard et différence	91
Annexe 4. La méthode du coin	92
Annexe 5. Estimation du maximum de vraisemblance	93
Annexe 6. généralités sur les réseaux de neurones artificiels	95
Annexe 7. Fonctions de transfert	99
Annexe 8. Les différentes règles d'apprentissage	100
Annexe 9. Algorithme de rétropropagation de l'erreur	105
Annexe 10. Tableaux de données	106

Annexe 11. Graphique de l'évolution des différents facteurs	110
Annexe 12. Fonction d'autocorrélation des résidus	111
Annexe 13. Résultats d'exécution du programme	112
Annexe 14. Détails du programme développé sous MATLAB	113
BIBLIOGRAPHIE	115

LISTE DES FIGURES

Figure I.1	Organigramme des méthodes de prévision	7
Figure I.2	Organigramme de la méthode de Box & Jenkins	22
Figure II.1	Neurone formel	32
Figure II.2	Réseau à couches	35
Figure II.3	Réseau à connexions locales	35
Figure II.4	Réseau à connexions récurrentes	36
Figure II.5	Réseau à connexions complètes	36
Figure II.6	Structure du perceptron linéaire	43
Figure II.7	Structure du réseau ADALINE	44
Figure II.8	Architecture d'un réseau multicouches	46
Figure III.1	Evolution des puissances mensuelles appelées maximales	54
Figure IV.1	Fonctions d'autocorrélations et d'autocorrélations partielles	65
Figure IV.2	Fonctions d'autocorrélations et d'autocorrélations partielles après différenciation saisonnière	65
Figure IV.3	Fonctions d'autocorrélations et d'autocorrélations partielles après différenciation	65
Figure IV.4	Valeurs observées et simulées de l'évolution de la charge électrique	72
Figure IV.5	Résultats de simulation après 5 itérations	77
Figure IV.6	Résultats de simulation après 20 itérations	78
Figure IV.7	Résultats de simulation après 100 itérations	78
Figure IV.8	Performances du réseau optimal	81
Figure IV.9	Valeurs observées et simulées de l'évolution de la charge électrique	82
Figure IV.10	Résultats simulés et prévisions par les deux méthodes	83

LISTE DES TABLEAUX

Tableau I.1	Valeur des déterminants	17
Tableau III.1	Matrice des corrélations	60
Tableau III.2	Variables introduites/éliminées	60
Tableau III.3	Synthèse de la sélection des variables	61
Tableau III.4	Paramètres du modèle	61
Tableau III.5	Coefficients normalisés	62
Tableau IV.1	Résultats de l'estimation du modèle $(1,1,1)(1,1,1)_{12}$	67
Tableau IV.2	Résultats de l'estimation du modèle $(0,1,1)(1,1,1)_{12}$	68
Tableau IV.3	Résultats de l'estimation du modèle $(0,1,1)(0,1,1)_{12}$	69
Tableau IV.4	Test de Box-Pierce	70
Tableau IV.5	Récapitulatifs des performances 1	72
Tableau IV.6	Résultats d'exécutions	80
Tableau IV.7	Récapitulatifs des performances 2	82
Tableau IV.8	Comparatif des résultats	83

Introduction générale

Depuis quelques années, il est accordé une attention de plus en plus importante à l'amélioration de la prise de décision. L'un des aspects fondamentaux de cette amélioration consiste à distinguer les différents éléments qui composent le processus de prise de décision. Ceci permettrait de les maîtriser individuellement dans un premier temps, puis globalement dans un second temps.

La prévision est identifiée comme étant l'un des éléments clés de ce processus. La prise de décision repose en effet sur des prévisions. La gestion d'un stock, d'une entreprise, d'un portefeuille ou de tout autre système dynamique repose en partie sur la faculté du décideur à faire des projections pertinentes de la situation présente.

Ainsi, tout manager doit être capable de prédire les circonstances qui entourent une situation particulière et de prendre la bonne décision qui se rapporte à cette situation. Une saine prise de décision est la conséquence d'une bonne prévision. La prévision étant une anticipation du futur, elle permet de faciliter la gestion de la production et de venir à bout d'éventuelles anomalies en les prévoyant. Son importance a entraîné le développement de méthodes conduisant à des résultats plus objectifs et plus sûrs.

Les méthodes qui permettent ce pré jugement ont considérablement évolué ces derniers temps. Le développement des approches scientifiques, la concurrence et la complexité croissante justifient ces progrès dans la théorie et la pratique.

Ainsi, sans pour autant délaisser les méthodes classiques, plusieurs recherches se sont orientées vers le domaine de l'intelligence artificielle, et cela s'est traduit, entre autres, par des avancées dans l'application des réseaux de neurones artificiels.

En effet, dans un premier temps, les réseaux de neurones ont été utilisés pour des tâches de reconnaissance des formes et de classification. Mais leur capacité à découvrir des formes récurrentes dans les séries dont la complexité pourrait échapper aux modèles classiques, les a rendus, dans un second temps, intéressants à exploiter dans le domaine de l'aide à la décision.

En partant de ce constat, nous nous proposons d'établir à travers ce présent travail une comparaison entre cette nouvelle approche de prévision et les méthodes prévisionnelles plus classiques. Le but réside à trouver le seuil au-delà duquel il est possible et utile de remplacer les outils de prévision existants par les réseaux de neurones. Ceci se concrétisera lors d'une application dans le cadre de la consommation électrique tout particulièrement.

Notre étude aborde ce sujet à travers deux parties :

La première partie est théorique. Nous redéfinirons tout d'abord la problématique qui fait l'objet de notre travail. Cette première partie comporte deux chapitres :

Le chapitre I vise à énumérer les différentes méthodes de prévision, leurs différents contextes d'utilisations ainsi que leurs limites. Nous développerons par la suite les fondements théoriques de l'une des méthodes les plus pertinentes pour notre étude, et qui représente surtout l'une des plus utilisées dans l'analyse de données temporelles.

Le chapitre II est consacré à la présentation des fondements théoriques, des propriétés fondamentales et des principaux types de réseaux de neurones.

La deuxième partie de notre travail est la partie pratique. Elle aborde la mise en œuvre d'un modèle de prévision selon les différentes démarches choisies, et ce, à travers deux chapitres. Nous justifierons, tout d'abord, dans le chapitre III le choix des méthodes utilisées par une synthèse de différents travaux effectués dans le domaines. Cette synthèse nous permettra d'aborder la première phase de conception du modèle de prévision : la collecte et le traitement des données relatives à l'évolution de la consommation d'électricité ainsi que les facteurs influençant ce phénomène

Le chapitre IV de ce document traite des démarches suivies pour la conception d'un modèle de prévision par les méthodes classiques puis par l'approche neuronale. Par la suite, nous essayerons de confronter les résultats de chacune des approches adoptées, d'établir une comparaison des résultats obtenus et d'en tirer les conclusions qui en découlent.

Enfin, en conclusion de ce document, nous présentons l'apport ainsi que les limites de chaque démarche.

Problématique

Nous assistons ces dernières années à une augmentation importante de la consommation d'électricité. Les centrales de production de l'énergie électrique sont continuellement confrontées aux diverses décisions à entreprendre pour une meilleure exploitation des capacités de production et ce, afin de répondre aux besoins des consommateurs.

Ainsi, pour une pleine satisfaction de la demande et une exploitation efficace, il faudrait mettre à niveau les ressources énergétiques nécessaires en tenant compte des influences environnementales.

La société nationale d'électricité et du gaz SONELGAZ a pour missions principales la production et la distribution de l'électricité à l'échelle nationale. SONELGAZ se doit d'assurer un équilibre constant entre la production et la consommation d'électricité.

Pour assurer cet équilibre, elle se doit de connaître continuellement la quantité d'énergie nécessaire à fournir pour couvrir la charge maximale pour une période donnée.

La prévision de la demande énergétique s'avère donc stratégique et ce, pour plusieurs raisons :

- La décision d'investir dans de nouvelles infrastructures dépend fortement de cette prévision.
- Ces prévisions permettront de vérifier et d'adapter les programmes de production de l'ensemble des unités afin d'assurer un équilibre constant production-consommation.
- Offrir aux consommateurs une meilleure qualité de service en maintenant une certaine stabilité du réseau et en évitant les surcharges.
- L'énergie électrique n'est pratiquement pas stockable directement sous forme d'électricité
- La SONELGAZ étant pour l'instant le seul producteur et distributeur sur le marché national algérien.

Par ailleurs, l'aspect principal d'une situation de prise de décision pour laquelle on doit établir une prévision est *l'horizon temporel* :

- Prévision à long terme ;
- Prévision à moyen terme ;
- Prévision à court terme.

Aussi, la consommation électrique dépend de plusieurs facteurs très différents : variables climatiques, macro-économiques,...etc. Selon l'horizon de prévision, les facteurs nécessaires à la modélisation de la demande électrique diffèrent, et dans certains cas les données relatives à ces facteurs ne peuvent exister, ce qui rend la conception d'un tel modèle très complexe.

L'analyse des séries chronologiques est une méthode de construction de modèle mathématiques généralement très efficace pour résoudre des problèmes complexes. Ces modèles sont utilisés pour identifier et prévoir le comportement d'un phénomène représenté par une séquence d'observations. Mais l'élaboration d'un tel modèle, prenant en compte d'éventuelles relations non linéaires entre les différents facteurs, s'avère très difficile.

Cependant, les réseaux de neurones ont été efficacement utilisés pour résoudre une grande variété de problèmes de ce type. Le succès de leur application aux séries temporelles réside dans leur capacité à approximer n'importe quel type de fonctions.

C'est enfin dans l'objectif de modéliser avec la plus grande précision la demande en énergie électrique, que nous allons développer un modèle de prévision à moyen terme en utilisant l'une des méthodes statistiques classiques ainsi que les réseaux de neurones artificiels.

En effet, La prévision à moyen terme est stratégiquement nécessaire pour palier à d'éventuels changements de probabilité de distribution de certains facteurs, et non pas à des événements ponctuels. Et cet horizon se prête parfaitement à l'objectif de notre étude qui consiste de transformer à l'aide de ces méthodes les prévisions climatiques et socio-économiques en prévisions énergétiques.

Ainsi, après avoir bien défini l'objectif de notre étude, il est nécessaire de développer les fondements théoriques des différentes méthodes que nous allons utiliser.

Chapitre I : Les méthodes classiques de prévision

Dans ce chapitre, nous allons présenter un ensemble de méthodes prévisionnelles. Le but n'est pas de faire un inventaire détaillé et précis des méthodes de prévisions. Ce ne serait d'ailleurs pas possible, étant donné la grande diversité des méthodes et l'importance de la recherche dans ce domaine. L'objet du présent chapitre est plutôt de donner un aperçu des différents modèles et techniques existants.

Mais tout d'abord, essayons de traiter brièvement du thème central de notre étude, et de recadrer le but de notre travail : la prévision.

I. La prévision : [BOU 98]

Une prévision vise à déterminer un événement futur à partir du regroupement systématique de données quantitatives. Le thème central de la prévision étant l'hypothèse selon laquelle on peut trouver un certain événement futur à partir de ce qui s'est passé auparavant Elle comporte généralement cinq (5) étapes :

- Conception et choix d'une théorie économique.
- Choix d'une méthode de traitements stratégiques.
- Rassemblement d'informations quantitatives appropriées.
- Elaboration des prévisions.
- Révision des prévisions en fonction des informations qualitatives et des jugements des prévisionnistes.

Le gestionnaire des opérations doit effectuer des prévisions pour la majorité des décisions à prendre. La prévision s'avère utile dans une multitude de contextes, parmi lesquels on citera :

- La planification de la stratégie globale ;
- La répartition des ressources internes ;
- La décision de l'accroissement de la capacité ;
- La détermination de la nature et de la quantité des stocks ;
- Le choix des technologies futures.

Pour répondre aux diverses situations où des prévisions sont nécessaires, un certain nombre de méthodes ou techniques a été développé. Ces modèles et techniques, aussi différents soient ils, ont en commun un même point de départ, se traduisant par une analyse au préalable d'une base de données appelée série chronologique ou série temporelle.

« Une série temporelle ou encore chronique est une succession d'observations au cours du temps, à des périodes ou à des instants donnés, représentant un phénomène économique ou autre ». [BOU 98]

II. Méthodes de prévision dans les chroniques : [DEG 03]

Les méthodes de prévision ont considérablement évolué dans les entreprises et les administrations au cours des deux dernières décennies. La complexité et la concurrence croissante de l'environnement des affaires expliquent ces progrès dans la théorie et dans la pratique. Des entreprises de toutes dimensions jugent nécessaire d'effectuer des prévisions dans les domaines où le hasard affecte leurs décisions et leurs résultats. Nous pouvons alors mettre en place des classes de techniques de prévision suivant l'organigramme ci-dessous :

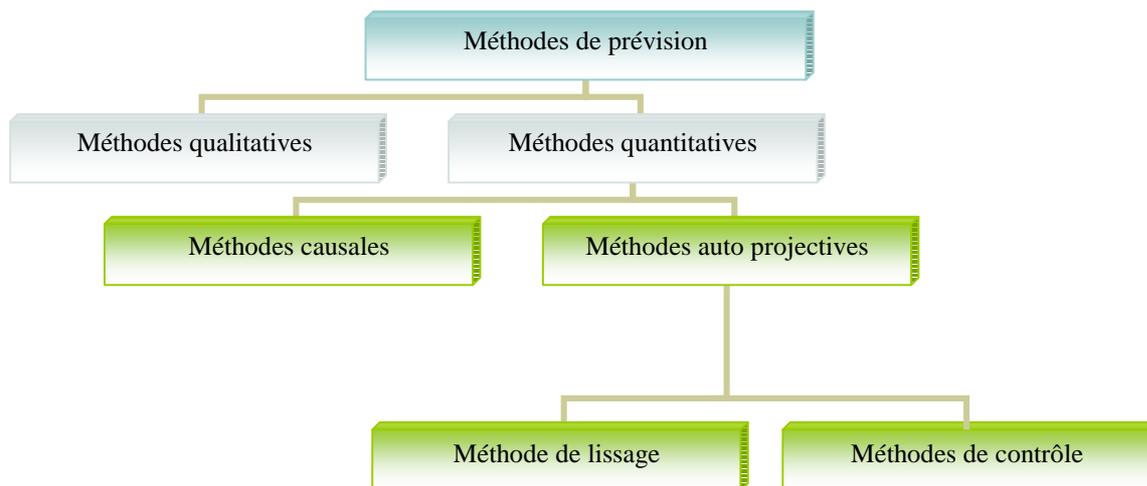


Figure.I.1- Organigramme des méthodes de prévision.

- les principaux critères de choix d'une méthode de prévision : [BOU 98]

Le choix d'une méthode de prévision dépend de plusieurs facteurs. Ce choix doit s'opérer en cherchant à minimiser un coût, mais il faut tenir compte également du type de chronique et de la finalité du traitement qui jouent souvent le rôle de contraintes dans cette recherche d'optimum.

- Le **coût** de traitement qui dépend de deux facteurs : d'une part l'algorithme de calcul inhérent à la technique de prévision choisie et de ce point de vue, certaines techniques impliquent l'usage d'ordinateurs puissants, et d'autre part, la longueur de l'historique qu'il faut conserver, ce qui se traduit par un coût de stockage des données et de traitement plus ou moins important. Ces deux facteurs ne sont pas indépendants, les techniques qui utilisent le maximum d'informations sont en général celles qui utilisent les algorithmes les plus complexes. Il ne faut jamais perdre de vue qu'un arbitrage doit toujours être effectué entre le coût de construction et de traitement d'une information, et les avantages que l'on en retire, et qu'il peut même être moins onéreux de ne pas traiter d'informations si l'activité peut être facilement régulée par un stock dont le coût de possession est négligeable.
- Le **type de série chronologique** : les techniques de prévision ne sont pas universelles, plusieurs paramètres sont à prendre en compte dans le choix d'une technique : longueur de l'historique possédé, existence d'une saisonnalité, importance relative des trois composantes d'une série.
- La **finalité du traitement** : certaines techniques de prévision ne sont appropriées que pour le court terme et fournissent des résultats catastrophiques si l'on tente de les utiliser pour le moyen ou long terme.

Il est en effet fort important de prendre en considération ces critères, qui, selon l'importance de chacun, peuvent modifier le choix de la méthode retenue.

Après examen des critères de choix, et après avoir passé en revue les différentes techniques de prévision (Un traitement complet des différentes méthodes de prévision peut être trouvé dans [BOU 98], [ECO 04] et [DEG 03]) nous avons arrêté notre choix, dans le cadre de notre étude, aux méthodes de Box et Jenkins (les modèles ARMA). Ces modèles présentent un avantage certain face aux autres méthodes quantitatives utilisées pour la prévision :

- Elles sont applicables lorsque la série présente une saisonnalité et/ou une tendance ;
- Les modèles ARMA se prêtent tout particulièrement aux prévisions à moyen terme ;
- Le coût d'élaboration de ces modèles est relativement faible. De nos jours, plusieurs logiciels prennent en charge ce type de modèles.

G.Box et G.Jenkins visent à formuler un modèle permettant de représenter une série. L'idée principale est le concept de parcimonie, ou de la minimisation du nombre de paramètres. En pratique, ces derniers étant inconnus, ils sont donc remplacés par leur valeur estimée : plus il y a de paramètres, plus nombreuses sont les chances de se tromper.

On propose, dans ce qui suit, de développer en détail les principes et concepts de la méthodologie de Box & Jenkins.

III. Prévision par la méthode de Box & Jenkins

Pour toutes les méthodes de prévision, il s'agit d'isoler au mieux la loi fondamentale sous-jacente représentée par les données historiques, qui manifeste un comportement aléatoire, afin de l'utiliser comme base de projection dans le futur. Cette loi fondamentale n'est pas toujours trouvée facilement car les situations réelles sont compliquées ; des tendances, des facteurs saisonniers, des facteurs cycliques et des fluctuations aléatoires engendrent cette loi.

La méthode de Box & Jenkins est la méthode la mieux adaptée pour les séries chronologiques complexes où la loi n'est pas immédiatement apparente.

Nous décrivons dans ce qui suit la méthodologie de Box & Jenkins et la modélisation stochastique ARIMA qui lui est souvent associée. Cette approche va nous permettre d'introduire bon nombre de concepts et techniques qui forment le fondement même de l'approche statistique.

1. Définitions et concepts

a. Processus stationnaire au sens strict : [UPM 04]

Soit un processus temporel aléatoire $(X_t, t \in \mathbb{Z})$:

Un processus est dit stationnaire au sens strict si, pour toute valeur j_1, j_2, \dots, j_n la distribution jointe de la suite $(X_t, X_{t+j_1}, \dots, X_{t+j_n})$ dépend uniquement des intervalles de temps j_1, j_2, \dots, j_n et est indépendante de la période t .

(comme complément, l'annexe 1 présente la définition d'un processus stationnaire au sens faible)

b. Processus Bruit Blanc (White noise) : [FIO 99]

Parmi les classes de processus stationnaires, il existe des processus particuliers que sont les processus bruit blanc (ou White noise). Ces processus sont très souvent utilisés en analyse des séries temporelles car ils constituent en quelque sorte les "briques élémentaires" de l'ensemble des processus temporels. En effet, tout processus stationnaire peut s'écrire comme une somme pondérée de bruits blancs (théorème de Wold).

Un bruit blanc est un processus stationnaire à accroissements indépendants. On parle aussi de processus i.i.d (variables indépendantes et identiquement distribuées). (cf. Annexe 2)

c. Fonction d'autocorrélation et fonction d'autocorrélation partielle : [BOU 98]

▪ Fonction d'autocorrélation (FAC) :

On appelle coefficient d'autocorrélation d'ordre 1 le coefficient de corrélation linéaire calculé entre la série chronologique et cette même série décalée d'une période de temps.

La fonction d'autocorrélation d'un processus X_t , de moyenne $E[X_t] = m$, noté $\rho(k)$ est définie par :

$$\forall k \in [0, K] : \quad \rho(k) = \frac{\gamma(k)}{\gamma(0)} \quad (\text{I.1})$$

(K étant le décalage maximum admissible)

Avec $\rho(k) \in [-1, 1]$, et où $\gamma(k)$ désigne la fonction d'autocovariance

$$\gamma(k) = E[(X_t - m)(X_{t-k} - m)] \quad (\text{I.2})$$

$\gamma(0)$ désigne la fonction de variance.

▪ **Fonction d'autocorrélation partielle (FACP) :**

L'autocorrélation partielle d'ordre k désigne la corrélation entre X_t et X_{t-k} obtenue lorsque l'influence des variables X_{t-k-i} ; avec $i < k$; a été retirée.

La fonction d'autocorrélation partielle d'un processus X_t , de moyenne $E[X_t] = m$, noté $r(k)$ est définie par le dernier coefficient de la projection linéaire de X_{t+1} sur ces k plus récentes valeurs c'est-à-dire :

$$X_{t+1} - m = C_1(X_t - m) + C_2(X_{t-1} - m) + \dots + C_{k-1}(X_{t-k+1} - m) + \rho(k)(X_{t-k+1} - m) \quad (\text{I.3})$$

Ou de façon générale, la fonction d'autocorrélation partielle d'un processus stationnaire X_t satisfait la relation :

$$r(k) = \frac{|r^*(k)|}{|r(k)|} \quad \forall k \in N \quad (\text{I.4})$$

avec :

$$r(k) = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(k-1) \\ \rho(1) & 1 & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \dots & \dots & 1 \end{pmatrix} \quad \text{et} \quad r^*(k) = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(1) \\ \rho(1) & 1 & \dots & \rho(2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \dots & \dots & \rho(k) \end{pmatrix}$$

d. Le processus autorégressif AR (p) : [BOU 98]**▪ Définition :**

Un processus autorégressif d'ordre p est un processus stationnaire X_t qui vérifie l'équation suivante :

$$X_t + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} = \varepsilon_t \quad (\text{I.5})$$

Les φ_i sont des réels et ε_t est un bruit blanc de variance σ^2 .

Cette relation peut s'écrire en fonction de l'opérateur retard L (cf. Annexe 3) sous la forme suivante :

$$(1 + \varphi_1 L + \varphi_2 L^2 + \dots + \varphi_p L^p) X_t = \varepsilon_t \quad (\text{I.6})$$

où :

$$\Phi(L)X_t = \varepsilon_t \quad (\text{I.7})$$

Un processus AR peut s'écrire sous la forme d'une moyenne mobile infinie.

En effet, si pour le polynôme $\Phi(L)$, toutes les racines de module différent de 1, on peut donc inverser l'opérateur $\Phi(L)$ défini sur les valeurs du processus stationnaire.

On obtiendra l'écriture suivante :

$$X_t = \Phi(L)^{-1} \varepsilon_t \quad (\text{I.8})$$

Si en plus le polynôme $\Phi(L)$ admet des racines de module strictement supérieur à 1 alors l'opérateur $\Phi(L)^{-1}$ admet un développement ne faisant intervenir que des puissances positives de L, on a alors :

$$\Phi(L)^{-1} = (1 + \varphi_1 L + \varphi_2 L^2 + \dots) \varepsilon_t \quad (\text{I.9})$$

▪ Les équations de Yule-Walker :

Etudions maintenant la fonction de covariance de X_t . En multipliant (I.5) par X_t et en prenant l'espérance, on obtient :

$$\gamma(0) = \sigma^2 - (\varphi_1 \gamma(1) + \dots + \varphi_p \gamma(p)) \quad (\text{I.10})$$

En multipliant par X_{t-h} ($h>0$), en prenant l'espérance et en divisant par $\gamma(0)$, on obtient :

$$\rho(h) + \varphi_1 \rho(h-1) + \dots + \varphi_p \rho(h-p) = 0 \quad \text{pour } h > 0 \quad (\text{I.11})$$

La suite d'autocorrélation satisfait donc une équation de récurrence linéaire homogène d'ordre p , dont les coefficients sont directement liés à ceux du polynôme autorégressif.

En écrivant l'équation (I.11) pour $h = 1, \dots, p$ et en tenant compte de la parité de $\rho(h)$, on obtient les équations de Yule-Walker :

$$\rho(h) = \varphi_1 \rho(h-1) + \varphi_2 \rho(h-2) + \dots + \varphi_p \rho(h-p) \quad (\text{I.12})$$

Ces équations sont bilinéaires en $\rho(h)$ et φ_i . Sous cette dernière forme, elles permettent d'obtenir, par inversion, les φ_i en fonction de $\rho(1), \rho(2), \dots, \rho(p)$.

On peut ainsi les mettre sous la forme d'un système linéaire en $\rho(1), \rho(2), \dots, \rho(p)$ et obtenir ainsi les quantités en fonction de $\varphi_1, \varphi_2, \dots, \varphi_p$. Les autres valeurs de $\rho(h)$ seront déterminées par l'équation de récurrence (I.11).

Cette équation de récurrence admet une seule suite de solutions pouvant s'interpréter comme une fonction d'autocorrélation. En effet les valeurs initiales $\rho(1), \rho(2), \dots, \rho(p)$ sont déterminées sans ambiguïté.

▪ Fonction d'autocorrélation partielle :

La fonction d'autocorrélation partielle $r(h)$ peut évidemment être calculée à partir des corrélations $\rho(h)$, à l'aide des méthodes générales suivantes :

$$X_t + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} = \varepsilon_t \quad (\text{I.13})$$

Il faut cependant souligner une propriété importante de cette fonction pour un processus AR(p), à savoir : $r(h) = 0$ pour $h > 0$.

Dans ce cas, comme $r(h) = \varphi_p$, on peut donc écrire :

$$X_t = \varepsilon_t - (\varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}) \quad (\text{I.14})$$

e. Le processus moyenne mobile MA (q) : [FIO 99]

▪ **Définition :**

On appelle processus moyenne mobile d'ordre q un processus X_t défini comme suit :

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (\text{I.15})$$

Où les θ_i sont des réels et ε_t est un bruit blanc de variance σ^2 .

La relation (I-15) peut encore s'écrire comme suit :

$$X_t = (1 + \theta_1 L + \dots + \theta_q L^q) \varepsilon_t \quad (\text{I.16})$$

où :

$$X_t = \Theta(L) \varepsilon_t \quad (\text{I.17})$$

Un tel processus est désigné dans littérature anglo-saxonne par MA(q) (MA sont les initiales de « Moving Average »).

Contrairement au processus AR (p), la définition de MA (q) est explicite et elle est clairement définie par l'équation (I-15) et est automatiquement stationnaire.

Avec la même démarche que celle pour un processus AR (p), on peut dire qu'un MA (q) peut être transformé en AR infini si les racines du polynôme $\Theta(L)$ ne sont pas de module égal à 1. Si en plus les racines ont un module strictement supérieur à 1, l'expression du AR (∞) ne contiendra que des termes positifs.

▪ **Fonction d'autocovariance :**

La fonction d'autocovariance est la suivante :

$$\gamma(h) = \begin{cases} E(X_t X_{t-h}) = (\theta_h + \theta_{h+1} \theta_1 + \theta_{h+2} \theta_2 + \dots + \theta_q \theta_{q-h}) \sigma^2 ; & \text{si } h \leq q \\ 0 & \text{si } h > q \end{cases} \quad (\text{I.18})$$

Pour $h = 0$, on obtient :

$$\gamma(0) = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2$$

La fonction d'autocorrélation deviendra donc :

$$\rho(h) = \begin{cases} (\theta_h + \theta_{h+1} \theta_1 + \theta_{h+2} \theta_2 + \dots + \theta_q \theta_{q-h}) / (1 + \theta_1^2 + \dots + \theta_q^2) ; & \text{si } h \leq q \\ 0 & ; \text{ si } h > q \end{cases} \quad (\text{I.19})$$

On constate que la fonction $\rho(h)$ s'annule pour $h > q$; il y a donc une certaine dualité entre la fonction d'autocorrélation d'un processus MA(q) et la fonction d'autocorrélation partielle $r(h)$ d'un processus AR(p).

f. Le processus ARMA (p, q) : [BOU 02]

▪ Généralités et définitions :

La modélisation ARMA généralise simultanément les modèles autorégressifs purs et moyennes mobiles pures étudiés précédemment.

Elle présente l'avantage d'être plus souple d'utilisation et de fournir généralement de bonnes approximations des séries réelles avec moins de paramètres que les modèles AR et MA.

Un processus stationnaire X_t admet une représentation ARMA (p, q) minimale s'il satisfait :

$$\phi(L)X_t = \Theta(L)\varepsilon_t \quad (\text{I.20})$$

$$\text{avec : } \begin{cases} \Phi(L) = 1 + \varphi_1 L + \dots + \varphi_p L^p \\ \Theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q \end{cases} \quad \text{et : } \varphi_p \neq 0 ; \theta_q \neq 0$$

Les polynômes Θ et Φ ont leurs racines de modules strictement supérieur à 1 et ils n'ont pas de racines communes et ε_t est un bruit blanc de variance σ^2 non nulle.

