

# Table des figures

2.1	Un perceptron. . . . .	21
2.2	Le Perceptron Multi-couche. . . . .	22
2.3	Le diagramme d'un Système Flou. . . . .	24
2.4	Les différents horizons de prévision. . . . .	25
2.5	Les secteurs concernés par la prévision de la charge électrique. . . . .	25
3.1	Variation horaire typique de la demande énergétique de différents types de charge électrique et la courbe totale d'une zone métropolitaine.[1] . . .	41
3.2	Représentation des différents facteurs utilisés dans les industries énergétiques. [2] . . . . .	42
3.3	Situation géographique de l'Algérie.[3] . . . . .	53
3.4	Les zones climatiques algériennes.[3] . . . . .	53
3.5	Disposition des gisements du sous-sol algérien.[3] . . . . .	54
3.6	Les principaux gisements algériens d'hydrocarbures.[3] . . . . .	55
3.7	Le planisphère des bassins de gaz et de pétrole de roche-mère.[3] . . . . .	56
3.8	Les principaux bassins de roche-mère en Algérie.[3] . . . . .	56
3.9	Les risques sur la disponibilité de l'eau.[3] . . . . .	57
3.10	Les différentes tensions dans un système de transmission et de distribution.[4] . . . . .	61
3.11	Production électrique par types d'énergie en Algérie. . . . .	64
3.12	La puissance installée par filière et par producteur pour l'année 2015. . .	67
4.1	La consommation énergétique algérienne globale de l'année 2000 à 2012. . .	75
4.2	La consommation moyenne de la clientèle BT de 2000 à 2012. . . . .	76
4.3	Comparaison des valeurs prédites et réelles de 2011 et 2012 du modèle linéaire SDE. . . . .	79
4.4	Prévision du modèle linéaire SDE à l'horizon 2023. . . . .	80
4.5	Prévision du modèle SDE en mode mixte à l'horizon 2030. . . . .	81
4.6	Le système de profils par saison. . . . .	87
4.7	Fenêtre de connexion des utilisateurs. . . . .	89
4.8	Le menu Accueil. . . . .	89
4.9	Dialogue de création d'une nouvelle étude. . . . .	90

4.10	Dialogue pour ouvrir une base de données existante. . . . .	91
4.11	Fenêtre de gestion des utilisateurs. . . . .	91
4.12	Le menu données. . . . .	92
4.13	Fenêtre d'ajout et de modification de l'historique. . . . .	93
4.14	Affichage d'une matrice de courbe de charge. . . . .	94
4.15	Calcul du taux d'occupation logement. . . . .	94
4.16	Fenêtre de création et de gestion des profils. . . . .	95
4.17	Le menu Analyse. . . . .	96
4.18	Profils de l'année 2012. . . . .	96
4.19	Menu prévision par niveau de tension. . . . .	97
4.20	Exemple de visualisation de la prévision R.I.N de la BT jusqu'à l'horizon 2030. . . . .	98
5.1	Le modèle X12-NAR-FFNN. . . . .	105
5.2	Évolution de la production en puissance algérienne entre 01/2000 et 12/2012.	106
5.3	Exemple de décomposition X12-ARIMA. . . . .	107
5.4	$D_{stat}$ des modèles pour l'horizon 24 mois. . . . .	110
5.5	Les erreurs MAPE de tous les modèles sur les différents horizons de pré- vision. . . . .	111
5.6	La prévision de la charge électrique algérienne de Janvier 2010 à Dé- cembre 2012 du modèle X12-NAR-FFNN. . . . .	113
A.1	Comptabilisation de l'erreur : écart entre Y observé et Y prédit par le modèle linéaire. . . . .	123

# Liste des tableaux

2.1	Tableau des poids les plus utilisés dans la méthode Moyenne Mobile pondérée. . . . .	12
3.1	La production en MW par énergie utilisée. . . . .	64
3.2	La clientèle et la consommation (GWh) par ville. . . . .	65
3.3	Evolution de la puissance produite en MW de chaque société de production du parc national algérien. . . . .	67
3.4	La puissance maximale appelée du RIN. . . . .	68
3.5	La puissance maximale appelée du PIAT. . . . .	68
3.6	La puissance maximale appelée du RIS. . . . .	68
4.1	L'échelle de temps et la rangée des données de la charge électrique algérienne. . . . .	73
4.2	Le taux de croissance annuel de la charge électrique algérienne. . . . .	74
4.3	Le taux de croissance annuel des trois tensions BT, MT et HT. . . . .	76
4.4	Statistiques de la régression. . . . .	77
4.5	Analyse de variance. . . . .	78
4.6	Récapitulatif du modèle linéaire de la SDE. . . . .	78
4.7	Tableau récapitulatif des différents modèles régressifs développés. . . . .	82
4.8	Profils Journaliers des quatre saisons de l'année 2012. . . . .	84
4.9	Profils hebdomadaires des quatre saisons de l'année 2012. . . . .	85
4.10	Profil annuel de l'année 2012. . . . .	86
4.11	Tableau récapitulatif des fonctions du menu Accueil. . . . .	90
4.12	Tableau récapitulatif des fonctions du menu Données. . . . .	92
5.1	Mesures de performance des modèles sur 3,6,12 et 24 mois. . . . .	112
A.1	Les valeurs nécessaires au calcul de la droite. . . . .	129

# Liste des abréviations

- ACP** Analyse des Composantes Principales.
- AGO** Accumulative generating operation.
- AGR** Secteur d'Activité de l'Agriculture.
- ANFIS** Adaptive NeuroFuzzy Inference System.
- ANN** Artificial Neural Network.
- AR** Auto-Régressifs.
- ARIMA** Autoregressive Integrated Moving Average.
- ARMA** Autoregressive Moving Average.
- AUT** Secteur d'Activité Autre.
- BP** Back Propagation.
- BPN** Back Propagation Network.
- BT** Basse Tension.
- BTALL** Basse Tension Globale.
- CC** Cycle Combiné.
- cli** Clientèle.
- consomoy** Consommation Moyenne.
- CREG** Commission de Régulation de l'Electricité et du Gaz.
- DSL** Diesel.
- E** Composante irrégulière.
- EeE** Secteur d'Activité Eau et Energie.
- EMD** Empirical Mode Decomposition.
- EMS** Energy Management System.
- ENR** Energie Renouvelable.
- EPR** Evolutionary Power Reactor.