Sous ces mêmes conditions on peut dire que :

- X_t admet la représentation MA (∞) i.e. $X_t = (\Theta(L) / \Phi(L)) \varepsilon_t$
- X_t admet la représentation AR (∞) i.e. $(\Phi(L) / \Theta(L)) X_t = \varepsilon_t$

On remarque que les processus purs sont des cas particuliers de la représentation ARMA. Ainsi, le processus moyenne mobile MA (q) est un processus ARMA (0,q) et un processus autorégressif AR(p) pur est un processus ARMA(p, 0).

Les seuls processus admettant simultanément une représentation MA pure et une représentation AR pure correspondent aux cas ARMA (0, 0) i.e. aux bruits blancs.

▪ **Les autocovariances :**

La fonction d'autocovariance d'un processus X_t satisfaisant une représentation ARMA (p, q) minimale est telle que :

$$\gamma(h) + (\varphi_1 \gamma(h-1) + \dots + \varphi_p \gamma(h-p)) = 0 \quad \text{pour } h > q+1 \quad (\text{I.21})$$

▪ **Caractérisation des ordres p et q : [UPM 04]**

Les ordres minimaux p et q sont apparus constamment dans les paragraphes précédents comme des indices à partir desquels certaines suites s'annulaient, comme des ordres d'équations de récurrence linéaire. On peut donc tirer des propriétés fournissant des caractérisations de ces ordres. On peut citer deux principales propriétés à savoir :

Propriété 1 :

Si un processus X_t admet une représentation ARMA (p, q) minimale alors :

- p est l'ordre minimum des équations de récurrence satisfaites par les coefficients du développement MA (∞), à partir d'un certain indice.
- q est l'ordre minimum des équations de récurrence satisfaites par les coefficients du développement AR (∞), à partir d'un certain indice.

Propriété 2 :

Un processus stationnaire régulier admet une représentation ARMA(p, q) minimale si et seulement si :

- $\Delta(i, j) = 0$ pour $i > p-1$ et $j > q-1$
- $\Delta(p, q-1) \neq 0$ et $\Delta(p-1, q) \neq 0$

Le résultat sera donc:

- $\Delta(i, q-1) \neq 0$ pour $i > p-2$
- $\Delta(p-1, j) \neq 0$ pour $j > q-2$

Cette dernière propriété conduit à construire un tableau à double entrée comportant en ligne l'indice i et en colonne l'indice j et ayant comme éléments les diverses valeurs des déterminants $\Delta(i, j)$ (tableau.I.1)

Le processus admet une représentation ARMA si et seulement si ce tableau contient un « coin » inférieur droit composé uniquement de 0. Ceci explique le nom de la méthode du coin attribué à cette façon de caractériser les ordres.

Tableau I.1 : Valeurs des déterminants

I \ J	0	1	2	q
0							
1							
2							
.							
P							

$q = 1$

$q = 2$

MA(2)

$p = 1$ ARMA(1, 1)

$p = 2$ AR(2)

g. Le processus ARIMA (p,d,q) : [CHA 03]

Il est évident que pour la plupart des séries économiques, l'hypothèse de stationnarité n'est pas toujours vérifiée. En revanche, si l'on considère par exemple des différences premières, ou plus généralement les différences d'ordre (d) de telles séries, l'hypothèse de stationnarité devient souvent vraisemblable.

Il est donc naturel de considérer la classe des processus dont la différence d'un certain ordre satisfait une représentation ARMA.

Notons $\Delta^d X_t$ la différence d'ordre d de X_t i.e. $(1-L)^d X_t$.(cf. Annexe 3)

Un processus X_t est un processus ARIMA (p, d, q) « autorégressif moyenne mobile intégré » s'il satisfait une équation du type :

$$\Phi(L)(1-L)^d X_t = \Theta(L)\varepsilon_t \quad (\text{I.22})$$

$$\text{avec : } \begin{cases} \Phi(L) = 1 + \varphi_1 L + \dots + \varphi_p L^p \\ \Theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q \end{cases} \quad \text{et : } \varphi_p \neq 0 ; \theta_q \neq 0$$

Ces polynômes ont des racines de module strictement supérieur à 1 et ε_t est un bruit blanc de variance σ^2 non nulle.

La propriété principale d'un processus ARIMA est qu'il tend vers un processus ARMA stationnaire.

h. Les modèles saisonniers : [DEL 00]

On sait que certaines séries mensuelles ont un « profil saisonnier » marqué, c'est à dire que des données relatives à un même mois de différentes années ont tendance à se situer de façon analogue par rapport à la moyenne annuelle. Ceci conduit à penser qu'il serait intéressant dans un modèle ARIMA d'introduire un décalage multiple de 12.

En théorie rien ne s'oppose à ce qu'on prenne des valeurs de p et q suffisamment grandes pour que ces retards soient pris en compte, seulement cela entraînerait une augmentation du nombre de paramètres et il sera très difficile de les estimer.

Pour résoudre ce problème, Box et Jenkins ont proposé un type particulier de modèles ARIMA saisonniers. Ces modèles vérifient la relation suivante :

X_t est appelé processus SARIMA $(p,d,q) (P,D,Q)_S$ Avec une période S , si le processus $\Delta^d \Delta_S^D X_t$ tel que :

$$\Delta^d \Delta_S^D X_t = (1-L)^d (1-L^S)^D X_t \quad (I.23)$$

est un processus ARMA stationnaire de la forme :

$$\Delta^d \Phi_p(L) \Delta_S^D \Phi_P(L^S) X_t = \Theta_q(L) \Theta_Q(L^S) \varepsilon_t \quad (I.24)$$

Avec :

- S est la période de la saisonnalité ($S = 12$ pour une saisonnalité annuelle, $S = 4$ pour une saisonnalité trimestrielle).
- $\Delta = 1-L$ et $\Delta_S = 1-L^S$ et $\Phi_p, \Phi_P, \Theta_q, \Theta_Q$ sont des polynômes de degrés p, P, q, Q , dont les racines sont de module supérieur à 1 et ε_t est un bruit blanc.

Il faut savoir qu'un processus satisfaisant une telle relation est appelée un processus **SARIMA** $s[(p, d, q) (P, D, Q)]$.

Le raisonnement conduisant à un tel modèle consiste à appliquer aux S séries la même transformation ; ces séries sont obtenues à partir de X_t en prenant pour chaque série des mois identiques.

La transformation est la suivante :

$$\frac{\Delta_S^D \Phi_P(L^S)}{\Theta_Q(L^S)}$$

On obtient ainsi la série suivante :

$$A_t = \left(\frac{\Delta_S^D \Phi_P(L^S)}{\Theta_Q(L^S)} \right) X_t \quad (I.25)$$

Si l'on suppose que cette série n'est pas stationnaire, elle est donc modélisable par un ARIMA (p,d,q) avec des valeurs faibles de p , d , q .

Le processus A_t vérifiera donc la relation suivante :

$$\Phi(L)(1-L)^d A_t = \Theta(L)\varepsilon_t$$

En fait, l'identification des paramètres P, D, Q des facteurs saisonniers se fait par des méthodes analogues à celles utilisées pour les paramètres p , d et q comme on le verra dans les parties suivantes. On ne retiendra que les valeurs de h multiples de S dans $\rho(h)$ et $r(h)$.

2. La méthode de Box & Jenkins :

▪ Description de la méthode : [CHA 03]

Box et Jenkins ont proposé en 1970, une démarche générale de prévision pour une série univariée. Cette notion est basée sur la notion de processus ARIMA.

Supposons que les observations disponibles X_1, \dots, X_t puissent être considérées comme compatibles avec un modèle ARIMA et que toutes les hypothèses de ce modèle sont satisfaites. Alors, à un horizon h , la prévision théorique d'une valeur future X_{t+h} est donnée par une régression linéaire de X_{t+h} sur X_t, X_{t-1}, \dots

Les paramètres de cette régression dépendent des paramètres du modèle. Une telle formule ne peut être directement utilisée puisque ses paramètres sont inconnus. Il faut donc au préalable les estimer.

Cette phase d'estimation est très délicate car les paramètres du modèle sont soumis à des contraintes différentes. En effet, on distingue les degrés inconnus p , q et d qui sont des valeurs entières, les autres paramètres, φ_i et θ_j -ceux des polynômes de retard où les modules des racines des polynômes Θ et Φ sont à l'extérieur du disque unité- et la variance des bruits blanc σ^2 .

Comme l'estimation de ces paramètres est très difficile, Box et Jenkins ont adopté une méthode qui consiste à séparer la recherche des degrés (phase d'identification) et celle des autres paramètres (phase d'estimation).

Cette démarche se présente de la façon suivante :

- a) Nous cherchons des valeurs plausibles des degrés (p, d, q) en utilisant des méthodes ne nécessitant pas l'estimation des autres paramètres. C'est l'*identification à priori* avant estimation. Cette phase conduit à retenir plusieurs triplets (p_j, d_j, q_j) avec $j = 1, \dots, J$
- b) Pour chaque modèle plausible, on cherche les valeurs correspondant aux autres paramètres φ_i et θ_j . C'est la phase *d'estimation* qui ressort avec J modèles estimés M_j .
- c) Dans cette étape, on examine si les modèles sont compatibles ou non avec les hypothèses sous-jacentes du modèle. On établit donc un certain nombre de barrages que devraient franchir les modèles compatibles avec les données. Après cette phase de *vérification*, on dispose d'un plus petit nombre de modèles estimés.
- d) A ce niveau les modèles retenus possèdent des propriétés semblables en vue de la prévision. Dans l'étape de choix du modèle on ne va retenir qu'un seul modèle des M_j . C'est l'étape de *vérification à posteriori*.
- e) On pourra alors effectuer les prévisions en utilisant la formule théorique de prévision après avoir remplacé les paramètres inconnus par les estimations obtenues pour M_j .

La démarche peut être résumée grâce à l'organigramme ci-après :

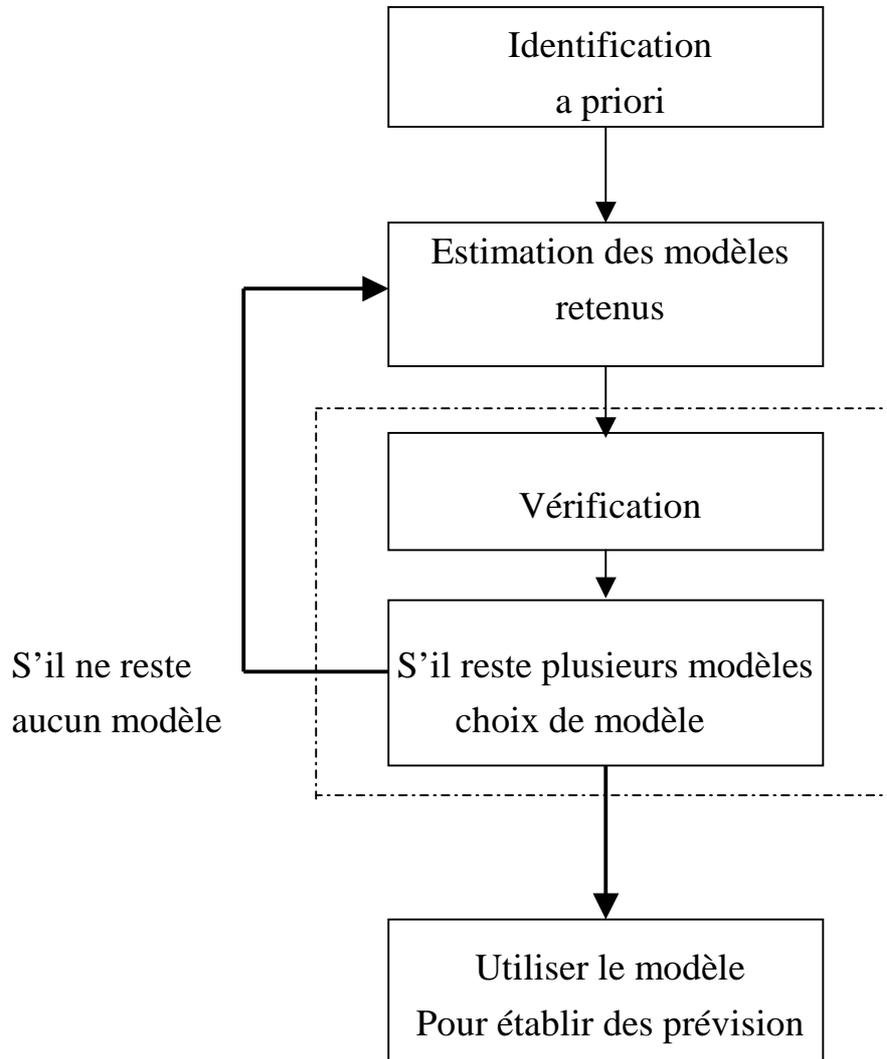


Figure.I.2- Organigramme de la méthode de Box & Jenkins

Développons maintenant chaque étape de la démarche de Box & Jenkins en détails :

1) Identification à priori : [FIO 99]

– Choix de d :

Si le processus X_t ayant généré les observations est un ARIMA avec $d > 0$, il est évident que ce processus n'étant pas stationnaire, on ne peut pas définir la fonction d'autocorrélation $\rho(h)$. Seulement cette fonction a toujours un sens et on peut voir que cette quantité tend vers 1 lorsque t tend vers ∞ .

On peut donc conclure qu'en pratique, si la fonction d'autocorrélation estimée $\rho_T(h)$ reste relativement proche de 1 pour un assez grand nombre de valeurs de h , on peut donc penser qu'il faille différencier la série pour la rendre stationnaire. Compte tenu des fluctuations d'échantillonnage de $\rho_T(h)$, le critère de proximité à 1 des premières valeurs de $\rho_T(h)$ doit être remplacé par la proximité de ces valeurs entre elles.

Quand la décision de différencier une fois la série a été prise, on applique le critère ci-dessus à la série différenciée pour savoir s'il convient de la différencier une seconde fois et ainsi de suite.

Pour les séries économiques, les valeurs de d qu'on est amené à retenir sont en fait 0, 1, 2. Les valeurs supérieures sont très rares.

– Le choix de q (resp. p) dans le cas d'un processus MA (resp. AR) :

Les guides pour les choix de p et q sont essentiellement les formes des estimateurs de fonctions d'autocorrélation $\rho_T(h)$ et d'autocorrélation partielles $r_T(h)$ de la série différenciée d fois.

On peut penser qu'une chute de $\rho_T(h)$ vers des valeurs proches de zéro pour $h > q$ indique un MA(q), ou qu'une chute de $r_T(h)$ vers des valeurs proches de zéro pour $h > p$ indique un AR(p).

Seulement pour juger si les $\rho_T(h)$ ou les $r_T(h)$ sont significativement différents de zéro, il est bon d'avoir une idée de leurs écart types.

Il est démontré que pour un MA(q), la variance de $\rho_T(h)$ est donnée par :

$$\text{Var} [\rho_T(h)] \approx [1 + 2(\rho(1)^2 + \dots + \rho(q)^2)] \left(\frac{1}{T}\right) \quad \text{pour } h > q \quad (\text{I.26})$$

On peut donc considérer l'estimateur de l'écart type comme étant :

$$\sigma[\rho_T(h)] = \frac{1}{\sqrt{T}} \sqrt{[1 + 2(\rho(1)^2 + \dots + \rho(q)^2)]} \quad (\text{I.27})$$

De même, on peut considérer $(1/\sqrt{T})$ comme écart type approximatif de $r_T(h)$ pour $h > p$.

Nous pouvons conclure que pour identifier l'ordre p d'un processus autorégressif, on représente la suite des autocorrélations partielles estimées $r_T(h)$ puis on examine à partir de quelle valeur de p celle-ci reste dans la bande $[-1.96/\sqrt{T}, +1.96/\sqrt{T}]$. Il faut savoir que cette bande est construite à partir d'un test de nullité des autocorrélations partielles $r(h)$ pour un seuil de 5%.

De la même manière, pour identifier l'ordre q d'un processus moyenne mobile, nous devons représenter la suite des autocorrélations estimées $\rho(h)$ et détecter à partir de quelle valeur de h tous les $\rho_T(h)$ restent dans l'intervalle d'extrémités $\pm \frac{1,96}{\sqrt{T}} [1 + 2\{\rho(1)^2 + \dots + \rho(h-1)^2\}]^{(1/2)}$.

Il existe une méthode de choix des paramètres p et q pour un processus ARMA appelée « la méthode du coin ». Nous développons cette méthode dans l'annexe 4.

2) Estimation : [FIO 99]

La phase d'estimation utilise les techniques classiques de statistique. Les paramètres $\varphi_i, \theta_j, \sigma^2$ sont généralement estimés en utilisant l'approche du maximum de vraisemblance ou la technique de moindres carrés. Ces méthodes sont appliquées en supposant les degrés p_j, d_j, q_j fixes, égaux aux valeurs trouvées dans l'étape d'identification à priori.

L'estimation des paramètres d'un modèle ARIMA (p,d,q) lorsque les ordres p et q sont supposés connus peut se réaliser par différentes méthodes dans le domaine temporel :

- Moindres Carrés Ordinaires 'MCO' (modèle sans composante MA, $q = 0$). Dans ce cas, on retrouve les équations de Yule Walker (I.12). En remplaçant les

autocorrélations théoriques par leurs estimateurs, on peut retrouver les estimateurs des MCO des paramètres du modèle par la résolution des équations de Yule Walker.

- Maximum de Vraisemblance approché (Box & Jenkins 1970)

- Maximum de Vraisemblance exacte (Newbold 1974, Harvey et Philips 1979, Harvey 1981).

Nous présentons en *annexe 5*, la démarche d'estimation des paramètres du modèle par le maximum de vraisemblance, réalisée à l'aide de l'algorithme d'optimisation non linéaire (Newton-Rahpson, méthode du simplex).

3) Vérification : [BOU 98]

Il s'agit, lors de cette étape, de vérifier statistiquement du choix des paramètres introduits dans la phase d'identification.

Les tests que l'on fait subir au modèle sont de deux types : les tests concernant les paramètres φ_i et θ_j du modèle et ceux concernant les hypothèses faites sur ε .

– Les tests concernant les paramètres :

Il s'agit de comparer deux formulations : ARMA (p,q) et une autre formulation ARMA(p₁, q₁).

On suppose que l'un des modèles est un cas particulier de l'autre et l'on démarre d'un modèle ARMA(p,q).

Dans un premier cas, il s'agit de voir si l'on peut diminuer d'une unité le nombre de retards intervenant dans la partie autorégressive (i.e. p₁ = p-1 et q₁ = q).

Cela revient à tester la significativité du paramètre φ_p , ce qui peut être fait par le biais d'un test de *Student*.

En effet, si l'on considère l'estimateur de φ_p dans la formulation ARMA (p,q) $\hat{\varphi}_p$ et $Var(\hat{\varphi}_p)$ comme étant sa variance estimée, alors on acceptera avec un risque de 5% l'hypothèse que le modèle est ARMA(p-1, q), si :

$$\frac{|\hat{\varphi}_p|}{\sqrt{Var(\hat{\varphi}_p)}} < Z_{\frac{\alpha}{2}} \quad (\text{I.28})$$

On peut effectuer un test similaire lorsque p₁ = p et q₁ = q-1.

Dans un deuxième cas, il s'agit de voir si l'on peut introduire un retard supplémentaire dans la partie autorégressive.

Il faut alors estimer un modèle ARMA $(p_1, q_1) = \text{ARMA}(p+1, q)$ et tester dans ce modèle la nullité de φ_{p+1} , ceci nous ramène au cas précédent. On peut faire la même chose concernant le modèle ARMA $(p, q + 1)$.

Dans un troisième cas, on étudie la possibilité d'augmenter simultanément les degrés des polynômes autorégressifs et moyenne mobile. Cependant on ne sait pas tester cette possibilité car les tests du rapport de maximum de vraisemblance ne sont pas utilisables.

– Les tests concernant le bruit blanc :

Ces tests ont pour but de vérifier que la série des résidus obtenue est cohérente avec l'hypothèse selon laquelle ces mêmes résidus forment un bruit blanc.

$$\varepsilon_t = \left(\frac{\Phi(L)(1-L)^d}{\Theta(L)} \right) X_t \quad (\text{I.29})$$

On peut citer le test du « porte manteau » qui a été proposé par Box-Pierce. Il est fondé sur la statistique :

$$Q = T \sum_{h=1}^K \hat{\rho}_{\varepsilon_t}^2(h) \quad (\text{I.30})$$

Q suit asymptotiquement une loi de χ^2 à $K-p-q$ degrés de liberté. On refuse l'hypothèse d'indépendance au niveau α si:

$$Q > \chi_{1-\alpha}^2[k-p-q] \quad (\text{I.31})$$

4) Choix du modèle : [CHA 03]

Une fois l'étape de vérification terminée, plusieurs modèles peuvent être retenus. A ce niveau, il faut repérer les modèles jumeaux qui sont des représentations différentes d'un même modèle et qui sont donc équivalents pour la prévision.

Le choix entre les différents modèles se fait sur la base de critères qui sont :

– **Les critères du pouvoir prédictif :**

Ces critères se basent sur l'erreur de prévision, ils proposent donc des modèles qui conduisent à une erreur de prévision suffisamment petite.

Plusieurs critères sont proposés, à savoir :

a) L'estimateur de la variance.

b) Le coefficient de détermination : $R^2 = 1 - \frac{\hat{\sigma}_T^2}{V}$ tel que V est la variance empirique de la série initiale différenciée d fois. Ce critère est la version normée du précédent.

c) Le coefficient de détermination modifié $\bar{R}^2 = 1 - \frac{\hat{\sigma}_T^2 / (T - p - q)}{V / (T - 1)}$, il permet de prendre en compte les degrés des polynômes autorégressifs et moyenne mobile.

d) La statistique de Fisher $F = \frac{(Var - \hat{\sigma}^2) / (p + q)}{\hat{\sigma}^2 / (T - p - q)}$. C'est une analogie avec le cas d'un modèle linéaire.

Les modèles choisis doivent remplir l'une des conditions suivantes, à savoir minimiser le critère a) ou maximiser les critères b), c), d).

– **Le critère d'information :**

C'est une approche qui consiste à supposer que les modèles ARMA (p,q) fournissent des approximations de la réalité et que la vraie loi inconnue des observations ne satisfait pas forcément un tel modèle.

On peut fonder le choix du modèle sur une mesure de l'écart entre la vraie loi inconnue et le modèle proposé.

La mesure habituelle retenue est la quantité d'information de Kullback¹ $I(f_0, F_{p,q})$.

$$I(f_0, F_{p,q}) = \underset{f \in F_{p,q}}{\text{Min}} \int \log \frac{f_0(x)}{f(x)} f_0(x) dx$$

La valeur de la quantité d'information $I(f_0, F_{p,q})$ est évidemment inconnue, mais si l'on possède un bon estimateur de $I(f_0, F_{p,q})$, il sera possible de l'utiliser comme critère. On retiendra alors $F_{p,q}$ qui conduira à la plus petite valeur de cet estimateur.

Les estimateurs de la quantité d'information les plus utilisés sont les suivants :

– Critère de Akaike : $AIC(p, q) = T \log(\hat{\sigma}_{\hat{\varepsilon}_t}^2) + 2(p + q)$

– Le critère d'information bayésien :

$$BIC(p, q) = T \log(\hat{\sigma}_{\hat{\varepsilon}_t}^2) - (n - p - q) \log \left[1 - \frac{(p + q)}{T} \right] + (p + q) \log(T) + \log \left[(p + q)^{-1} \left(\frac{\hat{\sigma}_{\hat{\varepsilon}_t}^2}{\hat{\sigma}_{\hat{\varepsilon}_t}^2 - 1} \right) \right]$$

5) Prévision : [BOU 98]

Lorsque la construction du modèle linéaire est terminée, nous sommes à même de générer des prévisions. Pour identifier le processus étudié à un processus ARMA; différentes transformations ont été appliquées; il est nécessaire lors de la phase de prévision de prendre en compte la transformation retenue et de « recolorer la prévision ». Plusieurs cas sont possibles :

- Si le processus contient une tendance déterministe, on extrait cette dernière par régression afin d'obtenir une série stationnaire lors de la phase d'estimation. Ensuite, lors de la phase de prévision, on adjoint aux prévisions réalisées sur la composante ARMA stationnaire, la projection de la tendance.
- Si la transformation résulte de l'application d'un filtre linéaire (de type par exemple différences premières), on réalise les prévisions sur la série filtrée stationnaire et l'on reconstruit ensuite par inversion du filtre les prévisions sur la série initiale.
- Dans le cas des modèles saisonniers, une différence saisonnière est appliquée. Lors de la prévision, cette différenciation est prise en compte.

▪ Extensions des modèles ARIMA : [FIO 99]

Dès leur introduction, les modèles ARIMA ont connu un large succès et la méthodologie de Box & Jenkins a été appliquée dans de nombreux domaines induisant l'obtention de prévisions ou la construction de modèles linéaires stochastiques. C'est le cas par exemple en économétrie ou encore en théorie du contrôle de processus, où des extensions ARMAX ont été proposées par les ingénieurs.

Les modèles ARMAX, ou ARIMA multi variés, occupent une place prépondérante en théorie du contrôle de processus, notamment en contrôle prédictif. Dans ce contexte, les variables exogènes X_{ti} sont appelées variables d'entrée ou de commande. La sortie du système Y_t est influencée par ses valeurs retardées Y_{t-j} (partie AR), les variables d'entrée ainsi que des perturbations aléatoires ε_{t-i} (partie MA) appelées entrées secondaires ou parasites.

Les conditions de stationnarité de ces modèles sont analogues à celles d'un processus ARMA uni varié. De même, l'estimation des paramètres du processus se fait de façon similaire à la démarche d'estimation d'un processus ARMA ou d'une série stationnarisée par différence et approchée à un processus ARMA. De plus, la plupart des logiciels ayant trait à l'analyse des séries chronologiques, prennent en charge les modèles ARIMA et leurs extensions.

Pour conclure, nous récapitulons le mécanisme général de cette méthode :

Box et Jenkins ont conçu cette méthode où il n'y a pas besoin de faire dès le départ l'hypothèse d'une loi rigide. Au lieu de cela, on commence par une loi expérimentale adaptée aux données en vue de minimiser l'erreur. Ensuite, la méthode fournit au manager des informations explicites qui lui donnent la possibilité de juger si la loi adaptée est correcte pour la situation considérée. Si elle l'est, on peut établir directement la prévision, et si elle ne l'est pas, la méthode de Box-Jenkins donne des indications supplémentaires permettant d'identifier la loi correcte. Ce processus d'itération permet d'apporter de proche en proche des améliorations à la méthode, pour arriver à un modèle de prévision qui assure l'optimisation de la loi de base et qui minimise l'erreur par rapport à cette loi. Ainsi, les avantages qu'offrent cette méthode pour le traitement des séries complexes font que cette dernière sera l'un des outils de prévisions sélectionnés lors de notre application en partie pratique de ce travail.

Chapitre II : Les réseaux de neurones artificiels

La démarche qui mène aux réseaux de neurones est semblable à celle qui mène à l'intelligence artificielle. Avec l'intelligence artificielle, les chercheurs ont voulu modéliser le raisonnement et la connaissance pour recréer de manière explicite les différentes étapes de la réflexion d'un expert du domaine considéré et obtenir à partir d'une question la réponse appropriée.

Les réseaux de neurones, fabriqués de structures cellulaires artificielles, constituent une approche permettant d'aborder sous des angles nouveaux les problèmes de perception, de mémoire, d'apprentissage et de raisonnement. Ils s'avèrent aussi des alternatives très prometteuses pour contourner certaines des limitations des ordinateurs classiques. Grâce à leur traitement parallèle de l'information et à leurs mécanismes inspirés des cellules nerveuses (neurones), ils infèrent des propriétés émergentes permettant de solutionner des problèmes jadis qualifiés de complexes.

L'hypothèse principale, à la base de l'essor des réseaux de neurones artificiels, est que le comportement intelligent est sous-tendu par un ensemble de mécanismes mentaux. Cela donc consiste à construire un modèle des mécanismes du cerveau. Ces mécanismes sont ceux qui font évoluer le cerveau et calculent la réponse de l'organisme à des stimuli extérieurs. Les réseaux de neurones artificiels sont issus de ce courant de pensée et des études menées depuis le début du siècle sur l'anatomie du cerveau.

Lors de l'émergence d'une nouvelle technique, l'ingénieur se demande naturellement en quoi elle peut lui être utile. La réponse à cette question doit être particulièrement précise et motivée. C'est la raison pour laquelle ce chapitre est consacré à un exposé des principes théoriques et pratiques qui justifient l'utilisation des réseaux de neurones.

Nous commencerons par introduire des définitions essentielles : qu'est ce qu'un neurone formel ? Qu'est ce qu'un réseau de neurones ? Nous établirons par la suite un cadre général de modélisation pour les réseaux de neurones.

Nous introduirons ensuite les propriétés des réseaux de neurones : l'approximation universelle et l'apprentissage.