**ES** Exponential Smoothing.

**FFNN** FeedForward Neural Network.

**FI** Fuzzy Intersection.

**FLR** Fuzzy Linear Regression.

**FNN** Fuzzy Neural Network.

**FOA** Fruit Fly Optimisation Algorithme.

**GM** Grey Model.

**GRNN** Generalized Regression Neural Network.

**HHNM** Hierarchical Hybrid Neural Model.

**HT** Haute Tension.

**HTALL** Haute Tension Globale.

**HYB** Hybride.

**HYD** Hydraulique.

**ID3** Iterative Dichotomiser 3.

**IEC** Commission Internationale électrotechnique.

**IEEE** Institute of Electrical and Electronics Engineers.

**IET** Institution of Engineering and Technology.

**IGCE** Secteur d'Activité des Industries Grosses Consommatrices d'Energie.

**IND** Secteur d'Activité des Industries diverses.

**IR** Irrégulière.

**JF** Jours Fériés.

**JO** Jours Ouvrables.

**LLFN** Locally Linear NeuroFuzzy.

**LM** Levenberg Marquardt.

**LoLiMoT** Locally Linear Model Tree.

**LS-SVM** Least Squares Support Vector Machine.

**LS-SVR** Least Squares Support Vector Regression.

**MA** Moyenne Mobile.

**MAD** Destruction Mutuelle Assurée.

**MAE** Mean Absolute Error.

**MAPE** Mean Absolute Percentage Error.

**MLP** MultiLayer Perceptron.

**MSE** Mean Squared Error.

**MT** Moyenne Tension.

**MTALL** Moyenne Tension Globale.

**NAR** Réseau de neurones autorégressif.

**P.I.B** Produit Intérieur Brut.

**P.M.E** Petites et Moyennes Entreprises.

**PIAT** Pôle In Salah-Adrar-Timimoune.

**PMA** Puissance Maximale Appelée.

**PNB** Produit National Brut.

**PSO** Particle Swarm Optimization.

**RBF** Radial Basis Function.

**RBMK** Reaktor Bolshoy Moshchnosti Kanalnyy.

**RES** Secteur d'Activité Résidentiel.

**RIN** Réseau Interconnecté du Nord.

**RIS** Réseau Isolé du Sud.

**RMA** Ratio Moving Average.

**RMSE** Root Mean Square Error.

**RNN** Recurent Neural Network.

**RVM** Relevance Vector Machine.

**S** Composante saisonnière.

**S.A.E** Secteur d'Activité Economique.

**SD** Société de Distribution.

**SD** Seasonal Decomposition.

**SDA** Société de distribution d'Alger.

**SDABT** Société de Distribution d'Alger de la Basse Tension.

**SDAHT** Société de Distribution d'Alger de la Haute Tension.

**SDC** Société de distribution Centre.

**SDCBT** Société de Distribution du Centre de la Basse Tension.

**SDCHT** Société de Distribution du Centre de la Haute Tension.

**SDE** Société de distribution Est.

**SDEBT** Société de Distribution Est de la Basse Tension.

**SDEHT** Société de Distribution Est de la Haute Tension.

**SDO** Société de distribution Ouest.

**SDOBT** Société de Distribution Ouest de la Basse Tension.

**SDOHT** Société de Distribution Ouest de la Haute Tension.

**SF** Fluctuation saisonnière.

**SKB** Shariket Kahraba Berrouaghia.

**SKD** Shariket Kahraba de Koudiet Edraouch.

**SKH** Shariket Kahraba Hadjret Ennouss.

**SKS** Shariket Kahraba Skikda.

**SKT** Shariket Kahraba Terga.

**SKTM** Shariket Kahraba wa Taket Moutadjadida.

**SOM** Self-Organizing Map.

**SPE** Société Algérienne de Production de l'Électricité.

**SPP1** Solar Power Plant One.

**SVM** Support Vector Machine.

**T** Composante tendance-cycle.

**TC** Tendance-Cycle.

**TER** Secteur d'Activité Tertiaire.

**TG** Thermique Gaz.

**TOL** Taux d'Occupation par Logement.

**TRA** Secteur d'Activité du Transport.

**TV** Thermique Vapeur.

**TVA** Taxe sur la valeur ajoutée.

**UHT** Ultra-Haute Tension.

**VDE** Vanguard Energy.

**VJO** Veilles Jours Ouvrables.

# Table des matières

Dédicace . . . . .	i
Remerciements . . . . .	ii
Résumé en Arabe . . . . .	iii
Abstract . . . . .	iv
Résumé . . . . .	vi
Publications . . . . .	viii
Table des figures . . . . .	ix
Liste des tableaux . . . . .	xi
Liste des abréviations . . . . .	xii
Table des matières . . . . .	xvi
<b>1 Introduction Générale</b>	<b>2</b>
1.1 Contributions . . . . .	5
1.2 Organisation de la thèse . . . . .	6
<b>2 État de l’art sur la prévision de la charge électrique</b>	<b>8</b>
2.1 Introduction . . . . .	9
2.2 Définition et analyse des séries chronologiques . . . . .	9
2.3 Décomposition saisonnière . . . . .	10
2.3.1 Décomposition classique . . . . .	11
2.3.1.1 Décomposition additive . . . . .	12
2.3.1.2 Décomposition multiplicative . . . . .	13
2.3.2 L’algorithme X12-ARIMA . . . . .	14
2.4 Approches utilisées . . . . .	15
2.4.1 Méthodes de prévision statistiques . . . . .	17
2.4.1.1 Analyse régressive . . . . .	17
2.4.1.2 La méthode Box-Jenkins . . . . .	18
2.4.2 Méthodes de prévision computationnelles . . . . .	20
2.4.2.1 Réseaux de Neurones artificiels . . . . .	21
2.4.2.2 La logique floue . . . . .	23
2.5 Horizons de prévision . . . . .	24