Les réseaux de neurones artificiels, une nouvelles approche de prévision :

Au début des années quatre-vingt, plusieurs recherches se sont orientées vers le développement de nouvelles techniques basées sur l'approximation de fonctions.

En effet, les réseaux de neurones artificiels habituellement utilisés en physique appliquée font leur entrée, depuis une dizaine d'années, dans le domaine économique en tant que méthode quantitative de prévision, à côté des méthodes statistiques classiques. Il existe plusieurs raisons qui nous ont conduit à nous intéresser à cet outil :

- Ils sont tout à fait adaptés pour traiter des problèmes complexes non structurés, c'est-à-dire des problèmes sur lesquels il est impossible à priori de spécifier la forme des relations entre variables utilisées. C'est grâce à des algorithmes, que ces systèmes apprennent par eux-mêmes les relations entre les variables à partir d'un ensemble de données, à l'image de ce que ferait le cerveau humain. Ainsi le réseau se paramètre de lui même à partir d'exemples qui lui sont fournis.[DRE 04]
- Il ne répète pas bêtement le passé et s'il existe effectivement une relation de cause à effet entre les descriptions en entrée et les valeurs à prédire, le réseau l'identifiera.[DRE 04]
- Le réseau de neurones est robuste. Il n'est pas handicapé par quelques exemples bruités ou faux : ceux-ci sont écartés car ils sont incohérents avec le reste. Les valeurs manquantes sont également gérées habilement et ne nuisent pas à la construction du modèle.[DRE 04]

Pour mieux expliquer les différentes raisons qui nous ont poussés à nous tourner vers cette nouvelle technique, nous allons introduire dans ce qui suit les fondements théoriques des réseaux de neurones.

I. Modèle mathématique :

Les réseaux de neurones biologiques (*cf. Annexe 6*) réalisent facilement un certain nombre d'applications telles que la reconnaissance de formes, le traitement du signal, l'apprentissage par l'exemple, la mémorisation, la généralisation. Ces applications sont pourtant, malgré tous les efforts déployés en algorithmique et en intelligence artificielle, à la limite des possibilités actuelles. C'est à partir de l'hypothèse que le comportement intelligent

émerge de la structure et du comportement des éléments de base du cerveau, que les réseaux de neurones artificiels se sont développés. Les réseaux de neurones artificiels sont des modèles, à ce titre, ils peuvent être décrits par leurs composants, leurs variables descriptives et les interactions des composants.[TOU 92]

1. Composant : le neurone formel (artificiel) :

Chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance de neurones amonts. A chacune de ces entrées est associé un poids w , abréviation de « *weight* » (poids en anglais), représentatif de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones avals. A chaque connexion est associée un poids.[TOU 92]

Un neurone formel représente donc « *une fonction non linéaire, paramétrée, à valeurs bornées* ».[DRE 04]

Et selon E.DAVALO, on peut définir un neurone formel par les éléments suivants :

- La nature de ses entrées ;
- La fonction d'entrée totale qui définit le prétraitement subi par les entrées ;
- La fonction d'activation du neurone et ;
- La nature de la sortie du neurone.

Finalement, le fonctionnement d'un neurone peut être décrit comme la réaction à une somme pondérée d'excitations.[DAV 93]

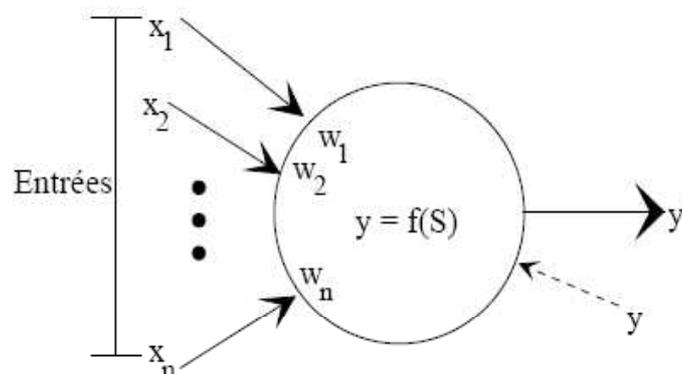


Figure.II.1. Neurone formel [TOU 92]

2. Comportement : [DRE 04]

On distingue deux phases. La première est habituellement le calcul de la somme pondérée des entrées (S) ou potentiel selon l'expression suivante : (figure.II.1)

$$S = \sum_i (w_i \cdot x_i) \quad (\text{II.1})$$

Où x_i : les variables d'entrée et w_i : les paramètres ou poids synaptiques ;

A partir de cette valeur, une fonction de transfert f (fonction d'activation) calcule la valeur de l'état du neurone (la sortie y). C'est cette valeur qui sera transmise aux neurones aval.

La fonction f peut être paramétrée d'une autre manière : les paramètres (w_i) interviennent directement dans la fonction.

Il existe de nombreuses formes possibles pour la fonction de transfert. Les trois fonctions les plus utilisées sont la fonction seuil, la fonction linéaire à seuil et la fonction sigmoïde (cf. *Annexe 7*). En théorie toute fonction croissante et impaire peut jouer le rôle de fonction de transfert. Le choix de cette fonction dépend de l'application.

Dans tous les cas, la fonction d'activation f doit être saturable, pour éviter des valeurs en sortie trop élevées qui peuvent déstabiliser le réseau .

La remarque importante qu'on peut faire sur la fonction d'activation est son caractère non linéaire, ce caractère avantageux donne la possibilité au réseau de simuler des fonctions complexes, non linéaires.

3. Les variables descriptives :

Ces variables décrivent l'état du système. Dans le cas des réseaux de neurones qui sont des systèmes non autonomes, un sous-ensemble des variables descriptives est constitué par les variables d'entrée, variables dont la valeur est déterminée extérieurement au modèle.[TOU 92]

4. Structure de connexions :

Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule

une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique de réseaux est évidemment un réseau.[TOU 92]

Comme nous venons de le voir, un neurone réalise simplement une fonction non linéaire, paramétrée, de ses variables d'entrée. L'intérêt des neurones réside dans les propriétés qui résultent de leur association en réseau, c'est à dire de la composition des fonctions non linéaires réalisées par chacun des neurones.[DRE 04]

Les connexions entre les neurones qui composent le réseau décrivent la topologie du modèle. Elle peut être quelconque, mais le plus souvent il est possible de distinguer une certaine régularité.

a- Réseau à couches (ou multicouche) : .[DRE 04]

Les neurones sont arrangés par couche. Il n'y a pas de connexion entre neurones d'une même couche et les connexions ne se font qu'avec les neurones des couches avales (figure.II.2). Habituellement, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante et celle-ci seulement. Ceci nous permet d'introduire la notion de sens de parcours de l'information (de l'activation) au sein d'un réseau et donc définir les concepts de neurone d'entrée, neurone de sortie. Par extension, on appelle couche d'entrée l'ensemble des neurones d'entrée, couche de sortie l'ensemble des neurones de sortie. Les couches intermédiaires n'ayant aucun contact avec l'extérieur sont appelés couches cachées.

L'expression « neurones d'entrée » est trompeuse. Nous l'utiliserons par abus de langage, mais les entrées ne sont pas des neurones : elles ne réalisent aucun traitement de l'information, puisqu'elles ne font que transmettre les valeurs des variables.

Les réseaux à une seule couche cachée sont appelés réseaux mono-couche, alors que les réseaux en possédant plusieurs sont appelés réseaux multicouches.

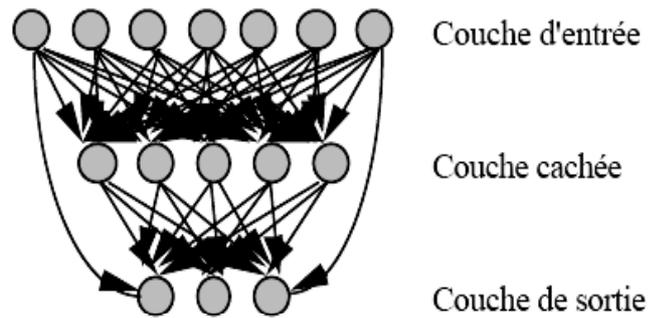


Figure.II.2. Réseau à couches [TOU 92]

b- Réseau à connexions locales : .[DRE 04]

Il s'agit d'une structure multicouche, mais qui, à l'image de la rétine, conserve une certaine topologie. Chaque neurone entretient des relations avec un nombre réduit et localisé de neurones de la couche avale (figure.II.3). Les connexions sont donc moins nombreuses que dans le cas d'un réseau multicouche classique.

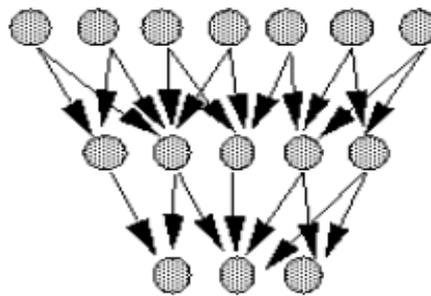


Figure.II.3. Réseau à connexions locales [TOU 92]

c- Réseau à connexions récurrentes : .[DRE 04]

Les connexions récurrentes ramènent l'information en arrière par rapport au sens de propagation défini dans un réseau multicouche. Ces connexions sont le plus souvent locales (figure.II.4).

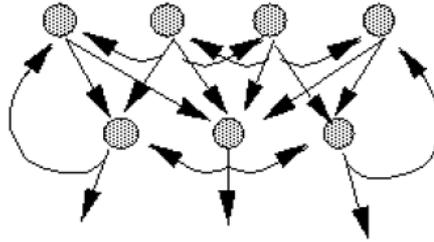


Figure.II.4. Réseau à connexions récurrentes [TOU 92]

d- Réseau à connexion complète : [DRE 04]

C'est la structure d'interconnexion la plus générale (figure.II.5). Chaque neurone est connecté à tous les neurones du réseau (et à lui-même).

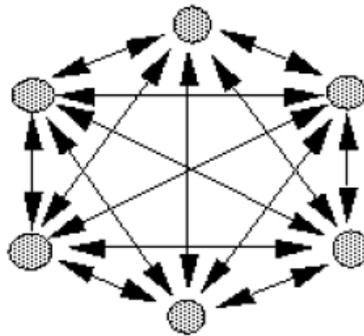


Figure.II.5. Réseau à connexions complètes [TOU 92]

Du point de vue circulation de l'information, on distingue deux types de réseaux de neurones : les réseaux non bouclés et les réseaux bouclés.

– **Les réseaux de neurones non bouclés :**

Un réseau de neurones non bouclé réalise une (ou plusieurs) fonction(s) algébrique(s) de ses entrées, par composition des fonctions réalisées par chacun de neurones.[DRE 04]

Un réseau de neurones non bouclé est donc représenté graphiquement par un ensemble de neurones « connectés » entre eux, l'information circulant des entrées vers les sorties sans

« retour en arrière » : si l'on représente le réseau comme un graphe dont les nœuds sont les neurones et les arêtes les « connexions » entre ceux-ci, le graphe d'un tel réseau est acyclique. On peut donc imaginer un grande variété de topologies pour ces réseaux. Néanmoins, la très grande majorité des applications des réseaux de neurones mettent en jeu des réseaux à couches.

Ce réseau de neurones non bouclé à n entrées, N_c neurones cachés et N_o neurones de sortie réalise N_o fonctions non linéaires des ses n variables d'entrée par composition des N_c fonctions algébriques réalisées par ses neurones cachés.[DRE 04]

– **Les réseaux de neurones bouclés :**

Un réseau de neurone bouclé à temps discret réalise une (ou plusieurs) équation(s) aux différences non linéaires, par composition des fonctions réalisées par chacun des neurones et des retards associés à chacune des connexions.[DRE 04]

C'est l'architecture la plus générale pour un réseau de neurones, dont le graphe des connexions est cyclique : lorsque l'on se déplace dans le réseau en suivant le sens des connexions, il est possible de trouver au moins un chemin qui revient à son point de départ. La sortie d'un neurone du réseau peut donc être fonction d'elle même ; cela n'est évidemment concevable que si la notion de temps est explicitement prise en considération.

Ainsi, à chaque connexion d'un réseau bouclé, est attaché, outre un poids comme pour les réseaux non bouclés, un retard, multiple entier (éventuellement nul) de l'unité de temps choisie. Une grandeur, à un instant donné, ne pouvant pas être fonction de sa propre valeur au même instant, tout cycle du graphe du réseau doit avoir un retard non nul.[DRE 04]

Propriété :

Tout réseau de neurones bouclé, aussi complexe soit-il, peut être mis sous une forme canonique, comportant un réseau de neurones non bouclé dont certaines sorties sont ramenées aux entrées par des bouclages de retard unité.[DRE 04]

En résumé, nous avons distingué et défini :

- les réseaux de neurones et leurs éléments constitutifs (le neurone formel), ainsi que l'analogie faite entre le modèle biologique et le modèle mathématique.
- les réseaux de neurones bouclés, statiques, qui réalisent des fonctions non linéaires ;

- les réseaux de neurones bouclés, qui réalisent des équations aux différences non linéaires.

Nous avons vu également que tout réseau bouclé peut être ramené à un réseau non bouclé. L'élément de base est donc le réseau de neurones non bouclé ; voilà pourquoi l'étude de ce dernier est fondamentale.

Après un exposé des définitions essentielles concernant les réseaux de neurones, le principe de leur utilisation reste à situer.

II. Principe de base :

L'idée principale des réseaux de neurones consiste à modéliser une grandeur mesurable (physique, chimique, économique...), dont on soupçonne qu'elle dépend d'un ou plusieurs facteurs qui constituent les composantes du vecteur d'entrée.

C'est pour cette raison que le domaine d'application des réseaux de neurones ne cesse de s'étendre : reconnaissance de formes, classification, analyse et diagnostic, prévision (économie, finance...).

L'objectif de la modélisation par les réseaux de neurones est de trouver, à partir des mesures disponibles, une relation, si elle existe, entre les variables d'entrées et la grandeur à modéliser. Le cadre conceptuel adapté à cette tâche est celui des statistiques.

En effet, le réseau n'a aucune idée à priori sur le modèle : on choisit une forme d'équation aussi générale que possible, et l'on ajuste les paramètres de cette équation, par apprentissage, de manière à lui conférer la meilleure capacité de généralisation sans s'intéresser particulièrement aux valeurs ou au sens de ces paramètres. On parle de modélisation « boîte noire ».[DRE 04]

Telles sont donc les propriétés fondamentales des réseaux de neurones : généralisation et apprentissage. Essayons de développer ces deux concepts.

III. Propriétés des réseaux de neurones :

1. Approximation universelle (capacité de généralisation) :

Les réseaux de neurones font partie de la classe de modèles appelés approximateurs universels au même titre que les systèmes flous, les fonctions Splines ou les polynômes.

La capacité de généralisation est l'aptitude du modèle à agir par construction ; il n'a aucune idée préconçue sur une nouvelle information, mais il peut la relier à des informations qu'il connaît déjà (l'apprentissage que nous allons développer plus tard).

Ces approximateurs ne supposent aucune hypothèse sur le modèle liant les variables explicatives aux variables expliquées.[DRE 04]

Définition :

Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones d'une certaine structure.[DRE 04]

L'avantage pratique des réseaux de neurones par rapport aux techniques classiques de modélisation non linéaire réside dans leur capacité à réaliser des modèles de précision équivalente avec moins de données expérimentales. Pour obtenir un modèle non linéaire de précision donnée, un réseau de neurones a besoin de moins de paramètres ajustables (principe de parcimonie).

2. L'apprentissage des réseaux de neurones :

Parmi les propriétés nécessaires pour un réseau de neurones, la plus fondamentale est sûrement la capacité d'apprendre de son environnement, d'améliorer sa performance à travers un processus d'apprentissage.

Dans le contexte des réseaux de neurones artificiels, nous adopterons un point de vue pragmatique en proposant la définition suivante :

« L'apprentissage est un processus dynamique et itératif permettant de modifier les paramètres d'un réseau en réaction avec les stimuli qu'il reçoit de son environnement. Le type d'apprentissage est déterminé par la manière dont les changements de paramètres surviennent ».[PAR 04]

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

On peut aussi définir l'apprentissage des réseaux de neurones comme étant :

« La procédure qui consiste à estimer les paramètres des neurones du réseau, afin que celui-ci remplisse au mieux la tâche qui lui est affectée ».[DRE 04]

Dans la plupart des architectures, l'apprentissage se traduit par une modification de l'efficacité synaptique, c'est-à-dire par un changement dans la valeur des poids qui relient les neurones d'une couche à l'autre. Soit le poids $w_{i,j}$ reliant le neurone i à son entrée j . Au temps t , un changement $\Delta w_{i,j}(t)$ de poids peut s'exprimer simplement de la façon suivante :

$$\Delta w_{i,j}(t) = w_{i,j}(t + 1) - w_{i,j}(t) \quad (\text{II.2})$$

et par conséquent :

$$w_{i,j}(t+1) = w_{i,j}(t) + \Delta w_{i,j}(t) \quad (\text{II.3})$$

avec $w_{i,j}(t+1)$ et $w_{i,j}(t)$ représentant respectivement les nouvelle et ancienne valeurs du poids $w_{i,j}$.[PAR 04]

Un ensemble de règles bien définies permettant de réaliser un tel processus d'adaptation des poids constitue ce qu'on appelle l'algorithme d'apprentissage du réseau.

Dans le cadre des définitions énoncées ci-dessus, on peut distinguer plusieurs types d'apprentissages et différents principes pouvant guider l'apprentissage d'un réseau de neurones. Du point de vue principe d'apprentissage, on distingue entre l'apprentissage « supervisé » et l'apprentissage « non supervisé ».

a- L'apprentissage supervisé :

Nous avons vu qu'un réseau non bouclé réalise une relation algébrique entre ses entrées et ses sorties. On peut donc affecter à un tel réseau la tâche qui consiste à réaliser une fonction algébrique non linéaire, pour laquelle on dispose de valeurs, en nombre fini, qui

résultent de mesures effectuées sur un processus physique, chimique, économique ou autres : le réseau réalise alors une tâche de modélisation statique ou une régression.

On connaît donc, en tout point ou seulement en certains points, les valeurs que doit avoir la sortie du réseau en fonction des entrées correspondantes : c'est en ce sens que l'apprentissage est « supervisé », métaphore qui signifie qu'un « professeur » peut fournir au réseau des « exemples » de ce que celui-ci doit faire. Pour cette raison, l'apprentissage supervisé est aussi qualifié d'apprentissage par des exemples.[DRE 04]

L'environnement est inconnu du réseau. Celui-ci produit un stimulus qui est acheminé à la fois au professeur et au réseau. Grâce à ses connaissances intrinsèques, le professeur produit une sortie désirée pour ce stimulus. On suppose que cette réponse est optimale. Elle est ensuite comparée (par soustraction) avec la sortie du réseau pour produire un signal d'erreur qui est ré-injecté dans le réseau pour modifier son comportement via une procédure itérative qui, éventuellement, lui permet de simuler la réponse du professeur. Autrement dit, la connaissance de l'environnement par le professeur est graduellement transférée vers le réseau jusqu'à l'atteinte d'un certain critère d'arrêt. Par la suite, on peut éliminer le professeur et laisser le réseau fonctionner de façon autonome.[PAR 04]

L'algorithme d'apprentissage le plus utilisé en pratique est celui de la rétropropagation des erreurs (algorithme de « backpropagation »)

b- L'apprentissage non supervisé :

L'apprentissage «non supervisé» ou encore «auto-organisé» est caractérisé par l'absence complète de professeur, c'est-à-dire qu'on ne dispose pas d'un signal d'erreur, comme dans le cas supervisé. Nous ne disposons donc que d'un environnement qui fournit des stimuli, et d'un réseau qui doit apprendre sans intervention externe. En assimilant les stimuli de l'environnement à une description de son état interne, la tâche du réseau est alors de modéliser cet état le mieux possible. Pour y arriver, il importe d'abord de définir une mesure de la qualité pour ce modèle, et de s'en servir par la suite pour optimiser les paramètres libres du réseau, c'est-à-dire ses poids synaptiques. A la fin de l'apprentissage, le réseau a développé une habilité à former des représentations internes des stimuli de l'environnement permettant d'encoder les caractéristiques de ceux-ci et, par conséquent, de créer automatiquement des classes de stimuli similaires.[PAR 04]

Nous venons de voir les principaux types d'apprentissage. Pour chacun de ces types, il existe plusieurs règles qui se distinguent par la manière dont les changements de paramètres surviennent. Les règles les plus utilisées sont les suivantes : (cf. Annexe 8)

- La règle de Hebb ;
- La règle de correction de l'erreur ;
- La règle de Widrow-Hoff
- La règle d'apprentissage compétitif.

IV. Les différents types de réseaux de neurones :

On peut classer les réseaux de neurones artificiels en deux grandes catégories que sont, les réseaux dit « feed-forward » ou « feed-back » (encore appelés réseaux récurrents). Comme leurs noms l'indiquent, dans les réseaux « feed-forward », l'information se propage de couche en couche sans retour en arrière possible, contrairement aux réseaux « feed-back »

1. Les réseaux « feed-forward » :

Ce sont les réseaux dans lesquels l'information circule à sens unique, ils peuvent se répartir comme suit :

a. Le Perceptron mono-couche : [DAV 93]

C'est historiquement le premier réseau de neurones artificiels, c'est le Perceptron de Rosenblatt. C'est un réseau simple, puisque il ne se compose que d'une couche d'entrée et d'une couche de sortie .

Il est calqué, à la base, sur le système visuel et de ce fait a été conçu dans un but premier de reconnaissances de formes. Cependant, il peut aussi être utilisé pour faire de la classification et pour résoudre des opérations logiques simples (telle « ET » ou « OU »). Sa principale limite est qu'il ne peut résoudre que des problèmes linéairement séparables. Il suit généralement un apprentissage supervisé selon la loi de Hebb.

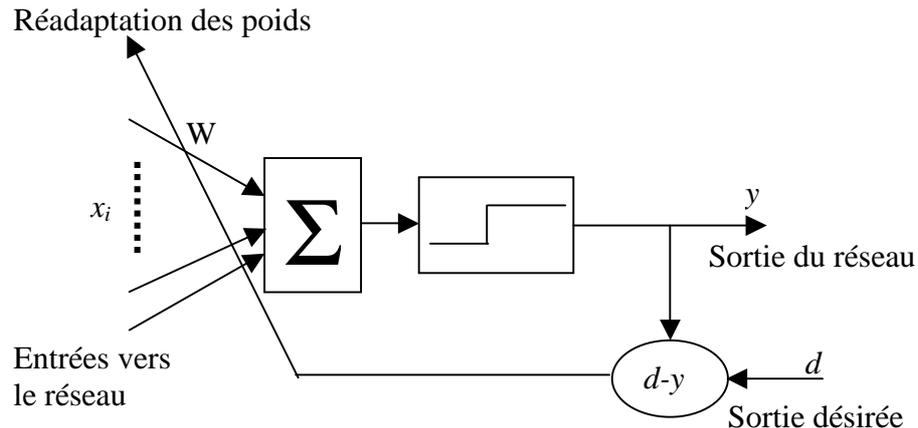


Figure.II.6. Structure du perceptron linéaire [DAV 93]

L'activité des neurones dans ce type de réseaux est définie par l'équation linéaire suivante :

$$net_i = \sum_{j=1}^n x_j w_{ij} + \theta_i \quad i=1 \dots n \quad (\text{II.4})$$

x_j représentent les entrées des neurones

w_{ij} le poids synaptique en provenance du $j^{\text{ème}}$ neurone de la couche d'entrée vers le $i^{\text{ème}}$ neurone de sortie.

θ_i est un paramètre qui agit sur le seuil en sortie.

La sortie est définie par la fonction seuil :

$$y = S_i(net_i) = \begin{cases} 1 & net_i > 0 \\ 0 & net_i \leq 0 \end{cases} \quad (\text{II.5})$$

L'adaptation des poids de ces réseaux se fait comme suit :

$$w(t+1) = w(t) + \Delta w(t) \quad (\text{II.6})$$

$$\text{Tel que :} \quad \Delta w(t) = \eta (d(t) - y(t)) \cdot X(t) \quad (\text{II.7})$$

Où :

η : est un paramètre qui contrôle le taux d'apprentissage.

$X(t)$: représente le vecteur présenté à l'entrée du réseau ; (x_1, \dots, x_n)

$y(t)$: est la sortie du réseau.

$d(t)$: est la sortie désirée.

L'entraînement du réseau se poursuit jusqu'à ce que chaque exemple soit correctement classé. Mais en 1969, Papert et Minsky démontrent les limites du perceptron. Ce dernier n'est efficace que pour des problèmes linéairement séparables.

Une amélioration de ce perceptron à une ou plusieurs couches de neurones cachés a donné naissance au perceptron multicouches.

b. Adaline, Madaline : [DAV93]

En 1960, B.Widrow, un automaticien, développe le modèle Adaline (Adaptative Linear Element). Dans sa structure, le modèle ressemble au Perceptron, une couche d'entrée connectée à une ou plusieurs sorties. Dans le cas de plusieurs sorties, il est alors appelé Madaline (Multiple Adaptive Linear Element). Ce type de réseau est destiné à l'approximation des fonctions.

Cependant la loi d'apprentissage est différente. Celle-ci est à l'origine de l'algorithme de rétropropagation de gradient très utilisé aujourd'hui avec les perceptrons multicouches.

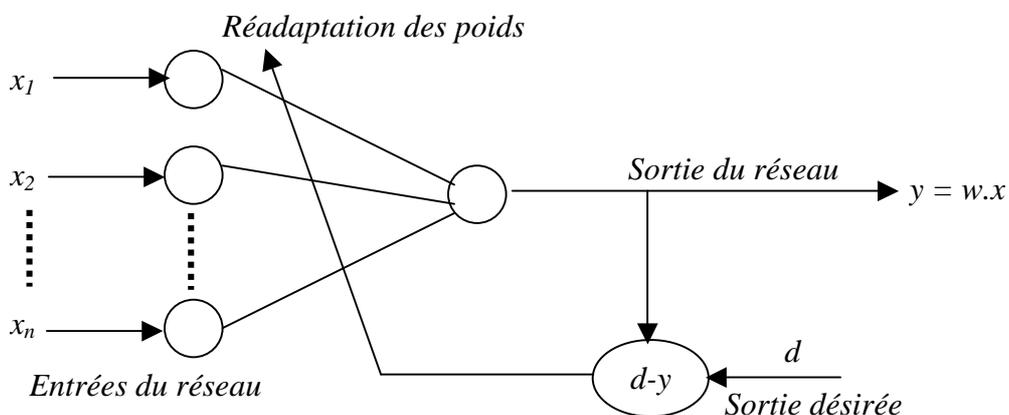


Fig.II.7. Structure du réseau ADALINE [DAV 93]

L'apprentissage du Madaline s'effectue grâce à la minimisation d'une fonction coût calculée à partir des exemples de la base d'apprentissage et de la sortie du réseau de neurones. Cette fonction détermine l'objectif à atteindre.

A la présentation d'une entrée x_i^k - k étant l'exemple présenté ($k=1, \dots, n_{ex}$) avec n_{ex} le nombre d'exemples - la sortie y_i^k est le produit scalaire de cette entrée avec un vecteur poids w_{ij} (Fig.II.7) :

$$y_i^k = \sum_{j=1}^n x_j w_{ij} + \theta_j \quad i=1, \dots, m \quad (\text{II.8})$$

x_j : l'entrée du $j^{\text{ème}}$ neurone.

w_{ij} : poids synaptique du $i^{\text{ème}}$ neurone de la couche de sortie provenant du $j^{\text{ème}}$ neurone de la couche d'entrée.

θ_j : biais de la $j^{\text{ème}}$ sortie.

On définit l'erreur instantanée e_j^k relative à chaque sortie par la différence entre la valeur de la sortie du réseau calculée y_j^k et la sortie désirée d_j^k par :

$$e_j^k = d_j^k - y_j^k \quad (\text{II.9})$$

Le problème réside dans l'ajustement des poids w_{ij} de telle sorte que l'écart entre la sortie désirée et la sortie calculée par le réseau soit nul pour tous les exemples. ($k = 1, \dots, n_{ex}$ avec n_{ex} le nombre d'exemples).

c. Le Perceptron multicouches (PMC) : [PAR 04]

C'est une extension du perceptron, avec une ou plusieurs couches cachées entre l'entrée et la sortie. Chaque neurone dans une couche est connecté à tous les neurones de la couche précédente et de la couche suivante (excepté pour les couches d'entrée et de sortie), et il n'y a pas de connexions entre les cellules d'une même couche. Les fonctions d'activation utilisées dans ce type de réseaux sont principalement les fonctions à seuil ou sigmoïdes. Il peut résoudre des problèmes non linéairement séparables et des problèmes logiques plus

compliqués, et notamment le fameux problème du XOR. Il suit aussi un apprentissage supervisé selon la règle de correction de l'erreur. Les principaux domaines d'application des PMC sont : la classification non linéaire, l'approximation de fonctions, le contrôle et la reconnaissance de formes et de paroles.