2.5.1	Moyen terme . . . . .	26
2.5.2	Long terme . . . . .	29
2.6	Conclusion . . . . .	33
<b>3</b>	<b>Spécificité de la charge électrique algérienne</b>	<b>35</b>
3.1	Introduction . . . . .	36
3.2	Découverte de l'électricité . . . . .	37
3.3	Production électrique . . . . .	39
3.4	Les différents types de centrales électriques . . . . .	43
3.4.1	Par chute d'eau . . . . .	43
3.4.2	Par énergie thermique . . . . .	45
3.4.2.1	Centrales à flamme . . . . .	46
3.4.2.2	Centrales nucléaires . . . . .	47
3.4.2.3	Autres centrales . . . . .	49
3.4.3	Par gaz naturel . . . . .	51
3.5	Présentation de l'Algérie . . . . .	52
3.6	Spécificité de la charge électrique algérienne . . . . .	57
3.6.1	Parc de Production National . . . . .	58
3.6.2	Données par énergie . . . . .	59
3.6.2.1	Haute Tension (HT) . . . . .	61
3.6.2.2	Basse Tension (BT) . . . . .	62
3.6.2.3	Moyenne Tension (MT) . . . . .	63
3.6.2.4	La production énergétique . . . . .	63
3.6.2.5	Infrastructures et distribution de l'électricité . . . . .	64
3.6.3	Données par puissance . . . . .	66
3.6.3.1	Le réseau interconnecté du Nord (RIN) . . . . .	67
3.6.3.2	Le pôle In Salah-Adrar-Timimoune (PIAT) . . . . .	68
3.6.3.3	Les réseaux isolés du Sud (RIS) . . . . .	68
3.6.4	Variables exogènes . . . . .	69
3.6.4.1	Produit Intérieur Brut . . . . .	69
3.6.4.2	Taux d'Occupation par Logement et clientèle . . . . .	69
3.7	Conclusion . . . . .	70
<b>4</b>	<b>Approches régressives pour la prévision à moyen long terme</b>	<b>71</b>
4.1	Introduction . . . . .	72
4.2	Disponibilité des données . . . . .	72
4.3	Étude de la charge électrique algérienne . . . . .	73
4.3.1	Prévision de la Société de distribution Est (SDE) . . . . .	77
4.4	Le système de profils . . . . .	83
4.5	PREVELEC-DZ . . . . .	86

4.5.1	Technologies utilisées et environnement de développement . . . .	87
4.5.2	Fonctionnalités de PREVELEC-DZ . . . . .	88
4.5.2.1	L'Accueil . . . . .	89
4.5.2.2	Menu données . . . . .	91
4.5.2.3	Menu Analyse . . . . .	95
4.5.2.4	Menu Pré.Niv.Tension . . . . .	97
4.6	Conclusion . . . . .	98
<b>5</b>	<b>Approches multi-modèle pour la prévision de la charge électrique à moyen long terme</b>	<b>100</b>
5.1	Introduction . . . . .	101
5.2	Problématique et contribution . . . . .	102
5.3	Résultats et Expérimentation . . . . .	104
5.3.1	Pré-traitement et données utilisées . . . . .	104
5.3.2	Mesures de performance . . . . .	107
5.4	Discussion et Expérimentation . . . . .	108
5.5	Conclusion . . . . .	113
<b>6</b>	<b>Conclusion</b>	<b>115</b>
6.1	Perspectives et travaux futurs . . . . .	118
<b>A</b>	<b>Calculs de la régression</b>	<b>119</b>
A.1	Les modèles régressifs . . . . .	120
A.1.1	La régression linéaire simple . . . . .	120
A.1.1.1	Hypothèses . . . . .	122
A.1.1.2	Estimation des moindres carrés ordinaires (MCO) . . .	122
A.1.1.3	Le point moyen . . . . .	124
A.1.1.4	Calcul des coefficients . . . . .	124
A.1.2	La régression linéaire multiple . . . . .	125
A.1.2.1	Les hypothèses . . . . .	126
A.1.2.2	Minimisation de la somme des carrés des erreurs . . .	127
A.2	La régression exponentielle . . . . .	128
A.3	Exemple . . . . .	129
	<b>Bibliographie</b>	<b>130</b>

# **Chapitre 1**

## **Introduction Générale**

L'impact de l'analyse des séries chronologiques sur les applications scientifiques peut être documenté par la production d'une liste non exhaustive des divers domaines dans lesquels d'importants problèmes de séries chronologiques peuvent survenir. Par exemple, de nombreuses séries chronologiques se produisent dans le domaine de l'économie, où nous sommes continuellement exposés aux cotations boursières ou aux chiffres mensuels du chômage. Les sociologues étudient les éléments relatifs à la population, tels que le taux de natalité ou les effectifs scolaires. Un épidémiologiste pourrait être intéressé par le nombre de cas de grippe observé sur une certaine période de temps. En médecine, les mesures de la pression artérielle tracées au fil du temps pourraient être utiles pour l'évaluation des médicaments utilisés dans le traitement de l'hypertension artérielle. Des séries chronologiques créées à partir d'imagerie d'ondes par résonance magnétique fonctionnelle du cerveau peuvent être utilisées pour étudier la façon dont le cerveau réagit à certains stimuli dans diverses conditions expérimentales.