Ce type de réseau a des aptitudes à la généralisation, c'est-à-dire qu'il peut produire une réponse correcte (ou du moins presque correcte) pour une entrée qui ne lui a pas été présentée pendant la phase d'apprentissage. Le réseau peut être assimilé à une fonction d'entrée-sortie non linéaire, l'apprentissage ayant pour but de mettre au point cette fonction à partir d'un ensemble (judicieusement) choisi d'exemples. Cet apprentissage se fait dans la plupart des cas selon l'algorithme de rétropropagation de l'erreur. (*cf. Annexe 9*)

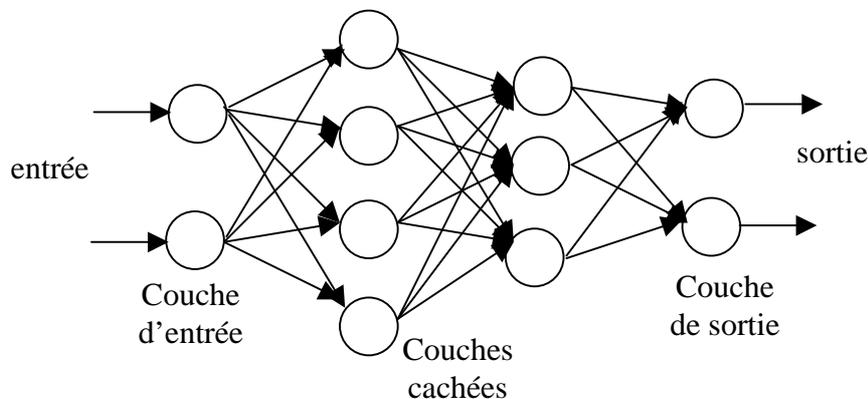


Fig.II.8. Architecture d'un réseau multicouches [PAR 04]

d. Les réseaux à fonction radiale : [DAV 93]

Ce sont les réseaux que l'on nomme aussi RBF (pour « Radial Basic Functions »). L'architecture est la même que pour les PMC cependant, les fonctions de base utilisées ici sont des fonctions Gaussiennes. Les RBF seront donc employés dans les mêmes types de problèmes que les PMC à savoir, en classification et en approximation de fonctions, particulièrement. L'apprentissage le plus utilisé pour les RBF est le mode hybride et les règles sont : soit, la règle de correction de l'erreur, soit la règle d'apprentissage par compétition. Les principaux domaines d'application des RBF sont : la classification non linéaire, l'approximation de fonctions, le contrôle et la reconnaissance de formes et de paroles.

2. Les réseaux « feed-back » :

Ce sont les réseaux dans lesquels l'information n'a pas de sens particulier de circulation fixé à l'avance et peut donc faire des allers-retours. Ils peuvent se répartir comme suit :

a. Les cartes auto-organisatrices de Kohonen : [PAR 04]

Ce type de réseaux de neurones vient de l'analogie faite avec le cerveau, qui démontre qu'il existe des zones dans le cerveau qui ont la même topologie que les capteurs auxquels elles sont reliées, par exemple deux zones proches dans le cortex visuel correspondront à deux zones proches dans la rétine .

Ce type de réseau, aussi appelé SOM (pour « Self Organized Maps »), est un réseau non supervisé à apprentissage compétitif qui établit une carte discrète, ordonnée topologiquement, en fonction de règles d'entrée. En effet une carte auto-organisatrice, comme son nom l'indique, s'organise par rapport aux stimuli présentés en respectant les contraintes topologiques de l'espace d'entrée. Il y a mise en correspondance de l'espace d'entrée avec l'espace du réseau. Les zones voisines de l'espace d'entrée sont voisines sur la carte auto-organisatrice. Les domaines d'application de ce type de réseaux est la projection non linéaire de données et la compression de données.

b. Les réseaux de Hopfield : [TOU 92]

Les réseaux de Hopfield sont des réseaux récurrents et entièrement connectés. Dans ce type de réseaux, chaque neurone est connecté à chaque autre neurone et il n'y a aucune différenciation entre les neurones d'entrée et de sortie. Ils fonctionnent comme une mémoire associative non-linéaire et sont capables de trouver un objet stocké en fonction de représentations partielles ou bruitées. L'application principale des réseaux de Hopfield est l'entrepôt de connaissances, mais aussi la résolution de problèmes d'optimisation. Le mode d'apprentissage utilisé ici est le mode non-supervisé.

c. Les ART : [TOU 92]

Les réseaux ART (pour « Adaptive Resonance Theory ») sont des réseaux à apprentissage par compétition. Le problème majeur qui se pose dans ce type de réseaux est le dilemme « stabilité /plasticité ». En effet, dans un apprentissage par compétition, rien ne garantit que les catégories formées vont rester stables. La seule possibilité, pour assurer la stabilité, serait que le coefficient d'apprentissage tende vers zéro, mais le réseau perdrait alors sa plasticité. Les ART ont été conçus spécifiquement pour contourner ce problème. Dans ce genre de réseaux, les vecteurs de poids ne seront adaptés que si l'entrée fournie est suffisamment proche, d'un prototype déjà connu par le réseau. On parlera alors de résonance.

A l'inverse, si l'entrée s'éloigne trop des prototypes existants, une nouvelle catégorie va alors se créer, avec pour prototype, l'entrée qui a engendré sa création. Il est à noter qu'il existe deux principaux types de réseaux ART : les ART-1 pour des entrées binaires et les ART-2 pour des entrées continues. Le mode d'apprentissage des ART peut être supervisé ou non.

V. Méthodologie de conception : [DRE 04]

« Pour une technique nouvelle, l'existence d'une méthodologie de mise en œuvre est un critère de maturité et une garantie de succès dans ses applications ».[DRE 04]

Pour cela, nous allons introduire une méthodologie de conception de modèle. Nous avons montré que les réseaux de neurones sont susceptibles de constituer des modèles non linéaires, statiques ou dynamiques, le plus souvent des modèles « boîte noire ».

Le problème de la modélisation est d'établir les relations entre la grandeur à modéliser et les facteurs mesurables qui la déterminent. On peut ainsi prédire l'évolution de cette grandeur si l'on connaît l'évolution de ses facteurs.

Le problème se ramène donc à l'estimation des paramètres du modèle le plus satisfaisant possible.

La conception d'un modèle non linéaire « boîte noire » comprend plusieurs tâches :

1. la sélection des entrées du modèle, c'est à dire la réduction de la dimension des données d'entrée et l'élimination des variables non pertinentes.
2. l'estimation des paramètres du modèle, c'est à dire l'apprentissage.
3. la sélection du meilleur modèle et l'estimation des performances des modèles.

L'ensemble de ces étapes constitue une méthodologie complète de conception de modèles non linéaires.

Pour peu que l'ingénieur soit capable d'exprimer le problème à résoudre sous la forme d'une fonction de mise en correspondance et qu'il dispose d'exemples pour l'apprentissage, les performances d'une modélisation par les réseaux de neurones artificiels dans le domaine de la généralisation du comportement à des situations inconnues (ou imprévues) sont tout à fait intéressantes. Citons pour mémoire le diagnostic, la prévision, la classification, la reconnaissance de formes, etc. Même lorsque les performances ne sont pas supérieures à celles d'une approche plus classique (algorithmique ou à base de connaissance), la facilité de programmation des modèles neuronaux par l'exemple autorise le développement d'applications sans requérir une connaissance très exhaustive du domaine par l'ingénieur.

Nous avons, à travers ce chapitre, développé les fondements de cet outil récent de modélisation ainsi que ses propriétés essentielles qui font de lui un outil d'aide à la décision très prometteur. En effet, dans le domaine des prévisions par exemple, il concurrence rudement les méthodes classiques. C'est alors dans ce but que, lors de notre application, nous confronterons cet outil de la famille de l'intelligence artificielle à une méthode classique de prévision puis nous tirerons les conclusions qui découleront de cette confrontation.

Chapitre III : Collecte et traitement des données

Après avoir présenté un aperçu sur les différentes techniques statistiques d'analyse et de prévision, ainsi que les nouvelles approches de prévision que représentent les réseaux de neurones artificiels, nous allons dans ce chapitre exposer la méthodologie de résolution du problème, objet de notre étude, défini comme étant l'élaboration d'un modèle de prévision.

Mais tout d'abord, intéressons nous à ce qui a été fait dans le domaine, les résultats obtenus ainsi que les différentes démarches entreprises qui nous serviront de guide pour l'élaboration d'un modèle de prévision.

Dans ce qui suit, nous passerons en revue quelques travaux entrepris dans le domaine de la prévision énergétique. Cela nous permettra de répondre à une multitude de questions concernant l'utilisation des réseaux de neurones pour la prévision (sélection des entrées, type de réseau à utiliser, type d'apprentissage,...) et de justifier la comparaison que nous essayons d'établir.

I. Synthèse de quelques travaux :

Nous nous sommes intéressés en entreprenant nos recherches au domaine de la prévision. Vaste domaine qui traite aussi bien des marchés boursiers que de la météorologie, en passant par le domaine énergétique, l'hydrologie ou même le trafic aérien ou ferroviaire.

Mais toutes ces études, en particulier les plus récentes, font ressortir l'intérêt croissant des réseaux de neurones dans la prévision de phénomènes difficilement modélisables par les approches classiques. En effet, l'analyse des chroniques, utilisant des méthodes statistiques ou se basant sur l'analyse d'experts dans le domaine traité, se tourne de plus en plus vers l'intelligence artificielle pour palier aux manques de ces différentes méthodes.

Nous citerons les travaux de J.YAO et C.L.TAN [YAO 00] qui se sont intéressés à l'application des réseaux de neurones artificiels pour la prévision des taux de change (différents indices de taux de change). Ces variables étant très aléatoires et répondant rarement à une loi bien précise, les auteurs ont prouvé que l'intelligence artificielle n'était pas

seulement applicable dans ce cas de figure, mais qu'elle donnait de bien meilleurs résultats en comparaison avec des modèles traditionnellement utilisés dans l'analyse de ce type de séries temporelles, les modèles ARMA.

L'engouement face à ce nouvel outil d'analyse est justifié. A.N.REFENS [REF 94] indique même que les méthodes statistiques conventionnelles ont atteint leur limite dans leur application face à des phénomènes économiques, entre autres, présentant de plus en plus de non-linéarité.

En effet, P.A.FISHWICK [FIS 99], en n'utilisant qu'un seul type de réseaux, le perceptron multicouche (PMC), et en appliquant ce dernier à des séries précédemment traitées par la méthode de Box & Jenkins, obtient sur 75% des séries traitées des résultats significativement meilleurs.

Ainsi, étant donné que la prévision n'est rien d'autre qu'une approximation de fonctions généralement inconnues, les PMC et les RBF qui sont des réseaux à couches et à apprentissage supervisé (*cf. Chapitre II*), sont les mieux placés pour le traitement de problèmes de prévision.[COU 99]

Pour arrêter notre choix sur le type de réseau que nous utiliserons dans le cadre de notre travail, nous nous sommes basés sur plusieurs travaux.

Nous citerons ainsi, une étude comparative de P.Coulibaly, F.Anctil et B.Bobee. Cette dernière rassemble plusieurs travaux qui traitent de l'application des réseaux de neurones dans le domaine de la prévision et à travers lesquels les auteurs ont tiré diverses conclusions, les plus importantes sont les suivantes:[COU 99]

- Pour des prévisions à moyen et long terme, les composantes non linéaires des systèmes et le nombre de variables explicatives prennent plus d'importance, ce qui expliquerait la bonne performance des PMC dans plus de 90% des cas d'études comparatives.
- De nombreuses études expérimentales ont démontré le potentiel des réseaux de neurones pour la prévision de certaines séries temporelles. Mais dans 95% des cas

d'application en prévision, seuls des réseaux non bouclés « feed-forward » de type PMC ont été utilisés avec généralement l'algorithme de rétropropagation de l'erreur.

De plus, les différentes études que nous avons consultées et qui traitent de la prévision dans le domaine énergétique par les réseaux de neurones artificiels optent, dans la plupart des cas, pour l'utilisation des PMC avec rétropropagation de l'erreur.

Nous citerons, à titre d'exemple, une étude faite par D.O.Arroyo M.K.Skov et Q.Huynh [ARR 02] qui relate les résultats d'une compétition mondiale organisée par l'EUNITE (the European Network of Excellence on Intelligent Technologies for Smart Adaptive Systems) sur la prévision de la charge électrique à court terme. Les meilleurs résultats obtenus lors de cette compétition ont été réalisés, entre autres, par l'application des PMC. De plus, les PMC se trouvent être plus faciles à manipuler et mieux adaptés à la modélisation statique, objet de notre application.

Dans le domaine énergétique, le thème le plus récurrent est la prévision de la charge électrique. De plus, l'énergie électrique, d'un point de vue théorique, présente un intérêt remarquable. En effet, les caractéristiques que comporte ce type de chroniques sont très intéressantes : tendance, saisonnalités annuelle et journalière, influence de variables externes et possible non-linéarités entre les variables.

Les différentes études concernant la prévision de la charge électrique relatent les limites des méthodes statistiques classiques. En effet, la consommation d'électricité augmente de façon importante et devient de plus en plus difficile à modéliser. Le nombre de facteurs influençant cette consommation ne cesse d'augmenter. Elle devient un phénomène complexe ayant des caractéristiques de plus en plus non linéaires. [IWA 99]

Mais quelque soit l'horizon temporel, ou la méthode utilisée, la modélisation reste avant tout une démarche de recherche et de préparation. Ainsi, les études consultées ont en commun une même approche de la modélisation. Ceci se concrétise par les étapes suivantes :

- Une analyse au préalable de la série.
- Une recherche et une sélection des variables ou facteurs influençant le phénomène étudié.
- L'application de la méthode retenue en prenant en compte les différents paramètres du modèle.

C'est donc la démarche que nous allons suivre pour la conception d'un modèle de prévision de la charge électrique. Nous allons dans ce qui suit développer cette démarche.

II. Sélection des entrées :

Quelque soit la méthode utilisée - dans notre cas la méthode de Box & Jenkins et les réseaux de neurones - la démarche de conception d'un modèle statique de prévision comprend une première tâche essentielle : la sélection des entrées du modèle. Le reste du chapitre est consacré à cette étape préliminaire de la modélisation qui, cependant, s'avère déterminante pour la suite de notre travail.

L'objet de notre étude est de modéliser à des fins de prévision la consommation électrique. Pour cela nous disposons d'un historique des puissances mensuelles maximales. Ces puissances maximales sont plus communément appelées « des pics ». Cet historique représente l'évolution de ces pics mensuels de Janvier 2000 à Décembre 2005 (cf. *Annexe 10*). Le graphe ci-dessous représente l'évolution de ces puissances appelées.

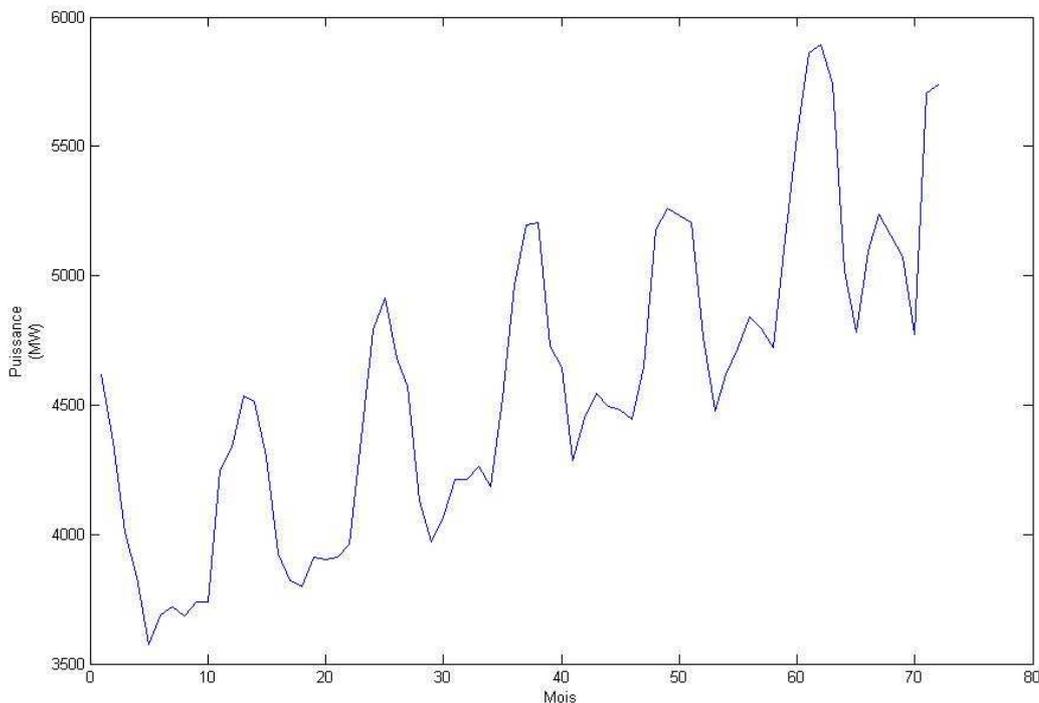


Figure.III.1 : Evolution des puissances mensuelles appelées maximales

En observant le graphique IV.1, nous pouvons en déduire les caractéristiques suivantes :

- La série présente une composante tendancielle croissante ;
- L'évolution des puissances appelées présente des variations saisonnières de période 12 (1 an). La série présente donc une composante saisonnière.

Ce premier traitement s'avère indispensable pour la construction d'un modèle de prévision. En effet, les caractéristiques de la série temporelle ainsi analysée, nous permettent de mieux comprendre la variable étudiée. De plus, cette analyse nous renseigne sur le type de facteurs influençant le phénomène à prédire.

Nous pouvons citer certains facteurs influençant la consommation électrique :

– *La météorologie :*

Les données météorologiques sont les premiers éléments influençant la consommation d'électricité. Ainsi, l'évolution des conditions météorologiques tout au long de l'année est, pour une bonne part, à l'origine des variations cycliques annuelles de la consommation. Nous pouvons utiliser une grandeur en particulier : la température. Ses variations se traduisent par des modifications de l'utilisation du chauffage électrique ou de la climatisation en été.

– *Population :*

Il est parait évident de penser que la demande en énergie électrique croit en proportion avec l'augmentation de la population.

– *L'activité économique :*

L'activité économique des entreprises a également une forte influence sur la consommation. Ses effets sont directement observables sur les courbes de consommation annuelle ou l'on remarque un creux de consommation au moment des vacances d'été.

Seulement, dans notre cas, cette variable n'est pas utilisable. En effet, les données relatives aux puissances appelées maximales sont des données mensuelles, ce qui ne correspond pas à plusieurs éventuelles variables explicatives.

La plupart des données économiques (indice de production, valeur ajoutée par secteur,...) ou même macro-économiques (parc logements, taux d'occupation des logements, niveau d'équipement des ménages,...) sont exprimées en pas de temps annuel ou trimestriel.

– *Comportement des consommateurs :*

Des entretiens avec certains cadres de la SONELGAZ nous ont amené à prendre en considération certains changements dans les habitudes des consommateurs.

En effet, nous avons remarqué ces dernières années, une augmentation du niveau d'équipement des ménages. En effet, des grands groupes industriels installent leur filiales sur le territoire national. On assiste aussi à une forte augmentation des ventes des produits électroménagers, ainsi qu'à une pénétration de plus en plus importante de l'outil informatique dans les foyers. De plus, le phénomène le plus répandu ces dernières années est le climatiseur. En effet, de plus en plus de foyers s'équipent de climatiseurs. De ce fait, cet équipement, grand consommateur d'électricité, est de plus en plus présent sur le marché national (*cf. Annexe 11 : Importations de climatiseurs*).

Nous avons donc essayé de prendre en considération ces nouvelles données. Nous nous sommes concentrés sur des facteurs s'exprimant mensuellement pour aborder le problème de la modélisation.

Ces différents facteurs ne sont probablement pas les seuls facteurs influençant la demande en énergie électrique, mais ils représentent les facteurs les plus déterminants à priori.

Afin de collecter les données relatives à cette première analyse, nous nous sommes rapprochés de la SONELGAZ, de l'Office National des Statistiques (ONS), du Centre National de l'Informatique et des Statistiques de la douane (CNIS), de l'Office National de la Météorologie (ONM), ...d'où nous avons pu rassembler les données relatives aux évolutions mensuelles de Janvier 2000 à Décembre 2005 des facteurs suivants :

- Population ;
- Population urbaine ;
- Nombre d'abonnés SONELGAZ ;
- Température moyenne ;
- Importation de climatiseurs.

Par ailleurs, une bonne méthode de prévision est celle qui peut établir la relation entre la consommation d'électricité et les variables exogènes dont elle dépend. Ainsi, lorsqu'on modélise un processus économique, une étape de sélection systématique des entrées est nécessaire.

En effet, à chaque entrée du modèle correspond un nombre de paramètres. Il est donc nécessaire que l'ensemble des entrées contienne tous les facteurs pertinents par rapport à la grandeur à modéliser, et seulement ceux-ci. La présence de variables qui ont, sur la grandeur à modéliser, moins d'influence que les perturbations ou le bruit, entraîne celle de paramètres superflus et introduit des variations non significatives sur les variables d'entrée qui constituent autant de sources d'erreur pour le modèle.[DRE 04]

Pour cela, la stratégie la plus naturelle consiste à démarrer d'un ensemble de variables candidates aussi grand que possible (modèle complet), à comparer les performances de celui-ci avec tous les modèles dont les entrées sont des sous-ensembles de l'ensemble des variables candidates puis, à choisir le meilleur sous-ensemble selon un critère bien défini.[DRE 04]

Nous avons choisi, pour la sélection des variables pertinentes parmi l'ensemble initial de facteurs, la méthode de régression pas à pas « *stepwise regression* ». Pour l'application de la régression « *stepwise* », nous avons utilisé le logiciel XLSTAT sous MS EXCEL qui intègre un sous programme dédié à cette méthode.

Le principe est de sélectionner un nombre réduit de variables qui expliquent pourtant une quantité raisonnable de variation. A chaque pas, nous sélectionnons la variable ayant la corrélation partielle¹ la plus élevée avec la variable dépendante. On calcule ensuite une valeur de test F partiel (*test de Fisher*) pour la variable et cette valeur est comparée à une valeur critique. Aussitôt que le F partiel pour la prochaine variable à entrer est inférieur à la valeur critique, l'analyse est terminée.

¹ Un coefficient de corrélation partielle mesure la liaison entre deux variables lorsque l'influence d'une troisième (ou de plusieurs autres) est gardée constante *sur les deux variables comparées*

1. Principe de la méthode :

Pour sélectionner les entrées pertinentes, nous supposons que le modèle obéit à l'équation suivant :

$$D = E \beta + b \quad (\text{III.1})$$

D : variable dépendante.

E : variables explicatives.

β : vecteur de dimension P des paramètres inconnus du modèle.

b : vecteur du bruit.

L'algorithme de régression pas à pas introduit les prédicteurs étape par étape dans le modèle, en commençant avec la variable indépendante la plus corrélée avec D .

Etape 1 :

L'algorithme sélectionne parmi les variables indépendantes restantes celle qui fournit la plus grande réduction dans la variance résiduelle (non expliquée) de la variable dépendante, c'est-à-dire la variable dont la corrélation partielle avec D est la plus grande.

Etape 2:

Le programme produit un F-test partiel pour voir si les variables ont une quantité de variation significative par rapport aux variables restantes dans la régression.

$$F = \frac{(R_{y p_i}^2 - R_{y p}^2)(N - p - 1)}{1 - R_{y p_i}^2} \quad (\text{III.2})$$

$R_{y p_i}^2$: R^2 pour l'ensemble des prédicteurs (P) déjà dans la régression avec le prédicteur i .

$R_{y p}^2$: R^2 pour l'ensemble des prédicteurs (P) déjà dans la régression.

$$R^2 = 1 - \frac{(D - \hat{D})^2}{(D - \bar{D})^2} \quad (\text{III.3})$$

\hat{D} : prévision de la variable dépendante.

\bar{D} : moyenne de la variable dépendante.

Le programme compare la F-valeur obtenue à une valeur tabulée (F-tabulée) que l'utilisateur définira comme critère de tolérance.

Le programme évalue si oui ou non la F-valeur obtenue à une étape donnée satisfait le minimum prédéfini, et s'il le satisfait, il entre la variable :

Si F-valeur est inférieure au F-tabulée, la variable est entrée. Sinon, le programme passe à une autre variable jusqu'à ce qu'il n'y ait plus de variables dans l'ensemble initial de facteurs.

De même, le programme décide à chaque étape si, oui ou non, chaque variable précédemment incluse satisfait encore un seuil de tolérance, et dans le cas contraire, il la supprime.

2. Application de la méthode de régression pas à pas :

Le programme XLSTAT intègre les statistiques suivantes :

- Coefficients de régression : le coefficient de régression β , l'erreur standard de β , la valeur t de β . Il affiche les intervalles de confiance à 95% pour chaque coefficient de régression, ou une matrice de covariances.
- Qualité de l'ajustement : Les variables entrées et supprimées du modèle sont listées et les statistiques de la qualité de l'ajustement suivantes sont affichées : R^2 et R^2 ajusté, erreur standard pour l'estimation. Toutes les variables doivent respecter le critère de tolérance pour être entrées dans l'équation. Le niveau de tolérance par défaut est 0,0001. Le seuil de tolérance par défaut est de 0,0001. Une variable n'est pas entrée si elle fait passer la tolérance d'une autre variable déjà entrée dans le modèle en dessous du seuil de tolérance.
- Mesure et corrélations partielles.

Nous adopterons dans ce qui suit les notations suivantes pour désigner les variables explicatives (variables d'entrées) que nous avons réunies :

- Population :POP ;
- Population urbaine : P. Urbaine ;
- Importation de climatiseurs : Import . CLIM ;
- Température moyenne : T° ;
- Nombre d'abonnés : Abonnés ;
- Puissance appelée maximale : Pmax.

Les résultats de l'application sont les suivants :

Tableau.III.1-Matrice de corrélation

Variables	Abonnés	POP	P. urbaine	Import . CLIM	T°	Pmax
Abonnés	1,000	0,998	0,997	0,294	-0,061	0,751
POP	0,998	1,000	1,000	0,299	-0,054	0,745
P. urbaine	0,997	1,000	1,000	0,302	-0,053	0,744
Import . CLIM	0,294	0,299	0,302	1,000	0,313	-0,011
T°	-0,061	-0,054	-0,053	0,313	1,000	-0,597
Pmax	0,751	0,745	0,744	-0,011	-0,597	1,000

On note du tableau III.1, les corrélations partielles de chaque variable explicative avec la variable dépendante. La variable la moins corrélée avec *Pmax* est *Import. CLIM* (-0,011). Les variables les plus corrélées avec la variable *Pmax* sont : *POP* (0,745), *POP. Urbaine* (0,744) et *Abonnés* (0,751).

Cependant ces trois variables sont très fortement corrélées entre elles. De ce fait, leur introduction ne paraît pas très pertinente.

Après calcul du F partiel des facteurs les plus corrélés avec *Pmax*, et en comparant ce dernier à la valeur critique, on obtient les résultats suivants :

Tableau.III.2-Variables introduites/éliminées

Modèle	Variables introduites	Variables éliminées	Méthode
1	Abonnés	-	Pas à pas (critère: Probabilité de F pour introduire $\leq 0,050$, Probabilité de F pour éliminer $\geq 0,100$).
2	T°	-	Pas à pas (critère: Probabilité de F pour introduire $\leq 0,050$, Probabilité de F pour éliminer $\geq 0,100$).

Tableau.III.3-Synthèse de la sélection des variables

Nb. de variables	Variables	Variable IN/OUT	Statut	MCE	R ²	R ² ajusté
1	Abonnés	Abonn	IN	127008,432	0,565	0,557
2	Abonnés / T°	T°	IN	38833,394	0,869	0,865

On remarque dans le tableau III.2 que les variables retenues dans le modèle sont les variables *Abonnés* et *T°*. Parmi toutes les autres variables introduites, ces dernières respectent le critère du test.

De plus le tableau III.3 relate pour les modèles retenus les valeurs du R² et R² ajustées. On remarque bien que le modèle *Abonnés/T°* est plus performant.

Tableau.III.4-Paramètres du modèle

Source	Valeur	Ecart-type	t	Pr > t	Borne inférieure (95%)	Borne supérieure (95%)
Constante	-1189,468	426,627	-2,788	0,007	-2043,146	-335,790
Abonnés	0,001	0,000	15,201	< 0,0001	0,001	0,001
POP	0,000	0,000				
P. urbaine	0,000	0,000				
Import .CLIM	0,000	0,000				
T°	-42,156	3,599	-11,715	< 0,0001	-49,357	-34,956

Les tableaux III.4 et III.5 donnent respectivement les valeurs des paramètres et les paramètres normalisés de régression du modèle. Nous remarquons que toutes les variables respectent le critère de tolérance (probabilité inférieure à 0,0001). Le tableau III.5 permet de mesurer l'effet neutraliser de grandeur des chaque variables (étant donné que les variables ont des grandeurs différentes).