La plupart des applications, les plus intensives et sophistiquées, des méthodes de séries chronologiques sont liées à des problèmes dans le domaine des sciences physiques et environnementales. Une des plus anciennes séries enregistrées est le nombre de tâches solaires mensuelles étudiées par Schuster (1906). D'autres recherches modernes peuvent montrer si un réchauffement est présent dans les mesures de température ou si les niveaux de pollution peuvent influencer sur la mortalité quotidienne. La modélisation de la série de la parole est un problème important lié à la transmission efficace des enregistrements vocaux. Les séries temporelles connues sous le nom de « spectre de puissance » sont utilisées pour aider les ordinateurs à reconnaître et à traduire la parole. Les séries chronologiques géophysiques, telles que celles produites par des dépôts annuels de différents types, peuvent fournir des procurations à long terme pour la température et les précipitations. Les enregistrements sismiques peuvent aider à cartographier les lignes de faille ou à faire la distinction entre les tremblements de terre et les explosions nucléaires. Les séries énergétiques, où l'on enregistre la consommation énergétique d'une région ou d'un pays, peuvent représenter la consommation ou la demande de tout le réseau, et la production en puissance des centrales électriques. L'étude de ce genre de séries chronologiques permet de prévoir les prises de décisions futures.

Le développement économique, à travers le monde, dépend directement de la disponibilité de l'énergie électrique, en particulier parce que la plupart des industries dépendent presque entièrement de son utilisation. La disponibilité d'une source d'énergie continue, pas chère et fiable est d'une importance économique avant tout. Ainsi, la prévision de la charge électrique est un outil important utilisé pour veiller à ce que l'énergie fournie par les services publics réponde à la demande en plus de l'énergie perdue dans le système. A cette fin, une équipe qualifiée est nécessaire pour mener à bien cette fonction. La prévision de la charge est fondamentalement définie comme étant la science ou l'art de prédire

la charge future sur un système donné, pour une période de temps déterminée à l'avance. Ces prédictions peuvent être justes pour une période d'une heure à des fins d'exploitation, ou jusqu'à 20 ans ou plus dans le futur à des fins de planification.

La prévision de la charge électrique peut être catégorisée en trois grands horizons, qui sont :

1. La prévision à longue échéance ; qui est la prévision de la charge jusqu'à 50 ans à l'avance pour que la planification et l'expansion puissent être facilitées.
2. La prévision à moyen terme ; qui est la prévision des pointes de la charge électrique sur un niveau hebdomadaire, mensuel ou annuel et peut aller jusqu'à 10 ans afin qu'une planification opérationnelle efficace puisse être effectuée.
3. La prévision à court terme ; qui est une prévision qui peut aller jusqu'à une semaine. L'intérêt de cette catégorie est de minimiser l'exécution quotidienne et les frais d'envoi. [5]

Dans les trois catégories précédentes, un modèle précis de la charge est nécessaire pour représenter mathématiquement la relation entre la charge et les variables influentes telles que : le temps, la météo, les facteurs économiques, etc. [6] La relation précise entre la charge et ces variables est généralement déterminée par leurs rôles dans le modèle de la charge. Une fois que le modèle mathématique est construit, ses paramètres seront déterminés par l'utilisation de techniques d'estimation.

La prévision consiste en l'extrapolation d'une relation mathématique sur un horizon donné ; ceci implique que les variables exogènes correspondantes soient disponibles ou prévisibles, même si ces dernières peuvent ne pas être obligatoires pour la construction du modèle mathématique, à cause de la difficulté de prévision avec précision pour le long terme des facteurs, tels que les conditions météorologiques et indices économiques qui sont de plus en plus difficiles à prédire. Plus le temps avance, moins précise la prédiction est susceptible d'être.

Au cours des 50 dernières années, les algorithmes d'estimation des paramètres utilisés dans la prévision de la charge étaient limités à ceux qui sont fondés sur le critère d'erreur de minimisation des moindres carrés, même si la théorie de l'estimation indique que les algorithmes basés sur les critères de valeurs absolues sont des alternatives viables. En outre, les réseaux de neurones artificiels ANN ont montré leur efficacité dans l'estimation de la charge à court terme. Cependant, l'ANN doit être recyclé pour convenir à une utilisation autre que le court terme.

À long terme, la prévision électrique de la charge est une question importante dans

une planification efficace. La sur ou sous-estimation peut grandement affecter les recettes de l'industrie de l'électricité. La surestimation de la charge peut conduire à dépenser plus d'argent dans la construction de nouvelles centrales électriques pour alimenter cette charge. Par ailleurs, la sous-estimation de la charge peut causer des problèmes dans sa fourniture ce qui peut conduire à une pénurie dans la réserve tournante du système. Par conséquent, un procédé précis est nécessaire pour prévoir la charge électrique, tout comme un modèle précis qui prend en compte les facteurs qui affectent la croissance de la charge sur un certain nombre d'années. De plus, un algorithme précis est nécessaire pour estimer les paramètres de ces modèles.

La croissance de la consommation d'électricité dans de nombreux pays en développement a dépassé les projections existantes et, par conséquent, les incertitudes de la prévision ont augmenté [7]. Des variables telles que la croissance économique, la population et les normes d'efficacité, associées à d'autres facteurs inhérents au développement mathématique des modèles de prévision, rendent difficile la projection d'une variable [7], [8]. Malheureusement, une prévision précise dépend du jugement du prévisionniste, et il est impossible de se fier uniquement aux procédures analytiques pour obtenir une prévision précise. [9]

## 1.1 Contributions

La thèse a pour objectif de proposer des solutions au problème de la prévision de la charge électrique algérienne pour le moyen-long terme, en utilisant des méthodes issues de la statistique ainsi que de l'intelligence artificielle. L'étude que nous proposons entre dans le cadre d'un projet communément initié par le laboratoire LabGed et l'entreprise nationale d'électricité et de gaz SONELGAZ avec lesquels nous avons collaboré dans des projets antérieurs, notamment pour la prévision à court terme de la charge électrique algérienne.