Tableau.III.5-Coefficients normalisés

Source	Valeur	Ecart-type	t	Pr > t	Borne inférieure (95%)	Borne supérieure (95%)
Abonnés	0,717	0,047	15,201	< 0,0001	0,623	0,812
POP	0,000	0,000				
P. urbaine	0,000	0,000				
Import . CLIM	0,000	0,000				
T°	-0,553	0,047	-11,715	< 0,0001	-0,647	-0,458

Après analyse des résultats de l'application de la méthode, le modèle final comporte les facteurs suivants :

- Le nombre d'abonnés SONELGAZ ;
- La température moyenne.

La sélection des entrées est une étape essentielle pour toute conception de modèle. Cette sélection nous a permis de réduire le nombre de variables explicatives de telle sorte que le modèle final ne contienne que les variables les plus pertinentes pour la modélisation.

En effet, puisque plusieurs variables peuvent présenter une certaine corrélation entre elles, donc sans sélection, des calculs inutiles seraient effectués.

Cette sélection étant effectuée, nous pouvons à présent passer à la conception du modèle de prévision. Nous commencerons par appliquer la démarche prévisionnelle définie dans le chapitre II, la méthode de Box & Jenkins, puis nous entamerons la conception du réseau neuronal que nous allons utiliser.

Chapitre IV : Conception des modèles de prévision

Habituellement, les modèles statistiques de type ARMA (ou ARIMA) dominent l'analyse des séries temporelles. Mais pour certaines applications, ces méthodes classiques ont atteint leur limites en application vis-à-vis de certaines non-linéarités que pourraient présenter des séries économiques. Les réseaux de neurones ont connu dès lors une application étendue au domaine économique.

Dans ce chapitre nous allons décrire la méthodologie adoptée afin d'aboutir à un modèle de prévision optimal et cela suivant les deux démarches présentées dans les chapitres I et II.

Nous allons tout d'abord appliquer la méthodologie de Box & Jenkins toujours dans le cadre de la prévision de la charge électrique.

I. Application de la méthode de Box & Jenkins :

La démarche de Box & Jenkins comme nous l'avons présentée au chapitre I, comporte quatre étapes :

- L'identification du processus ;
- L'estimation des paramètres ;
- La vérification et ;
- Le choix du modèle.

Cette méthodologie, s'applique au modèle ARIMA multivarié (*cf. Chapitre I*).

On note tout d'abord la série initiale comme X_t . Nous allons développer chaque étape de la démarche de Box et Jenkins.

1. Identification :

o Le choix de d et D :

Une série stationnaire fluctue autour d'une valeur moyenne et sa fonction d'autocorrélations décline rapidement vers zéro. Si une série présente des autocorrélations positives pour un grand nombre de décalages (par exemple 10 ou plus), alors elle nécessite d'être différenciée. La différenciation tend à introduire des autocorrélations négatives.

Si l'autocorrélation de décalage 1 est égale à 0 ou négative, la série n'a pas besoin d'être différenciée. Si l'autocorrélation de décalage 1 est inférieure à -0.5 , la série est surdifférenciée.[DEL 00]

L'examen du graphique (figure.III.1) fait apparaître les informations suivantes :

- Une tendance approximativement linéaire ;
- Une saisonnalité de période 12 est apparente.

Les fonctions d'autocorrélations et d'autocorrélations partielles (figure.IV.1) montrent que la série initiale ne peut être considérée comme engendrée par un processus stationnaire.

Nous devons donc procéder à une différenciation saisonnière puis régulière. La figure.IV.2 montre les fonctions d'autocorrélations et d'autocorrélations partielles engendrées par ces différenciations.

Après l'analyse des fonctions d'autocorrélations et d'autocorrélations partielles (figure.IV.3), le choix des paramètres d et D est donc comme suit: $d=1$ et $D=1$.

La série X_t s'écrira alors sous la forme I.23 en page 19, avec $d=1$; $D=1$ et $S=12$.

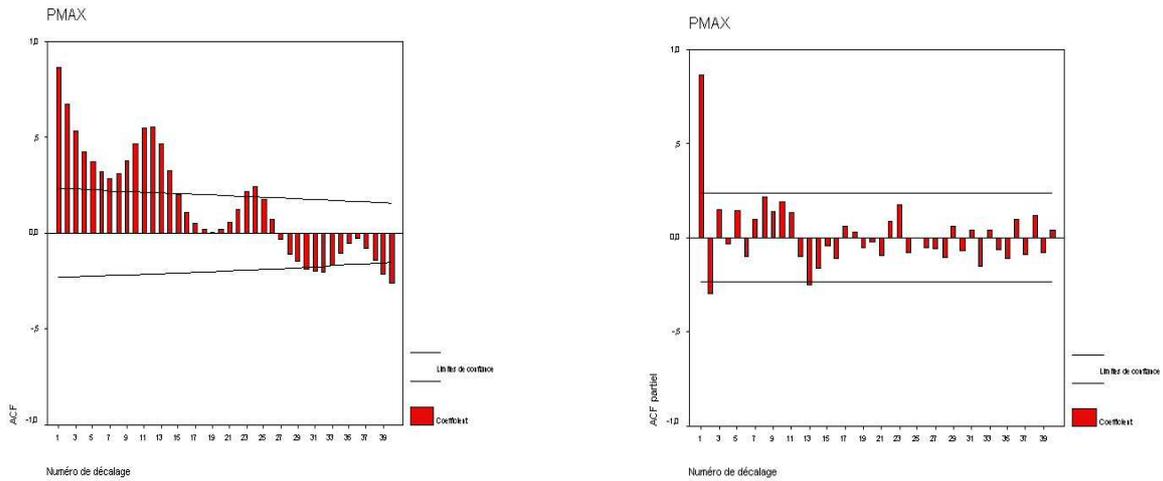


Figure.IV.1-Fonctions d'autocorrélations et d'autocorrélations partielles

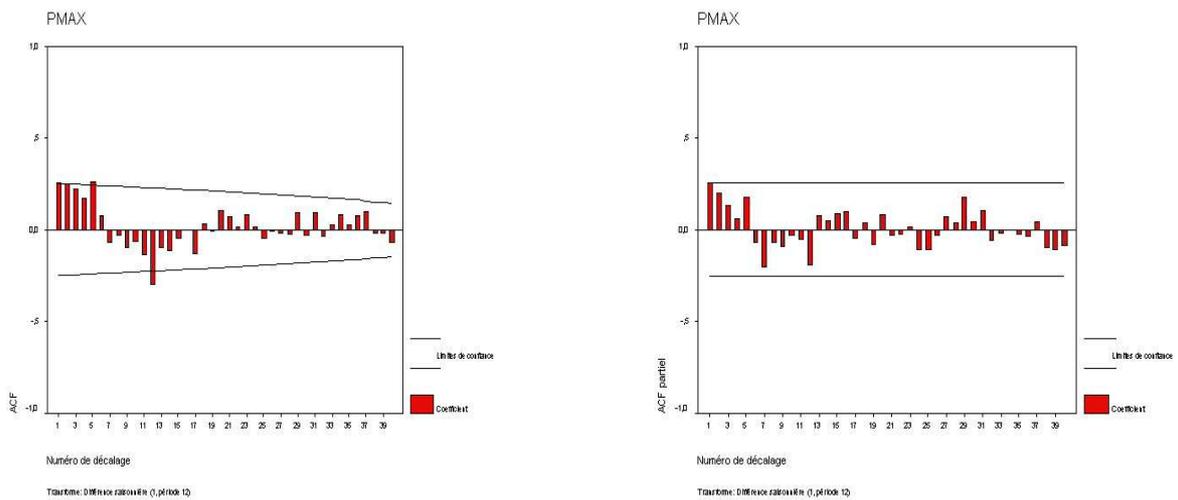


Figure.IV.2-Fonctions d'autocorrélations et d'autocorrélations partielles après différenciation saisonnière

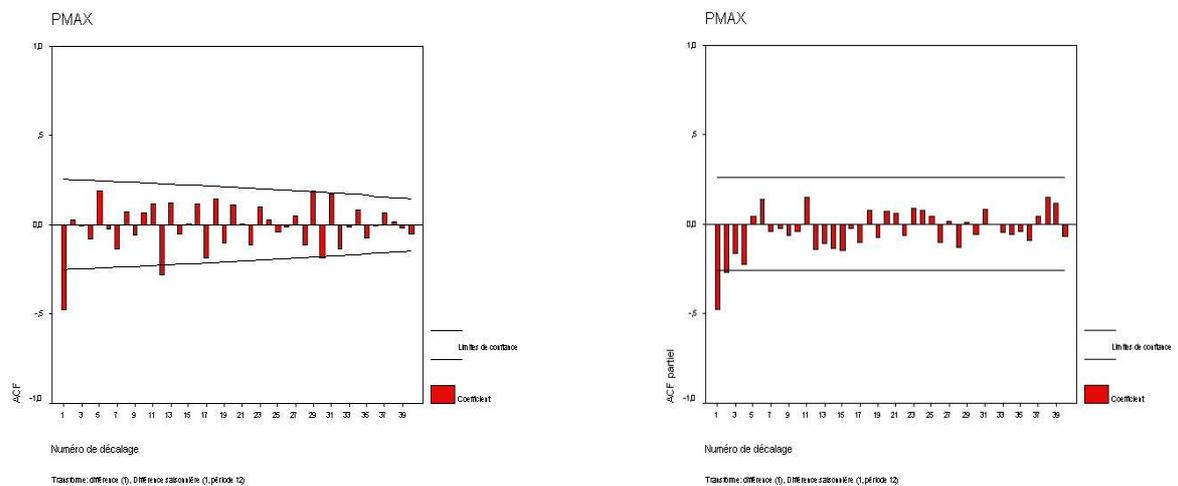


Figure.IV.3-Fonctions d'autocorrélations et d'autocorrélations partielles après différenciation.

- Choix des termes AR et MA :

Après que la série ait été stationnarisée, l'étape suivante consiste à identifier les termes AR et MA nécessaires pour corriger les autocorrélations résiduelles. Rappelons que cette analyse est basée sur l'examen des fonctions d'autocorrélations et d'autocorrélations partielles.

Les outils principaux utilisés lors de la phase d'identification sont donc les tracés de la série, les corrélogrammes d'autocorrélation, et d'autocorrélation partielle (*cf. chapitre I*). La décision n'est pas simple et les cas les plus atypiques requièrent, outre l'expérience, de nombreuses expérimentations avec des modèles différents (avec divers paramètres ARIMA).

A priori, après étude des corrélogrammes et corrélogrammes partiels estimés (Figure.V.3), ces fonctions sont typiques de la signature des processus de moyenne mobile [CHA 03]. Le pic unique est négatif pour la fonction d'autocorrélation indique un processus MA(1).

Une modélisation de type AR nécessiterait d'introduire un trop grand nombre de termes (les 4 premières valeurs de l'autocorrélogramme partiel étant significativement non-nulles).

Le modèle retenue serait un modèle SARIMA₁₂ (0,1,1) (0,1,1). Mais pour ne pas écarter certains modèles, nous allons aussi considérer le processus comme étant un modèle SARIMA₁₂ (1,1,1) (1,1,1).

2. Estimation des paramètres et vérification:

Nous avons utilisé pour l'estimation des paramètres du modèle le logiciel SSPS 8.0 qui prend en charge le calcul des paramètres des modèles ARIMA (la famille de modèles ARIMA). Ce logiciel permet donc de choisir les termes AR et MA du processus, saisonnier ou non. Grâce à ces fonctionnalités, les variables explicatives sont intégrées dans la procédure d'estimation. Le calcul se fait par le maximum de vraisemblance.

Les résultats de l'estimation du modèle $(1,1,1) (1,1,1)_{12}$ sont les suivants :

Tableau.IV.1-Résultats de l'estimation du modèle $(1,1,1)(1,1,1)_{12}$

	B¹	Ecart-Type B	T-RATIO	PROB.APPROX
AR1	-,062008	,171357	-,3618614	,71892196
MA1	,811226	,117117	6,9266365	,00000000
SAR1	-,149876	,377085	-,3974588	,69265601
SMA1	,525031	,428292	1,2258700	,22577049
T°	-23,658095	12,282713	-1,9261294	,05956294
Abonnés	,000061	,006570	,0092900	,99262334
Constante	2,636984	12,028807	,2192224	,82733505

¹ les paramètres du modèle

– Interprétations des résultats :

Comme nous l'avons expliqué au chapitre I (phase de vérification), nous allons tester la significativité des paramètres obtenus après estimation, et cela par le biais d'un test de *Student*.(I.28)

Les résultats du test sont donnés dans le tableau IV.1 ainsi que la valeur des estimateurs des paramètres du modèle et leur écarts-type.

On remarque que les coefficients AR(1), SAR(1) ne sont pas significativement différents de 0 au seuil 0,05. Leur probabilités respectives sont de 0,72 et 0,69 . De plus, la valeur du rapport de *Student* est inférieure à 1,96 (en valeur absolue) pour les deux coefficients.

On note également que le coefficient SMA(1) n'est pas significativement différent de 0 (probabilité approximative de 0,23). Le rapport de *Student* est lui aussi inférieur à 1,96.

Les paramètres du modèle $(1,1,1)(1,1,1)_{12}$ ne vérifient pas les critères de validation. Le modèle n'est pas de la forme $(1,1,1)(1,1,1)_{12}$.

La procédure à suivre, dans ce cas, est la suivante :

1. Eliminer un premier paramètre car il ne satisfait pas aux critères de validation ;
2. Refaire l'estimation avec les ordres restants ;
3. Analyser les résultats de nouveau et ;
4. Réitérer la procédure si l'un des paramètres ne satisfait pas les critères.

Nous allons éliminer le coefficient AR(1), puis recalculer les paramètres du modèle redéfini. Il s'agit alors d'un modèle $(0,1,1)(1,1,1)_{12}$.

Les résultats de l'estimation sont donnés dans le tableau suivant :

Tableau.IV.2- Résultats de l'estimation du modèle $(0,1,1)(1,1,1)_{12}$

	B	Ecart-type B	T-RATIO	PROB. APPROX
MA1	,988984	,424426	2,3301710	,02363694
SAR1	-,208359	,330611	-,6302239	,53125610
SMA1	,516576	,388844	1,3284918	,18970776
T°	-23,630845	12,676248	-1,8641830	,06783865
Abonnés	-,001802	,004737	-,3804796	,70511017
Constante	6,065853	8,981752	,6753530	,50238753

– Interprétation des résultats :

Après estimation des paramètres du nouveau modèle, on remarque que le coefficient SAR(1) n'est pas significativement différent de 0. Sa probabilité approximative est de 0,53. La valeur du rapport de *Student* est inférieure à 1,96 (le rapport de *Student* est de 0,63 en valeur absolue).

Comparativement, les coefficients MA(1) et SMA(1) sont assez significatifs, même si le rapport de *Student* pour le coefficient SMA(1) est inférieur à 1,96.

Puisque l'un des paramètres du modèle ne vérifie pas les conditions de validation, nous allons donc rejeter le modèle $(0,1,1)(1,1,1)_{12}$.

En suivant la même démarche, nous allons procéder à l'élimination du coefficient SAR(1), ne vérifiant pas les critères requis.

A présent, il faudrait prendre en compte le modèle $(0,1,1)(0,1,1)_{12}$ pour une nouvelle estimation.

En éliminant le coefficient SAR(1), nous obtenons les résultats suivants :

Tableau.IV.3- Résultats de l'estimation du modèle $(0,1,1)(0,1,1)_{12}$

	B	Ecart-type B	T-RATIO	PROB.APPROX
MA1	,966637	,144542	6,6875938	,00000001
SMA1	,696885	,260989	2,6701683	,00999519
T°	-24,597111	12,229750	-2,0112521	,04930149
Abonnés	-,001190	,004839	-,2458908	,80669902
Constante	4,928523	9,086391	,5424071	,58976968

– Interprétation de résultats :

On peut tout d'abord noter que les rapports de *Student* des deux paramètres MA(1) et SMA(1) sont respectivement 6,69 et 2,67 ($\geq 1,96$) avec des probabilités très faibles : ces deux coefficients sont significatifs.

Les tests concernant les paramètres introduits étant satisfaisants, nous passons à la dernière étape de vérification : les tests concernant le bruit blanc.

Une approche pratique consiste à examiner la série des résidus pour déterminer si cette dernière ne présente pas certaines tendances ou régularités qui feraient rejeter l'hypothèse de bruit blanc. Comme le montre la sortie SPSS présentée en *annexe 13*.

Nous allons regrouper pour les retards pertinents (multiple de S) les statistiques de Box-Pierce (cf. Chapitre I) dans le tableau IV.4.

Tableau.IV.4- Test de Box-Pierce

K	12	24	36
Q	8,890	22,235	34,929
$\chi_{1-\alpha}$	19,68	35,17	46

Les tests statistiques Q (équations I.30 et I.31 en page 26) sont vérifiés. $Q(h) \leq \chi_{1-\alpha}$, on accepte alors l'hypothèse de test de bruit blanc sur les résidus.

A l'issue de l'étape de vérification, le modèle SARIMA₁₂(0,1,1) (0,1,1) satisfait aux tests classiques. Le modèle est donc retenu pour la phase de prévision.

L'évolution de la charge électrique peut donc être approchée par un processus SARIMA₁₂ (0,1,1) (0,1,1). Les paramètres du modèle sont donnés dans le tableau IV.3.

3. Prévision :

Pour pouvoir tester les performance du modèle élaboré, nous avons décomposé l'ensemble des observations à notre disposition en deux parties. Les soixante premières observations représentent ce que l'on pourrait appeler l'ensemble d'apprentissage. Les observations restantes nous permettront de tester le modèle. Elles représentent un ensemble de validation (les 12 dernières observations).

Les données relatives aux différents facteurs (température moyenne, nombre d'abonnés) et les observations mensuelles des pics de charge sont présentées en *annexe10*.

Pour prévoir les valeurs de la consommation électrique, nous avons besoin des valeurs des différents facteurs d'entrées pour un même horizon de prévision. Ces valeur nous ont été fournies par les différents organismes dont nous nous sommes rapprochés (cf. *Annexe 10*).

Pour prévoir le nombre d'abonnés SONELGAZ, les cadres de cette dernière, responsables de cette tâche, utilisent des méthodes empirique. En effet, en tenant compte des objectifs définis pour chaque mois et en fonction des contraintes pratiques de connexion au réseau, ces cadres gênèrent leur prévision.

La prévision des températures moyennes à l'horizon 2007 n'est pas une tâche facile. Ce type de données (*cf. Annexe 11, figure.11.3*) ne pourrait être prédit par des méthodes classiques. Nous avons utilisé des prévisions saisonnières récoltées auprès de l'Office Nationale de la Météorologie (ONM). Ces prévisions sont la traduction de probabilités d'occurrence par rapport a une moyenne saisonnière.

Afin de générer les résultats de prévision, nous avons utilisé le logiciel SPSS, en premier lieu sur l'ensemble d'apprentissage. Une fois le critère de performance du modèle calculé sur l'ensemble de validation, nous avons généré des prévisions à l'horizon 2007.

Mais tout d'abord, définissons les critères de comparaison que nous allons utiliser dans le cadre de notre étude.

Les critères de comparaison :

Ces différents critères que l'on cherche à minimiser sont basés sur l'erreur de prévision. On utilisera les critères suivants :

- L'erreur absolue moyenne MAE ;
- L'erreur quadratique moyenne MSE ;
- L'écart absolu moyen en pourcentage MAPE ;

Les résultats de la prévision sont donnés en *annexe 10* . Pour visualiser ces résultats nous les avons représentés dans le graphe ci-dessous :

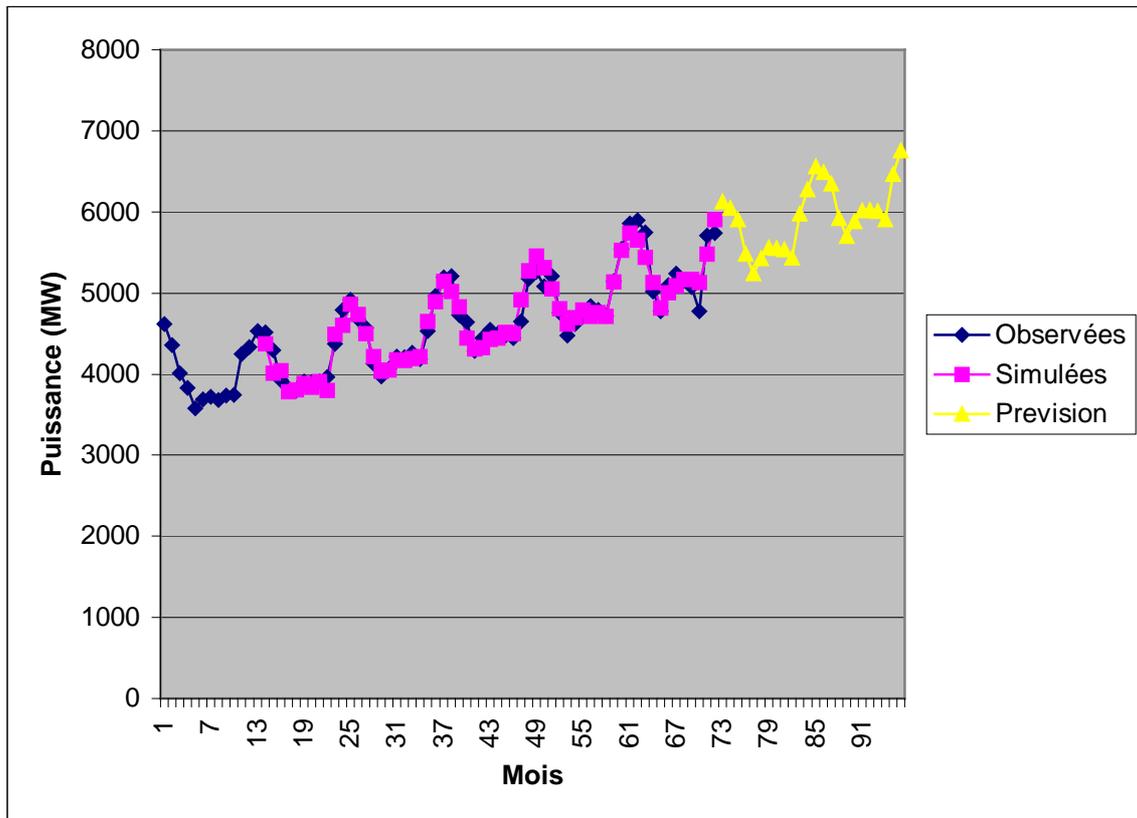


Figure.IV.4 : Valeurs observées et simulées de l'évolution de la charge électrique

Les statistiques de comparaison sont données dans le tableau suivant :

Tableau.IV.5- Récapitulatifs des performances 1

	Apprentissage	Validation	Pour les 72 obs.
MAE	85,28	171,32	102,94
MSE	11976,72	42607,97	18203,31
MAPE	1,86 %	3,18%	2,13%

Nous remarquons que les écarts absolus moyens en pourcentage ne dépassent pas les 10%, ce qui nous conduit à dire que les résultats de prévision sont satisfaisants.[MEL 90]

Après avoir suivi la démarche traditionnelle des modèles statistiques, nous allons à présent passer à la conception du modèle neuronal.

II. Application des réseaux de neurones artificiels :

Dans cette partie du chapitre, nous allons décrire la méthodologie que nous avons adoptée en élaborant un programme sous MATLAB afin d'aboutir au réseau de neurones optimal.

Rappelons que nous avons porté notre choix concernant le type de réseaux que nous allons utiliser sur le perceptron multicouche PMC, avec rétropropagation de l'erreur (*cf. Chapitre III*).

Comme nous l'avons souligné au chapitre II, la conception d'un modèle neuronal nécessite de suivre une certaine procédure. I.KAASTRA [KAA 96], a mis en place une méthode de conception de réseaux de neurones, de PMC à backpropagation tout particulièrement, pour la prévision de séries temporelles. Cette démarche s'articule en plusieurs étapes :

1. traitement des variables ;
2. échantillonnage des variables ;
3. propriétés du réseau ;
4. apprentissage.

Nous allons suivre cette démarche pour la conception du réseau. Cette procédure représente la première phase de notre application.

1. Conception du réseau :

a. Traitements des variables :

Pour cela nous disposons de 72 observations des variables explicatives (variables d'entrées du modèle) et autant de mesures de puissance maximale (variable de sortie du modèle).

Pour chaque mois m , les entrées correspondent aux valeurs, pour ce même mois, des descripteurs retenue dans le chapitre IV ainsi que la valeur de la charge maximale pendant le mois $m-1$. Ceci nous donne donc trois (03) entrées pour le modèle. La sortie du réseau correspond, bien entendu, à la valeur de la charge électrique au mois m .

Le traitement des données consiste à transformer les variables d'entrée et de sortie pour minimiser les bruit et optimiser l'apprentissage du réseau. En effet, ces différentes variables sont rarement de même grandeur, ce qui est le cas des données que nous allons utiliser. Il est donc nécessaire de normaliser et de centrer ces variables ou d'utiliser toute autre transformation qui pourrait uniformiser les grandeurs utilisées.

Nous avons donc choisi de transformer les données à notre disposition de telle sorte que ces dernières appartiennent à l'intervalle $[-1,1]$.

b. Echantillonnage des variables :

La technique la plus courante est de diviser la série temporelles (pour chaque variable) en deux parties. La première partie sera présentée au réseau. Elle sera donc utilisée lors de l'apprentissage. La seconde partie, quant à elle, nous permettra de juger des capacités de généralisation du réseau entraîné. Il n'y a pas de règles dans la façon de décomposer la série initiale, mais généralement les données d'apprentissage représente 80% de l'ensemble de l'historique.

Nous avons constitué deux groupes : les 60 premières observations serviront comme base d'apprentissage, tandis que les 12 dernières seront utilisées pour la validation

c. Propriété du réseau :

Les propriétés du réseau que nous devons définir sont les suivantes :

- nombre de couches cachées ;
- nombre de neurones dans chaque couche cachée ;
- la fonction de transfert pour chaque neurone ;

o *Choix du nombre de couches cachées :*

Pour le choix du nombre de couches cachées, nous nous sommes basés sur le théorème suivant : « *Toute fonction bornée suffisamment régulière peut être approchée uniformément avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire* ». [DRE 04]

cette propriété montre qu'une seule couche cachée est suffisante pour résoudre des problèmes d'approximation de fonctions et donc de prévision.

Nous allons donc utiliser dans notre application un perceptron à une seule couche cachée.

o *Choix du nombre de neurones dans la couche cachée :*

Les mesures de la variable à prédire étant entachées de bruit, il n'est pas souhaitable de chercher le modèle dans une famille de réseaux suffisamment complexes pour que le minimum de la fonction de coût (fonction à minimiser lors de l'apprentissage) soit nul : un tel modèle reproduirait le bruit et les perturbations, alors que le but de la modélisation est d'extraire tout ce qui est déterministe dans l'information présente dans les données.

Pour augmenter la fiabilité d'un modèle réseau de neurones il faut choisir la bonne configuration car de ce choix dépend la qualité de la prévision. En général, il est préférable d'utiliser une seule couche cachée et un nombre de neurones dans cette couche du même ordre de grandeur que le nombre de variables d'entrée. [DAV 93]

Il n'existe réellement aucune règle précise pour déterminer le nombre exact de neurones cachés. Cependant, sélectionner le meilleur nombre demande essentiellement de l'expérimentation.

J.O.Katz [KAA 96] indique qu'un nombre optimal de neurones cachés peut être généralement trouvé entre une à trois fois le nombre de variables d'entrée. O.Ersoy [KAA 96] propose de doubler le nombre de neurones jusqu'à ce que les performances en validation se détériorent.

Mais certaines approches, ayant fixé approximativement le nombre de neurones, privilégient une démarche constructive. Il s'agit là de faire varier le nombre de neurones cachés et de ce fait construire plusieurs réseaux et de comparer les résultats de chaque architecture.[KAA 96]

C'est la démarche que nous avons traduit dans notre programme MATLAB pour résoudre le problème du nombre de neurones dans la couche cachée. Nous avons pour cela procédé un de nombreux tests pour établir un ensemble d'architectures particulières pour lequel les performances paraissent nettement meilleur que des familles de réseaux de complexité supérieure.

- *La fonction de transfert :*

Nous allons utiliser, pour tous les neurones sauf celui de sortie, la fonction tangente hyperbolique, couramment utilisée et qui de plus présente un intervalle d'entrée entre $[-1,1]$.

d. L'apprentissage :

Rappelons que le principe de l'apprentissage est de trouver des valeurs numériques pour les composantes de la matrice des paramètres (des poids) telles que le modèle soit satisfaisant. L'apprentissage du réseau se fait comme nous l'avons dit par l'algorithme de rétropropagation de l'erreur. Pour cela, il faudrait définir avant tout un critère d'évaluation lors de l'apprentissage. Le plus couramment utilisé est l'erreur quadratique moyenne (Mean Squared Error MSE).

L'objectif de l'apprentissage est de minimiser l'erreur quadratique, et atteindre un minimum global. Cela se concrétise par l'apprentissage sur plusieurs itérations effectuées avec des poids générés aléatoirement de valeur minimale. Plusieurs études font mention d'un nombre d'itérations allant de 80 à 5000 itérations.

Le plus important est de déterminer expérimentalement ce nombre en tenant compte de l'évolution de l'erreur d'apprentissage. Nous avons pour cela procédé à plusieurs essais pour déterminer le nombre d'itérations nécessaires et pour lequel nous ne constatons aucune amélioration significative de l'erreur d'apprentissage. Ces essais ont été effectués en prenant en considération des architectures différentes.

Les graphes IV.5, IV.6 et IV.7 illustrent les résultats de simulation pendant l'apprentissage (à des itérations différentes) pour une architecture donnée :

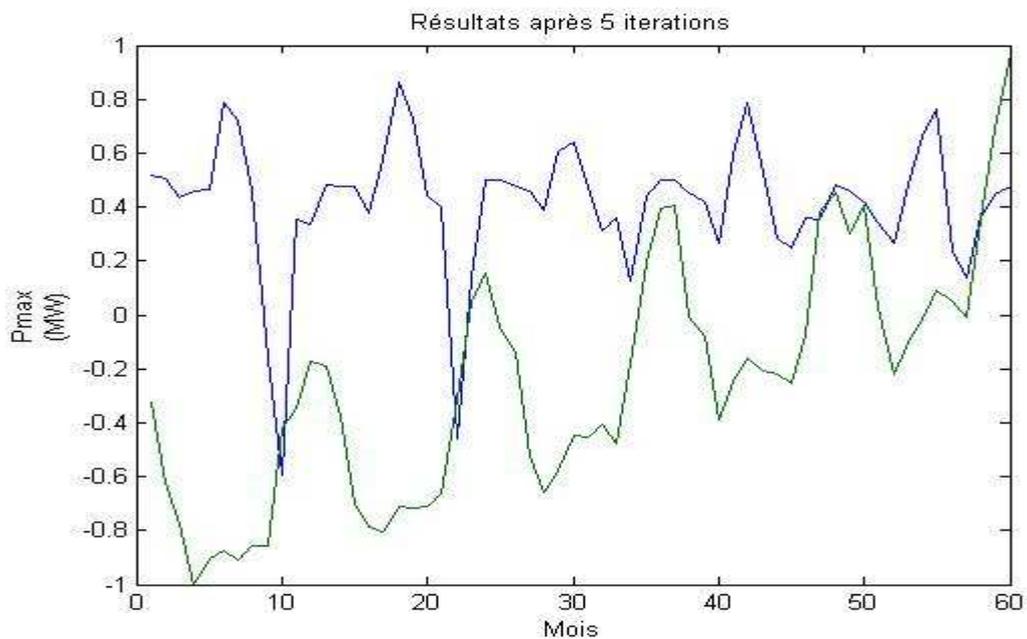


Figure.IV.5- Résultats de simulation après 5 itérations

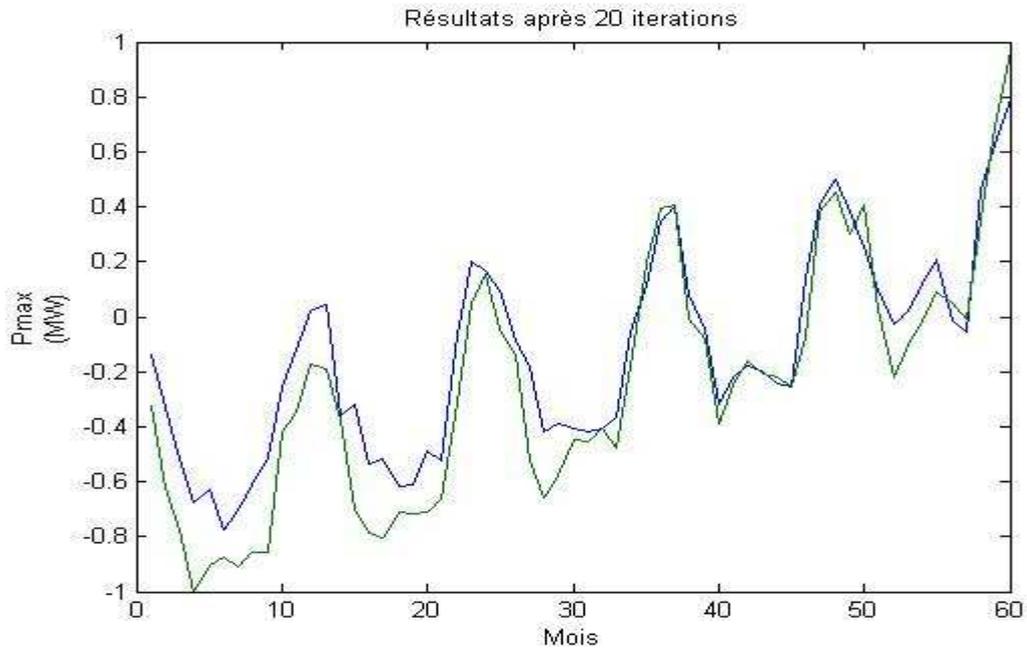


Figure.IV.6- Résultats de simulation après 20 itérations

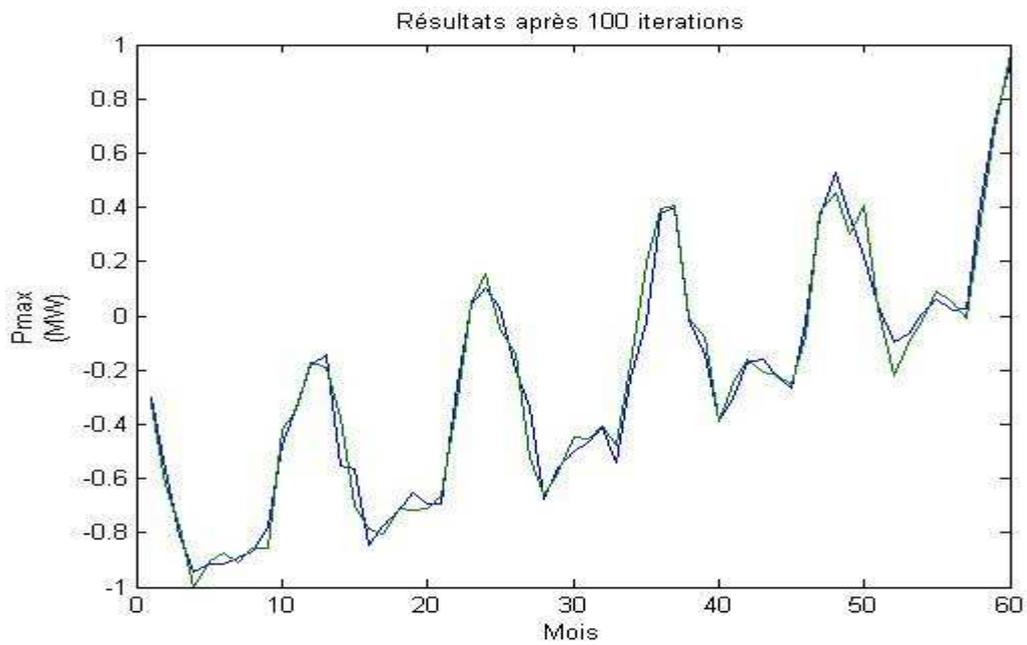


Figure.IV.7- Résultats de simulation après 100 itérations

Nous remarquons, au vue des essais pratique que le réseau reprenait assez vite l'allure de la courbe des observations initiales (objectif en sortie).

Ainsi, nous nous sommes fixés un nombre maximal d'itérations pour la convergence de l'apprentissage égale à 150.

Ces étapes de construction achevées, il nous reste encore à trouver le réseau qui modéliserait au mieux la variable de sortie. En d'autres termes, le réseau qui aurait les meilleures capacités de généralisation. Ce réseau minimiserait l'erreur en validation.

2. La détermination du réseau optimal :

La démarche que nous avons choisie pour déterminer le réseau optimal consiste chercher d'abord l'approximation dans la famille des réseaux à une seule couche cachée avec un nombre de neurones n allant de 1 à 7. Puis de comparer l'erreur de prévision en validation (une estimation de l'erreur de généralisation). La sélection sera alors fondée sur les données qui n'ont pas été utilisées pour l'estimation des paramètres.

Résumons la démarche développée sous MATLAB :

1. En partant de la structure la plus simple ($n=1$) et en rendant la structure progressivement plus complexe ($n=7$), nous calculons l'erreur en validation (erreur quadratique moyenne sur l'ensemble de validation).
2. Pour chaque structure nous effectuons 200 passages, autant d'apprentissages avec une réinitialisation des poids afin de minimiser l'erreur en validation du réseau. De plus pour chaque réseau, on calcule l'erreur quadratique moyenne normalisée NMSE (MSE divisée par la variance de la sortie). Un bon apprentissage doit vérifier $NMSE < 1$.
3. Nous comparons les réseaux optimaux obtenus pour chaque n .
4. réitérer la procédure si les résultats obtenus ne sont pas satisfaisant.

A titre d'exemple, les figures 14.1 et 14.2 en *annexe 14* illustrent le meilleur et le moins bon résultat obtenus lors d'une exécution du programme. Les différents graphiques renvoient à l'évolution de l'erreur durant l'apprentissage, et l'évolution des valeurs observées et simulées pendant l'apprentissage et la validation.

Nous ne retenons qu'un seul réseau optimal par architecture (nombre de neurones cachés) après les 200 passages. On obtient ainsi, à chaque exécution du programme, 7 réseaux optimaux chacun dans leur famille de réseaux. Pour une erreur en validation minimale, l'un des sept réseaux obtenus sera retenu comme étant le meilleur réseau pour cette exécution. Nous avons donc procédé à plusieurs exécutions pour déterminer l'ensemble de réseaux optimaux dans chaque architecture.

Après l'apprentissage des modèles avec des initialisations différentes des poids et le calcul de l'erreur quadratique moyenne MSE et l'erreur quadratique moyenne normalisée, nous avons retenu, après une cinquantaine d'exécutions, les meilleurs résultats :

Tableau.IV.6- Résultats d'exécutions

	Execution 1	Execution 2	Execution 3	Execution 4
MSE	0,0124	0,0106	0,0122	0,0158
NMSE	0,0505	0,0432	0,0497	0,0644
<i>n</i>	3	7	6	3

Le meilleur réseau est celui obtenu à l'exécution 2 avec 7 neurones cachés.

Le graphique ci-dessous illustre les performances de ce dernier :

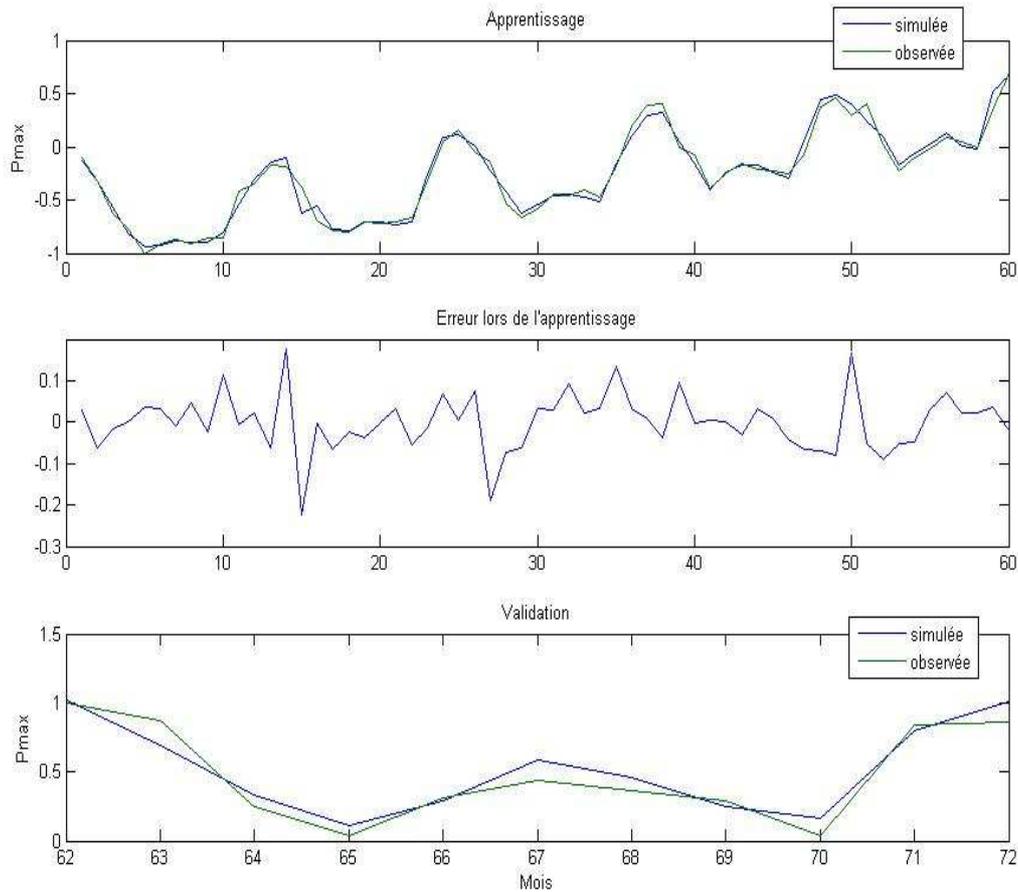


Figure.IV.8- Performances du réseau optimal

Nous notons à partir de la figure IV.8 que le réseau reproduit presque parfaitement la courbe de charge. De plus, il épouse parfaitement la tendance du graphique et prévoit même avec une grande précision les pics de consommation.

3. La prévision:

En utilisant le modèle retenu à l'itération 2, nous passons maintenant à la phase finale qui est la prévision de la charge électrique maximale. Nous présentons au réseau les nouvelles données relatives aux prévisions des descripteurs (cf. Annexe 10). Les valeurs simulées de l'évolution de la charge électrique ainsi que la prévision sont données en annexe 10.

Les graphes associés à la prévision de l'évolution de la consommation électrique sont représentés sur la figure.IV.9 :

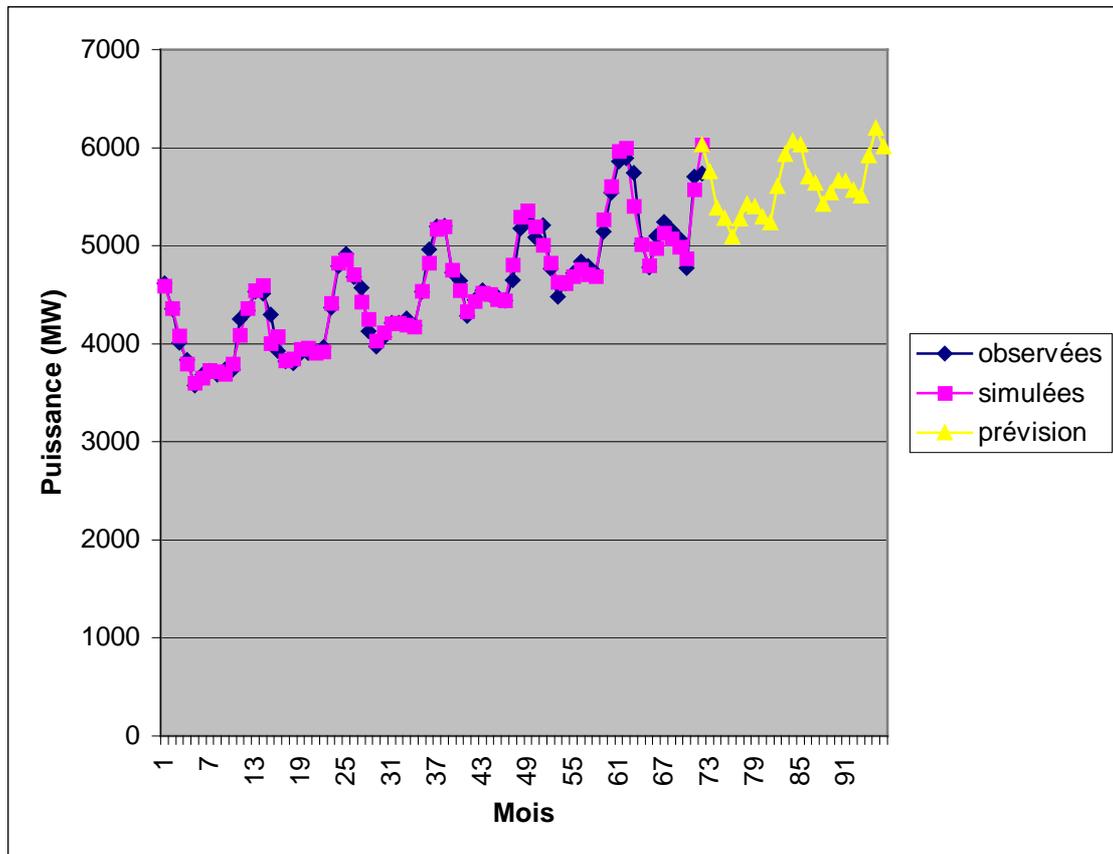


Figure.IV.9 : Valeurs observées et simulées de l'évolution de la charge électrique

Les statistiques de comparaison sont données dans le tableau suivant :

Tableau.IV.7- Récapitulatifs des performances 2

	Apprentissage	Validation	Pour les 72 obs.
MAE	60,34	124,06	70,96
MSE	6953,12	24307,40	9845,50
MAPE	1,35 %	2,26 %	1,50 %

On remarque que le réseau est très performant. La valeur de l'écart absolu moyen en pourcentage est bien inférieure à 10 %.

III. Comparaison des résultats :

Les résultats obtenus par les deux méthodes sont très satisfaisants (<10% en terme de MAPE). Mais en observant le tableau IV.8, nous remarquons nettement que les valeurs des écarts absolus moyens en pourcentage (MAPE) sont plus intéressantes par la modélisation neuronale (2,26 % en validation).

Tableau IV.8- Comparatif des résultats

	MAE	MSE	MAPE
RNA	124,06	24307,40	2,26 %
ARMA	171,32	42607,97	3,18%

Pour mieux visualiser les performances des deux méthodes, nous allons confronter les résultats dans un graphe (Figure.IV.10).

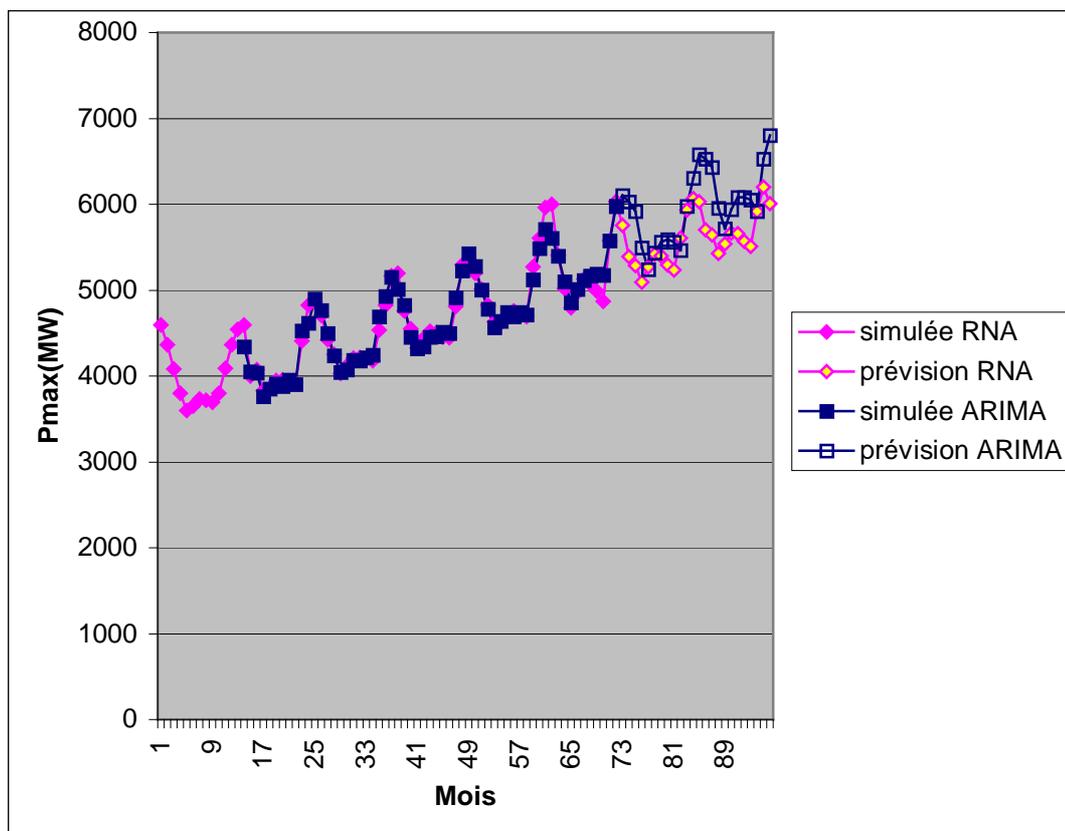


Figure.IV.10-Résultats simulés et prévisions par les deux méthodes

Nous remarquons nettement la différence entre les deux méthodes lors de la validation. Le dernier pic hivernal n'est pas aussi bien repris par le modèle ARMA. Nous notons aussi une tendance assez différente entre les deux modèles.

Cependant, les deux méthodes reprennent parfaitement le caractère saisonnier de la série initiale, et dénotent même un changement de consommation durant les mois d'été, un comportement qui tend à se généraliser.

Confrontation des méthodes :

Après avoir comparé les résultats obtenus par les deux méthodes, faisons une confrontation entre ces deux dernières et ce, selon certains critères : (complexité, traitement des données, temps de calcul, la taille de l'historique...)

- En terme de complexité de construction, les performances des deux méthodes sont pratiquement identiques. En effet, aussi bien la méthode de Box et Jenkins que celle des réseaux de neurones utilisent des logiciels assez performants intégrant des boîtes à outils facilitant la conception des différents modèles. Le logiciel SPSS permet, à travers son interface, de traiter aisément des séries temporelles complexes et d'appliquer la méthodologie de Box et Jenkins en toute simplicité. Aussi, mise à part l'aspect programmation sous MATLAB, la conception du modèle neuronal est facilitée grâce aux sous programmes intégrées dans la boîte à outils que fournit ce logiciel.
- Il faut signaler que les données d'entrée d'un réseau doivent fréquemment être retraitées afin de les normaliser, et de les stationnariser sous peine d'obtenir des résultats décevants. La normalisation vise à ramener la plage d'évolution des valeurs prises par les variables à l'intérieur d'un intervalle standardisé, fixé a priori.
- En ce qui concerne le temps de calcul, la méthode de Box et Jenkins donne des résultats dans des délais plus brefs que ceux obtenus par la méthode des réseaux de neurones. Ceci peut s'expliquer par le fait qu'il n'existe pas encore de théorie

permettant de déterminer l'architecture optimale d'un réseau de neurones. En particulier la détermination du nombre de couches cachées et de neurones dans chaque couche relève encore de l'intuition de l'utilisateur et de sa capacité d'expérimenter plusieurs architectures afin de retenir les meilleurs résultats.

- S'agissant de la taille de l'historique, les modèles ARMA se basent sur un échantillon de données beaucoup plus important. En effet, l'avantage pratique des réseaux de neurones, de part leur capacité de généralisation, réside dans leur capacité à réaliser des modèles de précision équivalentes aux modèles classiques. Les réseaux de neurones permettent de tirer le meilleur parti des données numériques disponibles pour construire des modèles.
- Les modèles ARMA requièrent des connaissances statistiques pour l'interprétation. En effet, l'obtention du modèle est conditionné par un traitement des résultats, opération se traduisant par l'application de plusieurs tests. Cette opération nécessite un certain temps considérable car il faut tester tous les paramètres. Pour ce qui est des réseaux de neurones artificiels, certes le temps d'obtention des résultats est important, mais ces derniers sont directement exploitables.
- Il faut noter que la modélisation par les réseaux de neurones permet d'exprimer d'éventuelles relations non linéaires entre les données ce qui n'est pas le cas des modèles ARMA.
- Les réseaux de neurones restent encore en partie aujourd'hui une « boîte noire » dans laquelle il reste difficile d'extraire des relations pertinentes entre les variables. De ce point de vue, le système neuronal n'a qu'un pouvoir explicatif médiocre contrairement aux modèles ARMA qui sont capables de retracer le cheminement suivi pour atteindre le résultats.

CONCLUSION GÉNÉRALE

La prévision de la demande énergétique est une activité cruciale pour les différentes entreprises qui sont, de près ou de loin concernées par les différentes activités de production, de transport ou de distribution d'énergie. Cruciale car cette prévision énergétique en générale, électrique en particulier, impose aux prévisionnistes des tolérances très réduites. Ceci est principalement dû au fait qu'il faut maintenir un équilibre constant entre la production et la consommation de l'énergie électrique. Cependant, nous assistons aujourd'hui à un accroissement significatif de la demande en matière d'électricité et cela rend les prévisions de plus en plus complexes. Les prévisionnistes se doivent à présent de prendre en considération de nouveaux paramètres.

C'est justement dans cette optique qu'a été orientée notre problématique. Nous nous sommes donc assignés, tout au long de notre travail, comme objectif de concevoir un modèle de prévision mieux adapté à la complexité du phénomène. Ceci a été effectué à travers un outil puissant de traitement des données : Les réseaux de neurones artificiels. Nous avons par la suite comparé ces résultats avec ceux obtenus par le biais d'un modèle de prévision plus conventionnel : la méthode de Box & Jenkins.

Ainsi, au chapitre I, a été détaillée la méthode de Box & Jenkins, jugée comme étant la méthode classique la plus appropriée à la série chronologique que nous devons traiter. Dans ce chapitre, nous avons abordés les différents concepts et définitions appropriés pour cette démarche. C'est dans cette optique que nous y avons détaillé les modèles ARMA associés à la méthodologie de Box & Jenkins.

Dans le second chapitre, nous nous sommes intéressés aux réseaux de neurones, à travers les principes de base qui leur sont associés pour mieux nous familiariser avec ce domaine. C'est en effet, par cette approche d'intelligence artificielle, que nous avons conçu notre second modèle de prévisions.

Toute modélisation à des fins de prévision passe par une analyse des différents facteurs influençant la consommation électrique. Pour cela, un prétraitement de ces variables explicatives a été effectué lors du chapitre III.

Arrivés à ce stade, la conception des modèles a pu être élaborée. Pour chaque méthode, nous avons suivi les démarches nécessaires à des fins de prévision. Les logiciels SPSS et MATLAB ont été utilisés pour la construction des modèles par les méthodes de Box & Jenkins ainsi que les réseaux de neurones respectivement. Ces logiciels ont l'avantage d'offrir des boîtes à outils et des sous programmes dédiés à ces différentes méthodes pour faciliter la construction du modèle.

Nous avons abouti à des prévisions pour chaque méthode, résultats que nous avons confrontés par la suite. Nous sommes donc arrivés à la conclusion que, pour la chronique que nous avons étudiée, la modélisation par la méthode des réseaux de neurones artificiels offre des résultats plus satisfaisants que ceux obtenus avec la méthode classique de Box et Jenkins.

Cette modélisation neuronale de la charge électrique, au vu des résultats très encourageants, constitue un outil d'aide à la décision très intéressant. En effet, ce modèle fournit aux responsables de la stratégie de SONELGAZ un support leur permettant de mieux définir leur objectifs et stratégies de développement, ainsi que de mieux adapter les programmes de production.

De plus, ce modèle pourra être mis à la disposition des managers pour l'élaboration de prévision plus fiables. En effet, notre modèle n'intègre sans doute pas l'ensemble des facteurs influençant la demande en énergie électrique. Cependant, la démarche adoptée permet de prendre en considération de nouvelles variables lors de la modélisation.

En ce qui nous concerne, cette étude nous a permis d'enrichir nos connaissances tant sur le plan théorique que sur la réalité du domaine de la prévision.

Les réseaux de neurones représentent, en effet, une alternative très prometteuse pour l'analyse des séries chronologiques. De plus, au vu des structures de réseaux et des procédures d'apprentissage utilisées lors de notre étude, nous pouvons penser que ce domaine présente un fort potentiel d'exploration. Ces outils puissants n'ont pas atteint leur plein développement.