Le déroulement de la thèse s'est fait en plusieurs étapes ; chacune avait pour but de répondre à un problème spécifique :

1. La première étape a consisté en la compréhension et l'analyse des données fournies par notre partenaire, c'est une étape indispensable dans tout projet prévisionnel. Une quinzaine de séries chronologiques ont été étudiées dans cette étape suite à laquelle les variables exogènes affectant chaque tension ont été désignées. En plus d'un état de l'art des méthodes et des modèles proposés dans la littérature pour la prévision de la charge à moyen et long terme.

2. Dans une seconde étape, on a procédé à la modélisation des modèles prévisionnels ; l'objectif était d'utiliser les outils issus de la statistique, comme la régression linéaire multiple et exponentielle et ceci sur une demande émise par Sonelgaz. La proposition d'un modèle mixte qui se compose d'un modèle exponentiel pour le court terme et d'un modèle linéaire pour le long terme. Quarante-quatre modèles ont été proposés pour une prévision allant jusqu'en 2030.
3. Par la suite, un outil a été développé. Ce dernier, dont la fonction principale est l'implémentation des modèles réalisés dans l'étape 2, intègre également d'autres fonctionnalités ; comme la gestion des utilisateurs, la modification et la visualisation des données, la création et la gestion des profils de charge etc. L'outil proposé est nommé PREVELEC-DZ qui est conçu pour la plateforme Microsoft Windows.
4. La dernière étape fût l'utilisation des méthodes de l'intelligence artificielle pour la prévision de la charge électrique à moyen-long terme. Les réseaux de neurones connus pour leur efficacité et robustesse en ce qui concerne la prévision des séries temporelles, sont également connus pour être gourmands en données d'apprentissage pour l'obtention de résultats satisfaisants. La contribution de cette étape fût la proposition d'un multi-modèle uni-varié, basé sur les réseaux de neurones, qui permet d'avoir la puissance des réseaux de neurones avec peu de données d'apprentissage ; et de réduire la complexité de la prévision liée à l'utilisation des variables exogènes.

## 1.2 Organisation de la thèse

La thèse proposée est organisée comme suit :

**Chapitre 2 :** Le chapitre 2 est un état de l'art de la prévision de la charge électrique, où les concepts de base des séries chronologiques sont décrits. L'accent est mis plus particulièrement sur les méthodologies utilisées dans le traitement des séries chronologiques, les différentes méthodes de décomposition d'une série, ainsi qu'une présentation non exhaustive des méthodes statistiques et computationnelles utilisées dans la littérature pour la modélisation et prévision des séries chronologiques. Le chapitre se termine par un état de l'art des travaux réalisés sur la prévision de la charge électrique à moyen et à long terme.

**Chapitre 3 :** Ce chapitre aborde l'histoire de l'électricité, tous les inventeurs qui ont contribué à sa découverte et à son développement tel qu'on le connaît aujourd'hui. Une description non exhaustive de quelques types de générateurs d'électricité est présentée dans ce chapitre, ainsi que des notions (facteurs) relatives aux motivations qui poussent

vers la construction de nouveaux dispositifs. Par la suite, une présentation de l'Algérie ; sa situation géographique, les ressources disponibles dans son sous-sol, ainsi que la présentation de la spécificité de la charge électrique algérienne ; en abordant le parc électrique et la distribution de l'électricité sur le territoire national. Pour conclure le chapitre, nous évoquerons les variables économiques (appelées exogènes) qui ont un rapport avec l'étude de l'électricité ; tel que le Produit Intérieur Brut (P.I.B), la clientèle, etc.

**Chapitre 4 :** Ce chapitre est consacré à la partie régressive de cette thèse, nous allons aborder les spécificités de la charge électrique algérienne, nous présenterons aussi les données disponibles, ainsi qu'une analyse des séries chronologiques et la sélection des variables exogènes. Nous exposerons quelques modèles prévisionnels et leurs résultats, quarante-quatre modèles ont été réalisés dans cette étude. Et pour finir, nous allons présenter l'outil développé PREVELEC-DZ en mettant en évidence les parties importantes de ce dernier.

**Chapitre 5 :** Dans ce chapitre nous allons parler de la problématique et de la contribution proposée, pour répondre aux problèmes liés à la complexité de la prévision d'une série temporelle fortement corrélée et l'obtention de résultats satisfaisants avec un ensemble de données qui est petit et inadéquat pour une prévision, en utilisant les méthodes computationnelles généralement gourmandes en données d'apprentissage.

**Conclusion :** Une conclusion sur les travaux réalisés, ainsi qu'une discussion sur les perspectives.



## **Chapitre 2**

# **État de l'art sur la prévision de la charge électrique**

## 2.1 Introduction

Les séries temporelles, appelées aussi séries chronologiques ou même chroniques, occupent une place importante dans tous les domaines de l'observation ou de la collecte de données. Le terme série est employé pour évoquer des objets (des nombres ou des mots) classés dans un certain ordre. L'ordre utilisé est le temps, plus précisément, on utilise une mesure du temps exprimée en : années, mois, jours, minutes ou n'importe quelle autre unité de mesure. En d'autres termes, les séries temporelles associent des objets divers à des marques temporelles successives plus ou moins équidistantes. La série est dite temporelle, parce qu'elle indexe l'objet ou valeur par le temps [10].

Dans ce chapitre, nous allons décrire les concepts de base des séries chronologiques. Nous mettrons l'accent plus particulièrement sur la méthodologie utilisée dans le traitement des séries chronologiques, les différentes méthodes de décomposition d'une série, ainsi que les méthodes statistiques et computationnelles utilisées pour la modélisation et la prévision d'une série chronologique. Nous terminerons ce chapitre par un état de l'art sur les travaux réalisés sur la prévision de la charge électrique à moyen et à long terme ; qui est considérée comme une série chronologique.