Annexe 1 : Processus stationnaire du second ordre [FIO 99]

- **Définition :**

Un processus $(X_t, t \in \mathbb{Z})$ est dit stationnaire au second ordre, ou stationnaire au sens faible, ou stationnaire d'ordre deux si les trois conditions suivantes sont satisfaites :

- $t \in \mathbb{Z}, E [X_t^2] < \infty.$
- $t \in \mathbb{Z}, E [X_t] = m,$ indépendant de $t.$
- $(t, h) \in \mathbb{Z}^2, cov [X_t, X_{t+h}] = E [(X_{t+h} - m) (X_t - m)] = \gamma (h),$ indépendant de $t.$

La première condition $E [X_t^2] < \infty$ garantit tout simplement l'existence (ou la convergence) des moments d'ordre deux. La seconde condition $E [X_t] = m$ porte sur les moments d'ordre un et signifie tout simplement que les variables aléatoires X_t doivent avoir la même espérance quelle que soit la date t . Autrement dit, l'espérance du processus X_t doit être indépendante du temps. Enfin, la troisième condition, $\gamma(h)$ indépendant de t , porte sur les moments d'ordre deux résumés par la fonction d'autocovariance. Cette condition implique que ces moments doivent être indépendants de la date considérée et ne doivent dépendre que de l'ordre des retards uniquement. Autrement dit, la fonction d'autocovariance du processus X_t doit être indépendante du temps.

Sous ces trois conditions, il existe alors une ou plusieurs représentations possibles de la série temporelle X_t , ces représentations ou processus ayant la même distribution dans le temps (même moments). Cette hypothèse d'invariance de la distribution dans le temps permet ainsi de se limiter à une certaine classe de processus.

En résumé, un processus est stationnaire au second ordre si l'ensemble de ses moments sont indépendants du temps.

Annexe 2 : Processus à accroissements indépendants [FIO 99]

- **Définition d'un Processus à accroissements indépendants**

Le processus $(X_t; t \in \mathbb{Z})$ est un processus stationnaire à accroissements indépendants si la loi de probabilité des accroissements (X_{t+h}, X_t) est indépendante de t , quelque soit $h \in \mathbb{Z}$. Mais dans la pratique, on retient plus généralement la définition suivante du bruit blanc.

$$E(X_t) = 0$$
$$\gamma(h) = E[X_t X_{t-h}] = \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases}$$

La première condition signifie tout simplement que l'espérance du processus est indépendante du temps, et de plus qu'elle est nulle. La seconde condition implique bien entendu l'indépendance de la fonction d'autocovariance par rapport au temps (stationnarité). Mais elle implique en outre que les termes d'autocovariances (pour $h \neq 0$) sont tous nuls. Seule la variance est non nulle.

Autrement dit, cela signifie que les bruits blancs sont des processus stationnaires particuliers sans "mémoire". Le niveau de la série considéré aujourd'hui n'a aucune incidence sur son niveau de demain, tout comme le niveau d'hier n'a aucune incidence sur le niveau d'aujourd'hui.

$$\varepsilon_t \text{ i.i.d. } \sim N(0, \sigma^2)$$

Annexe 3 : Les opérateurs retard et différence [BOU 98]

- **L'opérateur Retard :**

On aura souvent à considérer une variable en fonction de son passé. Il est donc commode de définir un opérateur qui transforme une variable X_t en sa valeur passée. C'est l'opérateur retard désigné par la lettre L et tel que:

$$L(X_t) = X_{t-1} \quad \text{et} \quad L^k(X_t) = X_{t-k}$$

Cet opérateur sera utilisé à l'intérieur de polynômes notés par exemple :

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \dots + \beta_s L^s \quad (3.1)$$

Alors :

$$B(L) = \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_s X_{t-s} \quad (3.2)$$

- **L'opérateur différence**

C'est l'opérateur Δ . il s'écrit en fonction de l'opérateur retard L .

$$\left\{ \begin{array}{l} \Delta = 1 - L \\ \Delta X_t = (1-L) X_t = X_t - X_{t-1} \\ \Delta^d X_t = (1-L)^d X_t = X_t - X_{t-d} \end{array} \right. \quad (3.3)$$

Annexe 4 : Choix de p et q dans le cas d'un processus ARMA par « la méthode du coin » [BOU 98]

Une méthode d'identification simultanée de p et q peut être déduite de la caractérisation des ordres définie précédemment. On va la définir à partir des autocorrélations et non des autocovariances. Il faut se rappeler que p et q sont tel que :

$$\begin{aligned}\hat{\Delta}(i, j) &= 0, \forall i \geq p \\ \hat{\Delta}(i, q-1) &\neq 0, \forall i \geq p-1 \\ \hat{\Delta}(p-1, j) &\neq 0, \forall j \geq q-1\end{aligned}$$

Avec :

$$\hat{\Delta}(i, j) = \det \begin{bmatrix} \hat{\rho}(j+1) & \cdots & \hat{\rho}(j+i+1) \\ \vdots & & \vdots \\ \hat{\rho}(j+1-i) & \cdots & \hat{\rho}(j+1) \end{bmatrix}$$

Les valeurs de $\Delta(i, j)$ sont inconnues, mais on peut les estimer par les $\hat{\Delta}(i, j)$; obtenus en remplaçant dans les déterminant les corrélations $\rho(h)$ par leurs estimateurs $\hat{\rho}(h)$.

Comme $\hat{\Delta}(i, j)$ sont des estimateurs convergeant de $\Delta(i, j)$. Il est donc vraisemblable qu'on verra apparaître dans un tableau, comportant les éléments $\hat{\Delta}(i, j)$, une rupture entre les lignes $i = p-1$ et $i = p$ et entre les colonnes $j = q-1$ et $j = q$.

Comme les estimateurs $\hat{\Delta}(i, j)$ de notre tableau présente des variabilité différentes, il est donc préférable de chercher les ruptures sur le tableau ayant pour élément des rapports de « Student » défini par :

$$\frac{|\hat{\Delta}(i, j)|}{\sqrt{\hat{V}(\hat{\Delta}(i, j))}}$$

En comparant les rapports du tableau à 1.96, on obtient un test asymptotique de nullité au seuil de 5%.

On, peut améliorer ce test de choix de p et q en testant la nullité simultanée de plusieurs déterminants, par exemple : $\hat{\Delta}(i, j)$, $\hat{\Delta}(i+1, j)$, $\hat{\Delta}(i, j+1)$. Ce test serait fondé sur une forme quadratique en $\hat{\Delta}(i, j)$, $\hat{\Delta}(i+1, j)$, $\hat{\Delta}(i, j+1)$ suivant asymptotiquement sous l'hypothèse nulle une loi du khi deux à 3 degrés de liberté.

Ce qui fait qu'on va donc comparer les éléments du tableau des corrélations canonique à 7.81 et on obtiendra un test asymptotique de nullité au seuil de 5%.

Annexe 5 : Estimation du maximum de vraisemblance [CHA 03]

Nous allons présenter ici brièvement la démarche de l'estimation par le maximum de vraisemblance. Cette maximisation est réalisée à l'aide d'algorithmes d'optimisation non linéaire (Newton-Raphson, méthode du simplex). Nous nous contenterons ici de montrer comment s'écrit le programme de maximisation de la vraisemblance permettant d'estimer les paramètres d'un modèle ARMA (p, q).

On considère un processus stationnaire X_t satisfaisant une représentation ARMA (p, q) telle que :

$$\begin{cases} \Phi(L) X_t = \Theta(L) \varepsilon_t \\ \Theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q \\ \Phi(L) = 1 + \varphi_1 L + \dots + \varphi_p L^p \end{cases}$$

Les polynômes Φ et Θ ont toutes leurs racines, de module >1 et n'ont pas de racines communes.

En plus de la définition standard d'un processus ARMA, on fait l'hypothèse de la normalité des résidus afin de spécifier une forme fonctionnelle à la vraisemblance du modèle.

$$\varepsilon_t \text{ i.i.d } \sim N(0, \sigma^2)$$

Ecrivons alors la vraisemblance associée au vecteur de réalisation (X_1, X_2, \dots, X_T)

$$\left(\frac{2\pi\sigma^2}{\varepsilon} \right)^{-T/2} \det [(\theta_i, \varphi_i)]^{-1/2} \exp \left\{ \frac{-1}{2\sigma^2} x' [(\theta_i, \varphi_i)]^{-1/2} x \right\}$$

- Equations de Yule-Walker

– Estimation des φ_i : On utilise les équations de Yule-Walker

$$\rho(k) = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p}$$

$$\begin{bmatrix} \rho_{q+1} \\ \rho_{q+2} \\ \vdots \\ \rho_{q+p} \end{bmatrix} = \begin{bmatrix} \rho_q & \dots & \rho_{q-p+1} \\ \rho_{q+1} & \dots & \rho_{q-p+2} \\ \vdots & & \\ \rho_{q-p+1} & \dots & \rho_q \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{bmatrix}$$

L'estimateur sera donc :

$$\hat{\varphi}_k = \hat{R}_k^{-1} \hat{\rho}(k) \quad k=1, \dots, p$$

– Estimation des θ_i

$$\begin{bmatrix} \rho'_{p+1} \\ \rho'_{p+2} \\ \vdots \\ \rho'_{p+q} \end{bmatrix} = \begin{bmatrix} \rho'_p \dots \rho'_{p-q+1} \\ \rho'_{p+1} \dots \rho'_{p-q+2} \\ \vdots \\ \rho'_{p+q-1} \dots \rho'_q \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_q \end{bmatrix}$$

$$\hat{\rho}'(k) = \frac{\sum_{h=0}^{p-k} \hat{\pi}_h \hat{\pi}_{h+k}}{\sum_{h=0}^p \hat{\pi}_h^2}$$

$$\hat{\theta}_k = \hat{R}_k^{-1} \hat{\rho}'(k) \quad k = 1, \dots, q$$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^2$$

Cette méthode Convergence rapidement vers la solution

Annexe 6 : généralités sur les réseaux de neurones artificiels

1. Historique : [DAV 93]

Dans la période de 1940 à 1956, les sciences et techniques de la cognition s'élaborèrent peu à peu et de l'ancienne cybernétique naquirent les sciences cognitives. La volonté de fonder une science « naturelle » de l'esprit émergea quasiment au même moment à cette époque, en Europe et aux Etats-Unis. Par les efforts conjugués de chercheurs d'origine très diverses (neurosciences, intelligence artificielle, linguistique, psychologie et épistémologie principalement), un nouveau champ d'investigations pluridisciplinaires, que sont les sciences cognitives, vit donc le jour. L'apparition des réseaux de neurones artificiels s'inscrit complètement dans cette genèse des sciences cognitives, et leur origine tient dans la volonté de modéliser, de façon mathématique, les neurones biologiques.

Les premiers à proposer un modèle, sont deux bio-physiciens de Chicago, McCulloch et Pitts, qui, en 1943 inventent le premier neurone formel qui portera leur noms (neurone de McCulloch-Pitts ou automate à seuil).

Quelques années plu tard, en 1949, Hebb, dans une perspective psychophysiologique, souligna l'importance du couplage synaptique dans le processus d'apprentissage.

C'est en 1958 que Ronsenblatt décrivit le premier modèle opérationnel de réseaux de neurones, mettant en œuvre les idées de Hebb, McCulloch et Pitts : le perceptron, inspiré du système visuel, capable d'apprendre à identifier des formes simples et à calculer certaines fonctions logiques en modifiant ses connexions synaptiques. Il constitue donc le premier système artificiel présentant une faculté jusque là réservée aux êtres vivants : la capacité d'apprendre par l'expérience.

Ce modèle suscita beaucoup de recherches, et sans doute trop d'espoir. Quand deux mathématiciens, Minsky et Papert, démontrèrent en 1969 les limites théoriques du perceptron, l'effet fut presque immédiat : chercheurs et investisseurs se désintéressèrent des réseaux de neurones pour se tourner vers l'approche symbolique de l'intelligence artificielle, qui semblait beaucoup plus prometteuse.

Le renouveau actuel des réseaux de neurones est dû à des contributions originales, comme celle de Hopfield en 1982 qui, montrant l'analogie des réseaux de neurones avec certains systèmes physiques, a permis de leur appliquer un formalisme riche et bien maîtrisé.

2. Fondements biologiques : [DAV 93]

• Le neurone :

Les cellules nerveuses sont les éléments de base du system nerveux central. Celui ci en posséderait environ cent milliards. Les neurones possèdent de nombreux points communs avec les autres cellules du corps humain mais elles présentent cependant des caractéristiques qui leur sont propres :

- Recevoir des signaux des neurones voisins ;
- Intégrer ces signaux et les interpréter ;
- Engendrer un influx nerveux ;
- Le conduire et ;
- Le transmettre à un autre neurone capable de le recevoir.

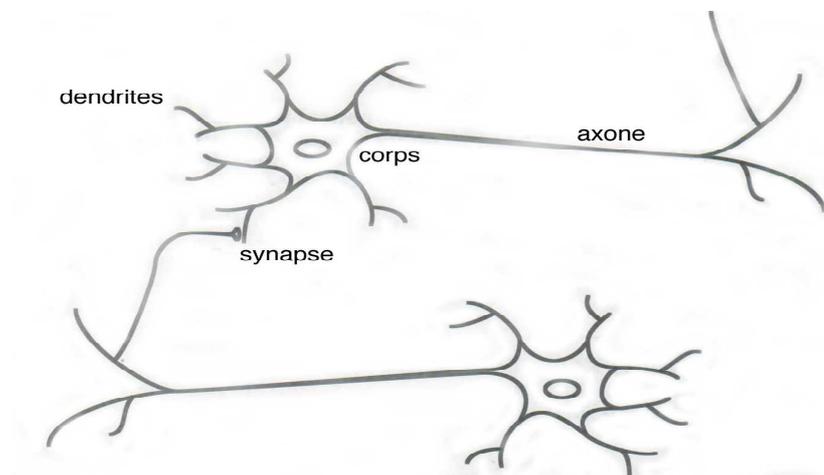


Figure6.1- Représentation d'un neurone [DAV 93]

- **Structure des neurones :**

Un neurone est constitué de trois parties :

- le corps cellulaire,
- les dendrites et
- l'axone.

Le corps cellulaire :

Il contient le noyau du neurone et effectue les transformations biochimiques nécessaires à la synthèse des enzymes et des autres molécules qui assurent la vie du neurone. Sa forme est pyramidale ou sphérique dans la plupart des cas. Ce corps cellulaire fait quelques microns de diamètre

Les dendrites :

Chaque neurone possède une « chevelure » de dendrites. Celles ci sont de fines extensions tubulaires, de quelques dixièmes de microns de diamètre et d'une longueur de quelques dizaines de microns. Elles se ramifient, ce qui les amène à former une espèce d'arborescence autour du corps cellulaire. Elles sont les récepteurs principaux du neurone pour capter les signaux qui lui parviennent (connexions entrantes).

L'axone :

L'axone, qui est à proprement parler la fibre nerveuse, sert de moyen de transport pour les signaux émis par le neurone (connexion sortante). Il se distingue des dendrites par sa forme et par les propriétés de sa membrane externe. En effet, il est généralement plus long (sa longueur varie d'un millimètre à plus d'un mètre) que les dendrites, et se ramifie à son extrémité, là où il communique avec d'autres neurones, alors que les ramifications des dendrites se produisent plutôt près du corps cellulaire.

Pour former le système nerveux, les neurones sont connectés les uns aux autres suivant des répartitions spatiales complexes. Les connexions entre deux neurones se font en des endroits appelés synapses où ils sont séparés par un petit espace synaptique de l'ordre d'un centième de microns.

- **Mode de fonctionnement :**

- le neurone capte les signaux qui lui sont transmis grâce à sa chevelure de dendrites ;
- le signal est traité au niveau du corps cellulaire ;
- le neurone réagit au signal reçu en déclenchant lui-même un signal qui lui est spécifique et qu'il transmet grâce à son axone aux autres neurones auxquels il est connecté.
- La connexion se fait par les synapses qui peuvent se trouver entre deux axones ou entre un axone et des dendrites.
- La transmission du signal se fait par la libération de substances chimiques dans l'espace synaptique, ces substances sont appelées neuro-transmetteurs.

Le corps cellulaire fait la somme de toutes les entrées, si la somme dépasse un certain seuil, le neurone réagit, il agit donc comme une fonction seuil. C'est cette particularité qui va permettre la modélisation du neurone naturel en un neurone formel.

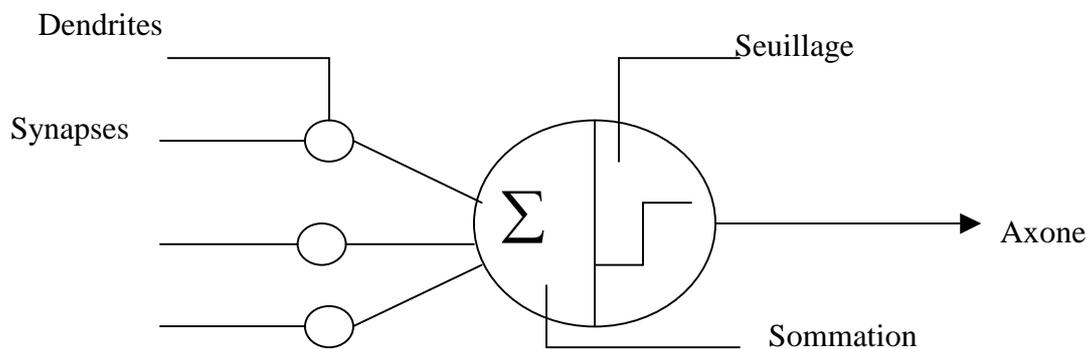


Figure6.2- Schématisation du neurone [DAV 93]

Annexe 7 : Fonctions de transfert [PAR 04]

Différentes fonctions de transfert pouvant être utilisées comme fonction d'activation du neurone. Les trois plus utilisées sont les fonctions «seuil» (en anglais «hard limit»), «linéaire» et «sigmoïde».

- Comme son nom l'indique, la fonction seuil applique un seuil sur son entrée. Plus précisément, une entrée négative ne passe pas le seuil, la fonction retourne alors la valeur 0 (on peut interpréter ce 0 comme signifiant *faux*), alors qu'une entrée positive ou nulle dépasse le seuil, et la fonction retourne 1 (vrai). Elle détermine l'emplacement du seuil sur l'axe $w^T p$, où la fonction passe de 0 à 1. Cette fonction permet de prendre des décisions binaires. Elle est illustrée par la figure.6.a.
- La fonction linéaire est très simple, elle affecte directement son entrée à sa sortie. Appliquée dans le contexte d'un neurone, cette fonction est illustrée à la figure.6.b. Dans ce cas, la sortie du neurone correspond à son niveau d'activation dont le passage à zéro se produit lorsque $w^T p = b$.
- La fonction de transfert sigmoïde est quant à elle illustrée à la figure.6.c. Son équation est donnée par :

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

Elle ressemble soit à la fonction seuil, soit à la fonction linéaire, selon que l'on est loin ou près de b , respectivement.. La sigmoïde est un compromis intéressant entre les deux précédentes. Notons finalement, que la fonction «tangente hyperbolique» est une version symétrique de la sigmoïde.

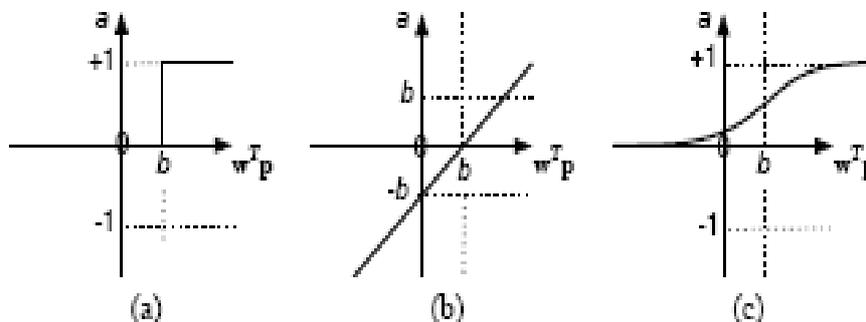


Figure.7.6. – Fonction de transfert : (a) du neurone «seuil» ; (b) du neurone «linéaire», et (c) du neurone «sigmoïde». [PAR 04]

Annexe 8 : Les différentes règles d'apprentissage

- **La règle de Hebb :**

La règle de Hebb (neurophysiologiste) est le premier mécanisme d'évolution proposé pour les synapses.

Son interprétation pour les réseaux de neurones formels est la suivante :

On considère que si deux neurones connectés entre eux sont activés au même moment, la connexion qui les relie doit être renforcée. Dans les cas contraires, elle n'est pas modifiée.[TOU 92]

Ceci traduit bien le fait que lorsque la connexion entre deux cellules est très forte. Quand la cellule émettrice s'active, la cellule réceptrice s'active aussi. Il faut donc, pour lui permettre de jouer ce rôle déterminant lors du mécanisme d'apprentissage, augmenter le poids de cette connexion.

Notons $W_{ij}(t)$ le poids de la connexion entre le neurone j et le neurone i à l'instant t .

Supposons qu'entre les instants t et $t+\delta t$, on ait soumis le neurone à un stimulus, et que ce stimulus ait provoqué l'activation A_i du neurone i , et l'activation A_j du neurone j . La règle de Hebb s'écrit :

$$W_{ij}(t+\delta t) = W_{ij}(t) + \mu A_i A_j \quad (8.1)$$

où μ ($\mu > 0$) est un paramètre de l'intensité de l'apprentissage.[DAV 93]

- **La règle de correction de l'erreur :** [PAR 04]

Cette règle est fondée sur la correction de l'erreur observée en sortie. Soit $a_i(t)$ la sortie que l'on obtient pour le neurone i au temps t . Cette sortie résulte d'un stimulus $p(t)$ que l'on applique aux entrées du réseau dont un des neurones correspond au neurone i . Soit $d_i(t)$ la sortie que l'on désire obtenir pour ce même neurone i au temps t . Alors, $a_i(t)$ et $d_i(t)$ seront généralement différents et il est naturel de calculer l'erreur $e_i(t)$ entre ce qu'on obtient et ce qu'on voudrait obtenir :

$$e_i(t) = d_i(t) - a_i(t) \quad (8.2)$$

et de chercher un moyen de réduire autant que possible cette erreur. Sous forme vectorielle, on obtient :

$$e(t) = d(t) - a(t) \quad (8.3)$$

avec $e(t) = [e_1(t); e_2(t); \dots; e_i(t); \dots; e_s(t)]$ qui désigne le vecteur des erreurs observées sur les S neurones de sortie du réseau. L'apprentissage par correction des erreurs consiste à minimiser un indice de performance F basé sur les signaux d'erreur $e_i(t)$, dans le but de faire converger les sorties du réseau avec ce qu'on voudrait qu'elles soient. Un critère très populaire est la somme des erreurs quadratiques :

$$F(e(t)) = \sum e_i^2(t) = e(t)^T e(t) \quad (8.4)$$

Maintenant, il importe de remarquer que les paramètres libres d'un réseau sont ses poids. Prenons l'ensemble de ces poids et assemblons les sous la forme d'un vecteur $w(t)$ au temps t .

Pour minimiser $F(e(t)) = F(w(t)) = F(t)$, nous allons commencer par choisir des poids initiaux ($t = 0$) au hasard, puis nous allons modifier ces poids de la manière suivante :

$$w(t + 1) = w(t) + \eta x(t) \quad (8.5)$$

où le vecteur $x(t)$ désigne la direction dans laquelle nous allons chercher le minimum et η est une constante positive déterminant l'amplitude du pas dans cette direction (la vitesse d'apprentissage).

L'objectif est de faire en sorte que $F(t + 1) < F(t)$. Pour choisir la direction x tout en respectant la condition précédente, considérons la série de Taylor de 1er ordre autour de $w(t)$:

$$F(t + 1) = F(t) + \nabla F(t)^T \Delta w(t) \quad (8.6)$$

où $\nabla F(t)$ désigne le gradient de F par rapport à ses paramètres libres (les poids w) au temps t , et $\Delta w(t) = w(t + 1) - w(t)$. Or, pour que $F(t + 1) < F(t)$, il faut que la condition suivante soit respectée :

$$\nabla F(t)^T \Delta w(t) = \eta \nabla F(t)^T x(t) < 0 \quad (8.7)$$

N'importe quel vecteur $x(t)$ qui respecte l'inégalité de l'équation 4.9 pointe donc dans une direction qui diminue F . On parle alors d'une direction de «descente». Pour obtenir une descente maximum, étant donnée $\eta > 0$, il faut que le vecteur $x(t)$ pointe dans le sens opposé au gradient car c'est dans ce cas que le produit scalaire sera minimum :

$$x(t) = -\nabla F(t) \quad (8.8)$$

Ce qui engendre la règle dite de «descente du gradient» :

$$\Delta w(t) = -\eta \nabla F(t) \quad (8.9)$$

L'expression exacte du gradient dépend de la fonction d'activation utilisée pour les neurones.

- **L'apprentissage de Widrow-Hoff** : [FRA 92]

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

$$w(t+1) = w(t) + \alpha \delta(t) x(t) \quad (8.10)$$

où:

- $w(t)$ est le poids à l'instant t ,
- $w(t+1)$ le poids à l'instant $t+1$,
- α est le facteur d'apprentissage,
- $\delta(t)$ caractérise la différence entre la sortie attendue et la sortie effective du neurone à l'instant t .
- $x(t)$ la valeur de l'entrée avec laquelle le poids w est associé à l'instant t .

Ainsi, si $\delta(t)$ et $x(t)$ sont positifs tous les deux, alors le poids doit être augmenté. La grandeur du changement dépend avant tout de la grandeur de $\delta(t)$ mais aussi de celle de $x(t)$. Le coefficient α sert à diminuer les changements pour éviter qu'ils deviennent trop grands, ce qui peut entraîner des oscillations du poids.

- **La règle d'apprentissage compétitif** : [PAR 04]

L'apprentissage compétitif, comme son nom l'indique, consiste à mettre en compétition les neurones d'un réseau pour déterminer lequel sera actif à un instant donné. Contrairement aux autres types d'apprentissage où, généralement, tous les neurones peuvent apprendre simultanément et de la même manière, l'apprentissage compétitif produit un «vainqueur» ainsi que, parfois, un ensemble de neurones «voisins» du vainqueur, et seuls ce vainqueur, et potentiellement son voisinage, bénéficient d'une adaptation de leur poids. On dit alors que l'apprentissage est local car limité à un sous-ensemble des neurones du réseau.

Une règle d'apprentissage compétitif comporte les éléments suivants :

- Un ensemble de neurones identiques (même type) sauf pour les valeurs de leurs poids synaptiques ;
- Une limite imposée à la «force» d'un neurone ;
- Un mécanisme permettant aux neurones de compétitionner pour le droit de répondre à un certain sous-ensemble des stimuli d'entrée, de manière à ce qu'un seul neurone de sortie soit actif à la fois.

Ainsi, les neurones individuels peuvent apprendre à se spécialiser sur des sous-ensembles de stimuli similaires pour devenir des détecteurs de caractéristiques.

Dans leur forme la plus simple, les réseaux de neurones qui utilisent l'apprentissage compétitif sont souvent constitués d'une seule couche de neurones de sortie, totalement connectée sur les entrées. Un neurone vainqueur modifiera ses poids synaptiques en les rapprochant (géométriquement) d'un stimulus d'entrée p pour lequel il a battu tous les autres neurones lors de la compétition :

$$\Delta w = \begin{cases} \eta(p - w) & \text{si le neurone est vainqueur} \\ 0 & \text{autrement} \end{cases} \quad (8.11)$$

où $0 < \eta < 1$ correspond à un taux d'apprentissage.

Un neurone qui ne gagne pas la compétition ne modifiera aucunement ses poids. Il ne sera donc pas affecté par le stimulus en question. Parfois, on définit également un voisinage

autour du neurone gagnant et on applique une règle similaire sur les voisins, mais avec un taux d'apprentissage différent :

$$\Delta w = \begin{cases} \eta_1(p - w) & \text{si le neurone est vainqueur ;} \\ \eta_2(p - w) & \text{si le neurone est voisin du vainqueur ;} \\ 0 & \text{autrement.} \end{cases} \quad (8.12)$$

avec $\eta_2 < \eta_1$

Annexe 9 : Algorithme de rétropropagation de l'erreur [PAR 04]

L'algorithme de rétropropagation standard se résume à la série d'étapes suivantes :

1. Initialiser tous les poids à de petites valeurs aléatoires dans l'intervalle $[-0.5, 0.5]$;
2. Normaliser les données d'entraînement ;
3. Permuter aléatoirement les données d'entraînement ;
4. Pour chaque donnée d'entraînement n :
 - (a) Calculer les sorties observées en propageant les entrées vers l'avant ;
 - (b) Ajuster les poids en rétropropageant l'erreur observée :

$$w_{ji}(n) = w_{ji}(n-1) + \eta \delta_j(n) y_i(n) \quad (9.1)$$

ou « le gradient local » est défini par :

$$\delta_j(n) = \begin{cases} e_j(n) y_j(n) [1 - y_j(n)] & ; \text{ si } j \in \text{couche de sortie} \\ y_j(n) [1 - y_j(n)] \sum_k \delta_k(n) w_{kj}(n) & ; \text{ si } j \in \text{couche cachée} \end{cases} \quad (9.2)$$

avec $0 \leq \eta \leq 1$ représentant le taux d'apprentissage et $y_i(n)$ représentant soit la sortie du neurone i sur la couche précédente, si celui-ci existe, soit l'entrée i autrement.

5. Répéter les étapes 3 et 4 jusqu'à un nombre maximum d'itérations ou jusqu'à ce que la racine de l'erreur quadratique moyenne (EQM) soit inférieure à un certain seuil.

Annexe 10 : Tableaux de données

Tableau.10.1-Evolution des différents facteurs et de la charge électrique mensuelle maximale

	Abonnés	POP (10⁶)	P. urbaine (10⁶)	Import CLIM (\$)	T° (°C)	Pmax (MW)
JAN/2000	4402062	30,003	16,847	595721702	5,8	4617
FEV	4414992	30,040	16,883	72807456	9,2	4357
MARS	4427921	30,078	16,919	100263895	12,1	4010
AVR	4440851	30,115	16,956	236198268	15,2	3834
MAI	4453781	30,153	16,992	266964023	20,2	3576
JUIN	4466711	30,190	17,028	265758702	22,8	3687
JUIL	4479640	30,228	17,064	316058341	27,1	3723
AOUT	4492570	30,266	17,101	346719913	26,5	3683
SEP	4505500	30,303	17,137	178293943	22,3	3739
OCT	4518430	30,341	17,173	200467296	16,1	3740
NOV	4531359	30,378	17,210	75675600	13,0	4251
DEC	4544289	30,416	17,246	209773595	10,6	4335
JAN/2001	4557928	30,455	17,288	130371054	9,0	4534
FEV	4571567	30,493	17,331	172389083	8,5	4515
MARS	4585207	30,532	17,373	177551936	15,0	4298
AVR	4598846	30,570	17,416	314820253	13,8	3927
MAI	4612485	30,609	17,458	206856861	17,8	3821
JUIN	4626124	30,647	17,501	168324724	24,5	3800
JUIL	4639763	30,686	17,543	468815443	27,2	3910
AOUT	4653402	30,725	17,586	251960844	27,2	3905
SEP	4667042	30,763	17,628	172206991	23,0	3913
OCT	4680681	30,802	17,670	268847210	21,3	3964
NOV	4694320	30,840	17,713	143815131	11,8	4372
DEC	4707959	30,879	17,755	241416853	8,1	4791
JAN/2002	4723681	30,919	17,797	158207931	8,1	4915
FEV	4739403	30,959	17,838	120372075	9,7	4683
MARS	4755124	30,998	17,879	214932259	12,5	4573
AVR	4770846	31,038	17,920	239169223	14,2	4131
MAI	4786568	31,078	17,961	439442623	18,6	3970
JUIN	4802290	31,118	18,003	572681450	24,4	4067
JUIL	4818011	31,158	18,044	731202759	25,8	4214
AOUT	4833733	31,198	18,085	235955560	25,3	4211
SEP	4849455	31,237	18,126	513916836	21,8	4261
OCT	4865177	31,277	18,167	233443232	18,4	4186
NOV	4880898	31,317	18,209	282055809	12,9	4529
DEC	4896620	31,357	18,250	471627996	10,6	4965
JAN/2003	4914733	31,398	18,290	221249957	7,5	5194
FEV	4932845	31,439	18,329	95527087	7,4	5206
MARS	4950958	31,480	18,369	419655716	11,7	4728
AVR	4969070	31,521	18,409	502487635	14,1	4646
MAI	4987183	31,562	18,448	601532340	18,2	4287
JUIN	5005295	31,603	18,488	592051892	25,3	4448
JUIL	5023408	31,643	18,528	414771598	28,6	4547
AOUT	5041521	31,684	18,568	355075656	27,5	4497
SEP	5059633	31,725	18,607	395436016	22,4	4480

OCT	5077746	31,766	18,647	170369756	19,0	4443
NOV	5095858	31,807	18,687	109130392	12,7	4651
DEC	5113971	31,848	18,727	297851718	8,5	5179
JAN/2004	5134123	31,891	18,768	278504571	8,1	5262
FEV	5154275	31,934	18,810	248567300	9,8	5084
MARS	5174427	31,977	18,851	374526422	11,8	5207
AVR	5194579	32,020	18,892	903302037	13,6	4764
MAI	5214731	32,063	18,934	869209300	16,3	4479
JUIN	5234883	32,106	18,975	1167065269	22,8	4619
JUIL	5255034	32,149	19,017	508133005	25,9	4719
AOUT	5275186	32,192	19,058	230688974	27,4	4841
SEP	5295338	32,235	19,100	363222409	22,7	4795
OCT	5315490	32,278	19,141	223672035	20,2	4722
NOV	5335642	32,321	19,183	188516214	10,8	5143
DEC	5355794	32,364	19,224	118811800	8,4	5541
JAN/2005	5376351	32,405	19,265	151885747	5,6	5859,2
FEV	5396908	32,445	19,305	291255175	5,7	5895
MARS	5417466	32,486	19,346	558331177	11,8	5744
AVR	5438023	32,527	19,387	535847384	14,2	5019
MAI	5458580	32,568	19,427	1169439342	20,6	4780
JUIN	5479137	32,608	19,468	901911846	23,8	5096
JUIL	5499694	32,649	19,508	848387718	27,7	5239
AOUT	5520251	32,690	19,549	668514347	26,1	5154
SEP	5540809	32,730	19,590	409807020	21,9	5074
OCT	5561366	32,771	19,630	258879094	18,8	4774
NOV	5581923	32,812	19,671	372968061	12,1	5707
DEC	5602480	32,853	19,712	319524551	9,0	5738

Tableau.10.4-Prévision de la charge électrique a l'horizon 2007

		Pmax sim (RNA)	Pmax sim (ARIMA)
2006	JAN	5754,8	6106,1
	FEV	5388,3	6029,0
	MARS	5283,5	5915,8
	AVR	5095,3	5492,3
	MAI	5273,7	5239,3
	JUIN	5423,4	5434,4
	JUIL	5395,6	5563,7
	AOUT	5297,2	5589,2
	SEP	5233,3	5563,1
	OCT	5607,1	5462,6
	NOV	5929,5	5976,2
	DEC	6067,5	6303,3
2007	JAN	6028,5	6582,5
	FEV	5701,5	6531,5
	MARS	5639,7	6427,7
	AVR	5426,3	5952,2
	MAI	5540,3	5719,4
	JUIN	5667,2	5936,8
	JUIL	5657,5	6080,4
	AOUT	5566,1	6084,4
	SEP	5505,7	6055,0
	OCT	5919,5	5921,4
	NOV	6202,2	6530,3
	DEC	6007,7	6802,7

Tableau.10.2-Prévision des facteurs

	Abonnés	T°
JAN/2006	5619376	7,3
FEV	5639623	8,4
MARS	5659870	12,5
AVR	5680117	14,2
MAI	5700364	18,6
JUIN	5720611	23,9
JUIL	5740858	27
AOUT	5761105	26,6
SEP	5781352	22,3
OCT	5801599	19
NOV	5821846	12,2
DEC	5842093	9,2
JAN/2007	5862340	7,3
FEV	5882587	8,4
MARS	5902834	12,5
AVR	5923081	14,2
MAI	5943328	18,6
JUIN	5963575	23,9
JUIL	5983822	27
AOUT	6004068	26,6
SEP	6024315	22,3
OCT	6044562	19
NOV	6064809	12,2
DEC	6085056	9,2

Tableau.10.3-Valeurs simulées de la charge électrique

		Pmax sim (RNA)	Pmax sim (ARIMA)			Pmax sim (RNA)	Pmax sim (ARIMA)
2000	JAN	-	-	2003	JAN	5169,8	5154,1
	FEV	4365	-		FEV	5199,4	5013,7
	MARS	4082,3	-		MARS	4755,4	4822,1
	AVR	3795,6	-		AVR	4547,7	4450,1
	MAI	3602,2	-		MAI	4327,7	4319,6
	JUIN	3650	-		JUIN	4435	4341,8
	JUIL	3728,9	-		JUIL	4519,9	4453,6
	AOUT	3718,1	-		AOUT	4503,8	4457,4
	SEP	3692,9	-		SEP	4453,1	4512,8
	OCT	3798,7	-		OCT	4443,4	4495,7
	NOV	4090,5	-		NOV	4807,9	4916,8
	DEC	4361,8	-		DEC	5295,2	5225,9
2001	JAN	4543,1	-	2004	JAN	5361,1	5424,5
	FEV	4598,1	4339,9		FEV	5198,8	5278,1
	MARS	4003,5	4054,3		MARS	5008,5	5002,2
	AVR	4076,9	4038,8		AVR	4824,5	4778,9
	MAI	3827,6	3761,8		MAI	4628,3	4567,0
	JUIN	3850,1	3854,3		JUIN	4618,9	4636,4
	JUIL	3947	3910,3		JUIL	4691,2	4740,3
	AOUT	3958,5	3879,9		AOUT	4761,1	4691,2
	SEP	3908,1	3955,7		SEP	4711,1	4744,4
	OCT	3919,9	3902,9		OCT	4688,3	4713,5
	NOV	4411,8	4526,2		NOV	5271	5120,6
	DEC	4822,8	4617,5		DEC	5608,1	5490,2
2002	JAN	4858,8	4900,4	2005	JAN	5961,9	5708,1
	FEV	4707,5	4764,5		FEV	5999,8	5605,8
	MARS	4427,3	4496,1		MARS	5403,4	5395,6
	AVR	4249,7	4240,8		AVR	5016,6	5100,7
	MAI	4028,4	4043,3		MAI	4798,1	4852,6
	JUIN	4116,9	4073,9		JUIN	4973,2	5007,8
	JUIL	4206,3	4186,9		JUIL	5133	5115,2
	AOUT	4211,9	4180,8		AOUT	5072	5163,6
	SEP	4195,4	4214,8		SEP	4989,2	5190,2
	OCT	4175,9	4247,4		OCT	4870,8	5178,3
	NOV	4537,5	4691,9		NOV	5573	5573,0
	DEC	4822,6	4925,7		DEC	6031,7	5974,4

Annexe 11 : Graphique de l'évolution des différents facteurs

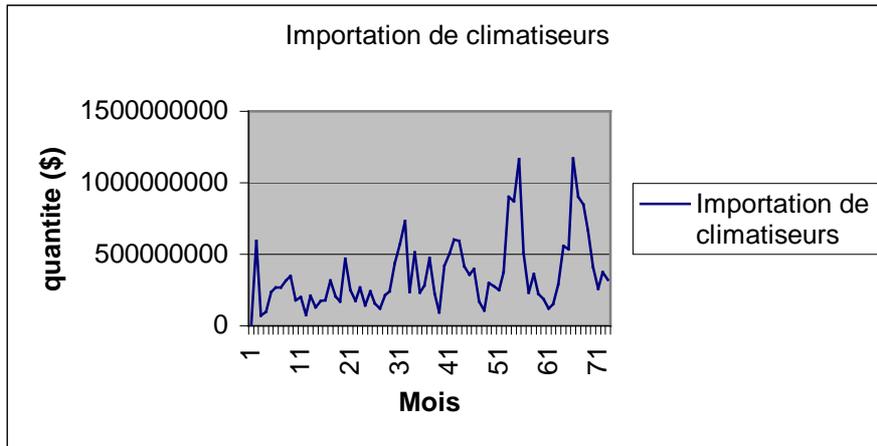


figure.11.1- Evolution des importations de climatiseurs

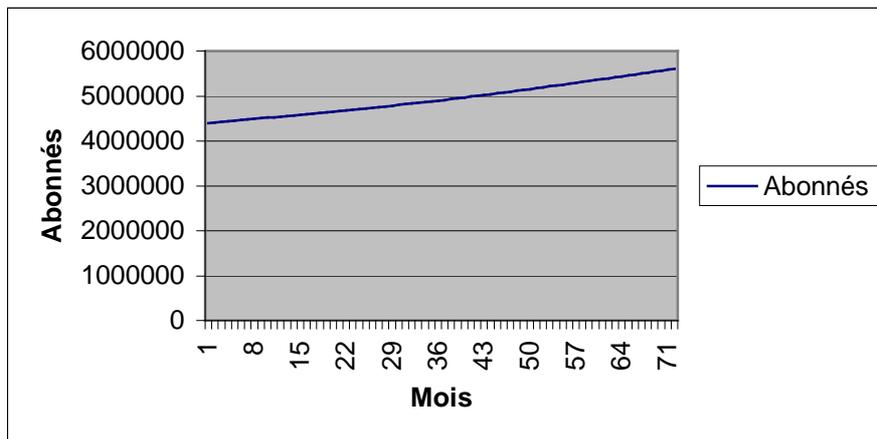


Figure.11.2-Evolution des abonnés SONELGAZ

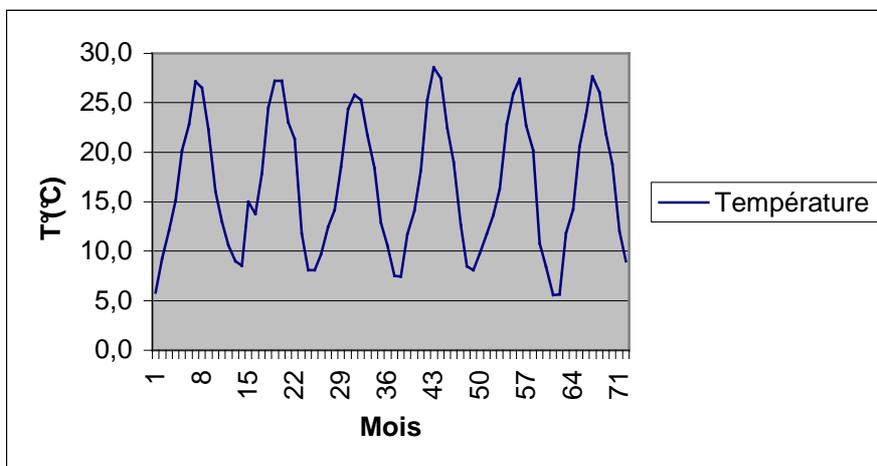


figure.11.3-Evolution des températures

Annexe 12 : Fonction d'autocorrélation des résidus

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Pierce	Prob.	
			+-----+-----+-----+-----+-----+-----+											
1	,024	,127					.	*				,037	,848	
2	,131	,126					.	I***				1,116	,572	
3	-,006	,125					.	*				1,119	,773	
4	,034	,124					.	I*				1,192	,879	
5	,143	,122					.	I***				2,548	,769	
6	,028	,121					.	I*				2,601	,857	
7	-,159	,120					.	***I				4,359	,738	
8	-,143	,119					.	***I				5,800	,670	
9	-,163	,118					.	***I				7,710	,564	
10	-,090	,117					.	**I				8,306	,599	
11	-,031	,115					.	*I				8,376	,679	
12	-,082	,114					.	**I				8,890	,712	
13	-,177	,113					.	****I				11,330	,583	
14	-,159	,112					.	***I				13,361	,498	
15	-,052	,111					.	*I				13,583	,557	
16	-,015	,109					.	*				13,603	,628	
17	-,068	,108					.	*I				13,997	,667	
18	,123	,107					.	I**				15,325	,640	
19	,073	,105					.	I*				15,807	,670	
20	,154	,104					.	I***				17,982	,589	
21	,061	,103					.	I*				18,339	,627	
22	,042	,101					.	I*				18,512	,675	
23	,192	,100					.	I****				22,213	,507	
24	,015	,099					.	*				22,235	,565	
25	,099	,097					.	I**				23,265	,562	
26	-,040	,096					.	*I				23,441	,608	
27	-,022	,094					.	*				23,494	,658	
28	-,132	,093					.	***I				25,530	,599	
29	,070	,091					.	I*				26,124	,619	
30	-,073	,090					.	*I				26,793	,634	
31	,064	,088					.	I*				27,312	,656	
32	-,201	,087					.	* **I				32,682	,433	
33	-,069	,085					.	*I				33,340	,451	
34	-,086	,083					.	**I				34,407	,448	
35	-,058	,082					.	*I				34,903	,473	
36	,013	,080					.	*				34,929	,519	
37	,013	,078					.	*				34,956	,565	
38	-,006	,076					.	*				34,962	,611	
39	-,013	,075					.	*				34,994	,653	
40	-,063	,073					.	*I				35,751	,662	

Annexe 13 : Résultats d'exécution du programme

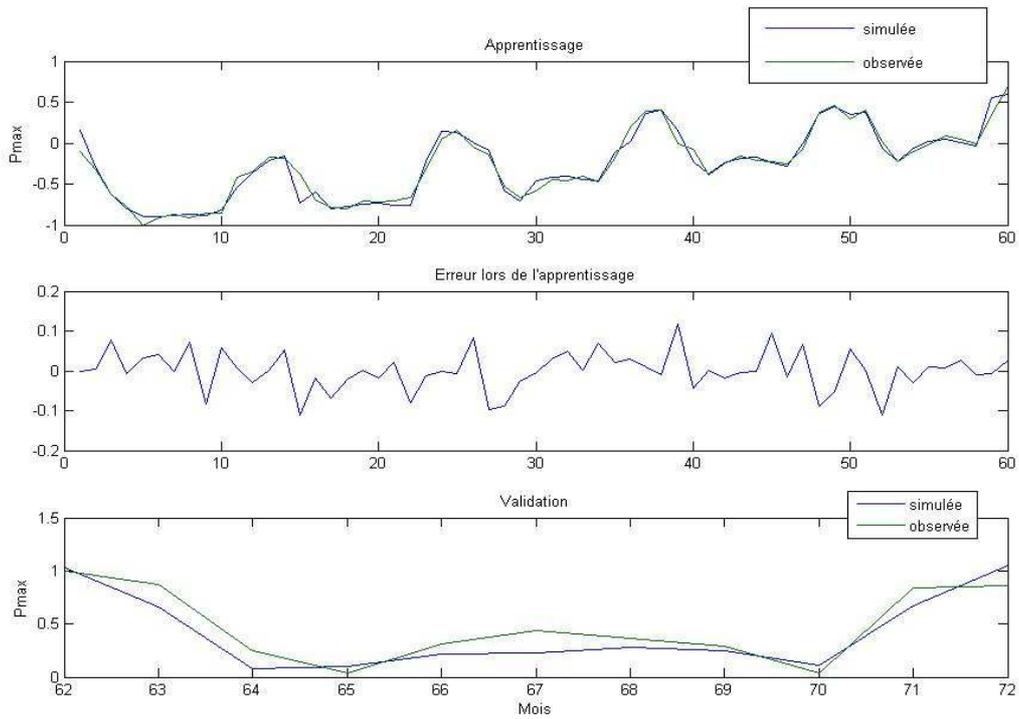


Figure.14.1- Performance du meilleur réseau lors de l'exécution « exemple »

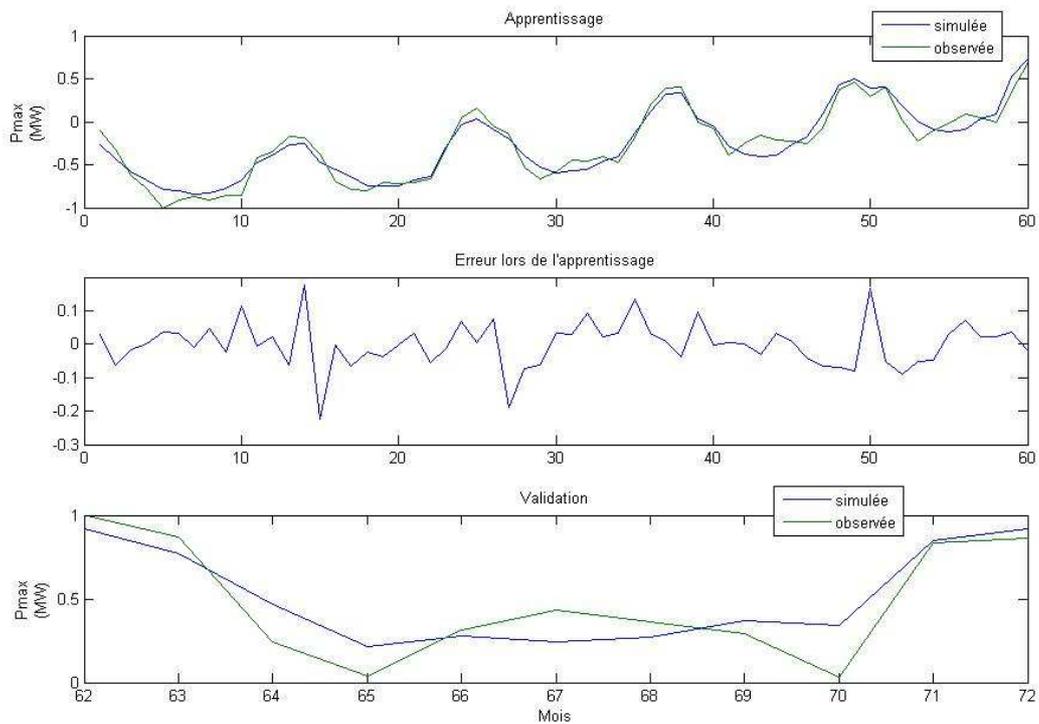


figure.14.1-Performance du réseau le moins bon de l'exécution « exemple »

Annexe 14 : Détails du programme développé sous MATLAB

```

%%%%%%%%Modélisation des puissances appelées mensuelles%%%%%%%%
load donnees.mat
for n=1:7
E=[-ones(3,1) ones(3,1)];
M=[Abonnn temperature Pmax]
Tn=Pmax(2:61);
Pr=[M(1,2:61);M(2,2:61);M(3,1:60)];
%%%%%%%%construction du réseau%%%%%%%%
net=newff;
net=newff(E,[1 1],{'tansig' 'purelin'},'trainlm')
net.trainParam.show = 5;
net.trainParam.epochs = 150;
net.trainParam.goal = 0;
%-----
NMSE=2;
MSE=1;
k=0;
MSE_opt=2;
%-----
while (MSE >= 0.01 | (NMSE > 1)
k=k+1;
%%%%%%%%phase d'apprentissage%%%%%%%%
net=init(net);
[net,tr,Y,Er] = train(net,Pr,Tn');
%plot(t(1:60),Y,t(1:60),Tn')
%plot(Er)
net.IW
net.b
%%%%%%%%phase de validation%%%%%%%%
Tn1=Pmax(62:72);
Pr1=M(:,62:72);
Y1=sim(net,Pr1);
%plot(t(62:72),Y1,t(62:72),Tn1')
var=mean((Pmax-mean(Pmax)).^2)
NMSE=(mean((Y1-Tn1').^2))/var ;
MSE=mean((Y1-Tn1').^2);
MAPE=mean((abs(Y1-Tn1'))/abs(Tn1'));
%%%%%%%%recherche du réseau optimal%%%%%%%%
if MSE_opt >= MSE
    net_opt=net;
    MSE_opt=MSE;
    MAPE_opt=MAPE;
end
if k > 200
    break
end
subplot(3,1,1);plot(t(1:60),Y,t(1:60),Tn')
subplot(3,1,2);plot(Er)
subplot(3,1,3);plot(t(62:72),Y1,t(62:72),Tn1')
end

```

```

%%%%%%%%sauvegarde des résultats%%%%%%%%
switch n
    case 1
        net1_opt=net_opt;
        MSE1=MSE_opt
        MAPE1=MAPE_opt
        n
    case 2
        net2_opt=net_opt;
        MSE2=MSE_opt
        MAPE2=MAPE_opt
        n
    case 3
        net3_opt=net_opt;
        MSE3=MSE_opt
        MAPE3=MAPE_opt
        n
    case 4
        net4_opt=net_opt;
        MSE4=MSE_opt
        MAPE4=MAPE_opt
        n
    case 5
        net5_opt=net_opt;
        MSE5=MSE_opt
        MAPE5=MAPE_opt
        n
    case 6
        net6_opt=net_opt;
        MSE6=MSE_opt
        MAPE6=MAPE_opt
        n
    otherwise
        net7_opt=net_opt;
        MSE7=MSE_opt
        MAPE7=MAPE_opt
        n
end
NMSE
MSE
k
% prevision
Y=sim(net_opt,M);
y=Y(72)
%test=y
YY=y
for s=73:96
    Pr_pre=[M(1,s);M(2,s);y];
    y=sim(net_opt,Pr_pre);
    %-----yy vecteur de sauvegarde des previsions y
    yy=[yy y ]
    %y=prevision;
end
end

```

Références bibliographiques :

- [ARR 02]: D.O.Arroyo M.K.Skov et Q.Huynh, “Accurate electricity load forecasting with artificial neural networks”, 2002. (<http://www.imi.aau.dk/~msk/publications/LFANN-CIMCA2005.pdf>)
- [BOU 98] : R.BOURBONNAIS et M.TERRAZA, “Analyse des séries temporelles en économie”, Ed Economie, Paris, 1998.
- [BOU 02] : R.BOURBONNAIS, “Econométrie”, Ed Dunod, Paris, 2002.
- [CHA 03] : A.CHARPENTIER, “Cours de séries temporelles : théorie et applications”, Université Paris Dauphine, 2003. (<http://www.crest.fr/pageperso/charpent/TS2.pdf>)
- [CHE 01]: M.W.CHANG, B.J.CHEN et C.J.LIN, “EUNITE network competition: electricity load forecasting”, 2001. (<http://citeseer.ist.psu.edu/677671.html>)
- [COU 99]: P.COULIBALY, F.ANCTIL et B.BOBEE, “Prévision hydrologique par réseaux de neurones artificiels : état de l’art”, CNRC Canada, vol 26, 293-304, 1999.
- [DAV 93] : E.DAVALO et P.NAIM, “Des réseaux de neurones”, Ed Eyrolles, Paris, 1993.
- [DEL 00] : D.DELIGNIERES, “Séries temporelles : modèle ARIMA”, Séminaire EA, Paris, 2000.
- [DEG 03] : S.DEGERINE, “Séries chronologiques”, E-books, 2003.
- [DRE 04] : G.DREYFUS, “Réseaux de neurones : méthodologie et applications”, Ed Eyrolles, Paris, 2004.
- [ECO 04]: Cours d’économétrie , Département du Génie Industriel, Ecole Nationale Polytechnique, Alger, 2004.
- [FIO 99] : A.FIORDALISO, “Systèmes flous et prévision de séries temporelles”, Ed Hermes Science, Paris, 1999.
- [FIS 99]: P.A.FISHWICK et Z.TANG, “Feed forward neural nets as model for time series forecasting”, TR 91-008, computer and information sciences, University of Florida,1999. (<http://citeseer.ist.psu.edu/cs>)

[FOU 04] : T.FOUCART, “Colinéarité et régression linéaire”, Université Pierre et Marie Curie, Paris, 2004.

[FRA 92] : K.FRANLING, “Les réseaux de neurones comme outils d’aide à la décision floue”, Rapport de DEA, I.N.S.A de Lyon, 1992.

[IWA 99]: H.IWAMIYA et B.KERMANSHAHI, “Long-term load forecasting using neural nets”, IEEE Transaction on Power Systems, vol 12, 13-17,1999.

[KAA 96]: I.KAASTRA et M.BOYD, “Designing a neural network for forecasting financial and economic time series”, Neurocomputing 10 (1996) 215-236, 1996. (<http://citeseer.ist.psu.edu/kaastra96designing.html>)

[MEL 90] : G.MELARD, “Méthodes de prévision à court terme”, Ed Ellipses, Paris, 1990.

[PAR 04] : M.PARIZEAU, “Réseaux de neurones”, Université de LAVAL, Canada, 2004.

[REF 94]: A.N. REFENES, A. ZAPRANIS et G. FRANCIS, “Stock performance modeling using neural networks: a comparative study with regression models”, Neural Network, 5 (1994) 961-970, 1994.

[SEN 96] : A.SENOUSSEI et M.T.BOURBON, “Elaboration d’un modèle de prévision de la charge électrique nationale en temps réel”, mémoire de projet de fin d’études, Département du Génie Industriel, Ecole Nationale Polytechnique, Alger, 1996.

[TAY 02]: J.W.TAYLOR et R.BUIZZA, “Neural network load forecasting with weather ensemble predictions”, IEEE Transaction on Power Systems, vol 17, 626-632, 2002. (<http://ieeexplore.ieee.org>)

[TOU 92] : C.TOUZET, “les réseaux de neurones artificiels : introduction au connexionisme”, E-books,1992.

[UPM 04] : “Cours d’économétrie”, Université Pierre et Marie Curie, Paris, 2004.

[YAO 00] : J.YAO et C.L.TAN, “A case study on using neural network to perform technical forecasting on forex”, Neurocomputing 34 (2000) 79-98, 2000.