## 2.2 Définition et analyse des séries chronologiques

Soit  $(\Omega, \mathcal{A}, P)$  un espace de probabilité, et  $T$  un ensemble d'index. Une série temporelle ou processus stochastique est une fonction  $X(t, w)$  définie sur  $T \times \Omega$  tel que pour chaque valeur de  $t$ ,  $X(t, w)$  une variable aléatoire de l'espace de probabilité  $(\Omega, \mathcal{A}, P)$ . La fonction  $X(t, w)$  est aussi écrite  $X_t(w)$  ou  $X_t$ . Ainsi, une série temporelle peut être considérée comme une collection de variables aléatoires  $X_t : t \in T$ . [11] L'étude d'une série chronologique permet d'analyser, de décrire et d'expliquer un phénomène au cours du temps et d'en tirer des conséquences pour des prises de décisions. L'un des objectifs principaux de l'étude d'une série chronologique est la prévision ; qui consiste à extrapoler les valeurs observées dans un futur plus ou moins lointain.

Les principales étapes de traitement d'une série chronologique sont les suivantes :

1. Correction des données.
2. Observation de la série.
3. Modélisation.

#### 4. Prévision.

Une règle générale en statistique descriptive consiste à commencer par analyser les données et tracer un graphique représentant ces dernières. L'observation de ce graphique est souvent une aide à la modélisation de la série temporelle. La modélisation d'une série consiste en la création d'un modèle, qui est une image simplifiée de la réalité, qui vise à traduire les mécanismes de fonctionnement du phénomène étudié et permet de mieux les comprendre. Qu'ils soient mathématiques ou issus de l'intelligence artificielle, différents modèles sont possibles pour une série temporelle. Grâce à ces modèles, une prévision des futures valeurs est possible.[12]

### 2.3 Décomposition saisonnière

Les séries temporelles peuvent montrer une variété de patterns qu'il serait efficace de catégoriser pour pouvoir extraire des caractéristiques spécifiques à chacun. La décomposition de la série temporelle en plusieurs composantes, chacune représentant une catégorie de patterns, est l'une des manières de distinguer les caractéristiques spécifiques à une série.

Chaque série  $y_t$  comprend 3 composantes :

- Une composante saisonnière notée  $S$ .
- Une composante tendance-cycle (contenant la tendance et le cycle) notée  $T$ .
- Une composante irrégulière (contenant le reste de la série temporelle) notée  $E$  [13].

Si on prend en considération le modèle additif pour une série temporelle, on peut écrire la formule 2.1 :

$$y_t = S_t + T_t + E_t \quad (2.1)$$

où  $y_t$  représente les données au moment  $t$ , Composante saisonnière ( $S$ ) au moment  $t$ ,  $T_t$  la Composante tendance-cycle ( $T$ ) composante à la période  $t$  et  $E$  la Composante irrégulière au moment  $t$ . De la même manière un modèle multiplicatif s'écrit comme montré dans l'équation 2.2

$$y_t = S_t * T_t * E_t \quad (2.2)$$

Le modèle additif est plus approprié pour les séries temporelles dont la magnitude de la fluctuation saisonnière ou la variation autour de la Composante tendance-cycle (T) ne varie pas avec le niveau de la série temporelle. Autrement, le modèle multiplicatif est à prendre en considération. [13]

Une alternative existe au modèle multiplicatif, et consiste en une transformation de la série jusqu'à la rendre stable, du point de vue de la variation, sur toute la période où la série est définie. Suite à cela, un modèle additif est suffisant. La transformation logarithmique est équivalente à l'utilisation d'un modèle multiplicatif (Voir équation 2.3 ).

$$y_t = S_t * T_t * E_t \Leftrightarrow \log(y_t) = \log(S_t) * \log(T_t) * \log(E_t) \quad (2.3)$$

Dans certains cas, la composante tendance-cycle est appelée simplement "Tendance" et contient la composante cyclique.

### 2.3.1 Décomposition classique

La méthode classique de la décomposition a été développée dans les années 1920 et a beaucoup été utilisée jusqu'en 1950. Cette méthode est à la base des autres techniques de décomposition des séries temporelles. C'est pour cela qu'il est important de comprendre la décomposition classique en premier lieu. La première étape de cette méthode est la Moyenne Mobile (MA) qui permet d'estimer la tendance. La moyenne mobile d'ordre  $m$  peut-être écrite comme dans l'équation 2.4

$$\hat{T}_t = m \sum_{j=-k}^k y_{t+j} \quad (2.4)$$

où  $m = 2k + 1$ . L'estimation de la Composante tendance-cycle au moment  $t$  est obtenue par la moyenne de la série temporelle dans l'intervalle  $k$ . Les observations qui sont proches en temps sont généralement proches dans leurs valeurs, et la moyenne mobile appliquée à une partie des données, qui donne une composante tendance-cycle bien lissée,  $m - MA$  est la notation pour la moyenne mobile d'ordre  $m$ . Plus cette dernière est grande plus la courbe est lisse.

Dans le but d'avoir une symétrie du même ordre, une moyenne mobile de la moyenne mobile est possible, cela donne une moyenne mobile pondérée qui est décrite avec l'équation 2.5 :

$$\hat{T}_t = \sum_{j=-k}^k a_j y_{t+j} \quad (2.5)$$

où  $k = (m - 1)/2$  et les poids  $[a_{-k}, \dots, a_k]$ . La somme de tous les poids doit être égale à 1 et symétrique, si bien que  $a_j = a_{-j}$ .  $m - MA$  est un cas spécial où tous les poids sont égaux à  $1/m$ . L'avantage de cette méthode est qu'elle permet d'avoir une meilleure estimation de la tendance. Le tableau 2.1 expose les combinaisons de poids qui sont le plus souvent utilisées.

où  $S$  et  $H$  représentent respectivement la moyenne mobile pondérée de type Spencer et Henderson.

Il existe deux types de décomposition : une additive et une autre multiplicative, celles-ci sont décrites pour des séries temporelles avec période  $m$ . On assume que la composante saisonnière est constante sur toute la période de la série,  $m$  est aussi appelée "Indice saisonnier".

Name	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$
<b>3-MA</b>	.333	.333	-	-	-	-	-	-	-	-	-	-
<b>5 MA</b>	.200	.200	.200	-	-	-	-	-	-	-	-	-
<b>2x12-MA</b>	.083	.083	.083	.083	.083	.083	.042	-	-	-	-	-
<b>3x3-MA</b>	.333	.222	.111	-	-	-	-	-	-	-	-	-
<b>3x5-MA</b>	.200	.200	.133	.067	-	-	-	-	-	-	-	-
<b>S15-MA</b>	.231	.209	.144	.066	.009	-.016	-.019	-.009	-	-	-	-
<b>S21-MA</b>	.171	.163	.134	.037	.051	.017	-.006	-.014	-.014	-.009	-.003	-
<b>H5-MA</b>	.558	.294	-.073	-	-	-	-	-	-	-	-	-
<b>H9-MA</b>	.330	.267	.119	-.010	-.041	-	-	-	-	-	-	-
<b>H13-MA</b>	.240	.214	.147	.066	.000	-.028	-.019	-	-	-	-	-
<b>H23-MA</b>	.148	.138	.122	.097	.068	.039	.013	-.005	-.015	-.016	-.011	-.004

Tableau 2.1 – Tableau des poids les plus utilisés dans la méthode Moyenne Mobile pondérée.

### 2.3.1.1 Décomposition additive

La décomposition additive se fait suivant les étapes citées ci-dessous :

**Etape 1 :** Si  $m$  est un nombre paire, on calcule la tendance via  $2 * m - MA$  pour obtenir  $\hat{T}_t$ . Sinon, le calcul de la tendance se fait avec  $m - MA$  pour l'obtention de  $\hat{T}_t$ .

**Etape 2 :** Déduire la série sans tendance, il suffit de soustraire la série originale de la tendance.

**Etape 3 :** Pour estimer la Composante saisonnière, pour chaque mois, il suffit de faire

la moyenne de tous les mois de la série dépourvue de tendance. Les indexes sont ensuite ajustés puis liés ensemble pour obtenir la composante saisonnière  $\hat{S}_t$ .

**Etape 4 :** La Composante irrégulière  $\hat{E}_t$  est obtenue via l'équation 2.6

$$\hat{E}_t = y_t - \hat{S}_t - \hat{T}_t \quad (2.6)$$

### 2.3.1.2 Décomposition multiplicative

La décomposition multiplicative classique est similaire à la méthode additive mis à part que la soustraction est remplacée par la division.

**Etape 1 :** Si  $m$  est un nombre paire, on calcule la tendance via  $2 * m - MA$  pour obtenir  $\hat{T}_t$ . Sinon, le calcul de la tendance se fait avec  $m - MA$  pour l'obtention de  $\hat{T}_t$ .

**Etape 2 :** Déduire la série sans tendance en divisant la série originale de la tendance.

**Etape 3 :** Pour estimer la Composante saisonnière (S), pour chaque mois, il suffit de faire la moyenne de tous les mois de la série dépourvue de tendance. Les indexes sont ensuite ajustés puis liés ensemble pour obtenir la composante saisonnière  $\hat{S}_t$ .

**Etape 4 :** La composante irrégulière  $\hat{E}_t$  est obtenue via l'équation 2.7

$$\hat{E}_t = \frac{y_t}{\hat{S}_t * \hat{T}_t} \quad (2.7)$$

La décomposition classique reste très utilisée, néanmoins, elle n'est pas recommandée pour les raisons suivantes :

- L'estimation de la tendance n'est pas possible pour les premières et dernières observations de la série, ce qui implique une absence de composante irrégulière pour ces mêmes observations.
- Pour plusieurs séries ; l'hypothèse que la composante saisonnière se répète d'année en année est juste, néanmoins pour des séries qui sont définies sur une très longue période de temps cela n'est pas toujours vrai. La demande électrique est un bon exemple ; en effet, elle a beaucoup changé avec l'utilisation de la climatisation, et la décomposition classique ne peut capturer ce changement d'habitudes.
- La méthode n'est pas robuste pour les valeurs inhabituelles.

### 2.3.2 L'algorithme X12-ARIMA

La méthode de décomposition X12-ARIMA, très utilisée pour les données mensuelles et trimestrielles, tire ses origines des méthodes développées par le Bureau Américain du Recensement. Elle est très utilisée par les bureaux et agences du gouvernement à travers le monde. Les versions antérieures de cette méthode incluent X11 et X11-ARIMA. [14]

La méthode X12-ARIMA est basée sur la décomposition classique, mais avec de nombreuses étapes et des fonctionnalités supplémentaires pour pallier aux inconvénients de la décomposition classique qui ont été discutés précédemment. X12-ARIMA permet l'estimation de la tendance pour toutes les observations, et la composante saisonnière est autorisée à varier lentement dans le temps. Aussi, la méthode est relativement robuste en ce qui concerne les observations inhabituelles. X12-ARIMA n'est autorisée que pour les données mensuelles et trimestrielles, elle supporte la décomposition additive et multiplicative.

La partie « ARIMA » de la méthode fait référence au modèle Autoregressive Integrated Moving Average, connu dans la littérature, qui offre la possibilité de prévoir le futur d'une série temporelle ainsi que son passé. Suite à l'obtention de la tendance-cyclique grâce à la moyenne mobile, aucune perte d'informations, sur les premières et dernières observations de la série chronologique, n'est observée.

L'algorithme X12-ARIMA commence d'une manière similaire à la décomposition classique, et les composantes sont affinées à travers plusieurs itérations. La méthode est similaire pour une décomposition d'additifs et des données trimestrielles. L'algorithme de la méthode est décrit, dans ce qui suit, pour une décomposition multiplicative appliquée aux données mensuelles [15] :

1. Calculer la moyenne mobile de  $2 * 12$  sur les données originales pour obtenir une estimation grossière de la tendance-cyclique  $\hat{T}_t$  sur toute la période de la série temporelle.
2. Calculer ce qu'on appelle "centred ratio" ou le ratio centré via l'équation 2.8.

$$\frac{y_t}{\hat{T}_t} \quad (2.8)$$

3. Pour chaque ratio centré d'un mois, on calcule 3x3MA pour former une estimation de la composante saisonnière  $\hat{S}_t$ .
4. Ensuite pour obtenir la composante irrégulière  $\hat{E}_t$  on divise le ratio centré par  $\hat{S}_t$ .
5. Réduire les valeurs extrêmes de  $E_t$  pour obtenir une estimation de  $\hat{E}_t$  modifiée.

6. Multiplier l'estimation modifiée de  $\hat{E}_t$  par  $\hat{S}_t$  pour obtenir un ratio centré modifié.
7. Répéter l'étape 3 pour obtenir une composante saisonnière  $\hat{S}_t$  révisée.
8. Diviser les données originales par la dernière estimation de  $\hat{S}_t$  pour obtenir une série préliminaire ajustée saisonnièrement.
9. La tendance-cyclique  $\hat{T}_t$  est estimée en appliquant la moyenne mobile pondérée de type Henderson sur les valeurs préliminaires ajustées saisonnièrement. Pour les séries mensuelles, une moyenne mobile Henderson d'ordres 9, 13 ou 23 est utilisée.
10. Répéter l'étape 2 ; un nouveau ratio est obtenu en divisant les données originales par la dernière estimation de  $\hat{T}_t$ .
11. Répéter les étapes 3 jusqu'à 6 avec les valeurs des nouveaux ratios et utiliser 3x5MA à la place de 3x3MA.
12. Répéter l'étape 7 avec la moyenne mobile 3x3MA.
13. Répéter l'étape 8.
14. La composante irrégulière  $\hat{E}_t$  est obtenue en divisant les données de l'étape 13 par la tendance-cyclique obtenue dans l'étape 9.
15. Les valeurs extrêmes de la composante irrégulière sont remplacées de la même façon que dans l'étape 5.
16. Une nouvelle série est obtenue en multipliant les trois composantes : saisonnière, tendance-cyclique et irrégulière.

L'algorithme est répété deux fois en utilisant les données obtenues dans l'étape 16. Dans la dernière itération, la moyenne mobile 3x5MA de l'étape 11 et 12 est remplacée par respectivement 3x3, 3x5 ou 3x9 selon la variabilité des données.

X12-ARIMA propose en outre un module très complet, appelé Reg-ARIMA, permettant de corriger la série initiale de toute sorte d'effets indésirables. L'estimation de ces effets se fait grâce à l'utilisation de modèles de régression à erreurs ARIMA, en plus d'être connu pour sa précision et son efficacité pour les petites séries temporelles. [16]

## 2.4 Approches utilisées

Une fois le modèle de la série choisi et ses paramètres estimés, il peut être utilisé pour prévoir les valeurs futures en utilisant la méthode de prévision adéquate. Dans le passé, le terme intelligence computationnelle était étroitement lié au terme prédiction. Les



dernières recherches, qui ont pour intérêt de déterminer les valeurs futures d'une fonction empirique sur la base d'un ensemble de données, ont remplacé le terme prévision par prédiction. Néanmoins, le terme prévision reste généralement associé aux problèmes de l'analyse des séries temporelles, le terme prédiction reste préférable pour les systèmes et l'ingénierie de contrôle.

La prévision des valeurs futures d'une série temporelle est définie de la façon suivante :

- Etant donné un ensemble de valeurs  $x_1, x_2, x_3, \dots, x_n$  d'une série temporelle, l'estimation des valeurs  $x_{n+1}, x_{n+2}, \dots$ , doit être réalisable.
- La prévision de q-pas en avant, calculé à un point  $n$ , est notée  $\hat{x}_n(q)$ .

Généralement, les approches de prévision peuvent être classifiées en :

- Prévision objective ; basée sur ce qui est subjectif, comme : le jugement, l'intuition, les connaissances en commerce ainsi que toute autre information pertinente.
- Prévision uni-variée ; basée entièrement sur l'approximation d'un modèle à une dimension, les données collectées et l'extrapolation de la série temporelle.
- Prévision multi-variée ; basée sur l'observation simultanée de deux variables ou plus et sur un modèle de série temporelle multi-variée.

Dans la pratique, une approche de prévision peut combiner deux des approches mentionnées plus haut. Avant de sélectionner une méthode, il est essentiel de considérer la façon avec laquelle elle sera utilisée, l'exactitude de la prévision, les ressources disponibles, combien de variables seront prédites, le nombre de données disponibles et enfin le nombre de pas en avant nécessaire. Mis à part cela, certaines méthodes produisent des points de prévision, mais dans certains cas, il serait préférable d'avoir un intervalle de prévision.

Pour toutes les approches de prévision, il est recommandé comme première étape d'analyser la courbe graphique de la série, car certaines informations peuvent être obtenues suite à l'examen visuel de la courbe, ce qui aide à sélectionner la méthode de prévision appropriée. En plus du choix de la meilleure prévision, il est également important d'indiquer la précision avec laquelle les prévisions ont été déterminées, de sorte que le risque associé à des décisions basées sur des prévisions puisse être calculé.

## 2.4.1 Méthodes de prévision statistiques

Nous allons voir dans cette section une liste non exhaustive des méthodes de modélisation et prévision statistiques

### 2.4.1.1 Analyse régressive

La régression est un outil mathématique qui permet d'étudier la relation entre les variables observées. L'objectif principal est d'estimer et de prédire les valeurs d'une variable en prenant en considération les valeurs d'autres variables dites exogènes. Cependant, avant d'utiliser cette technique pour la prévision d'une variable, les variables exogènes reliées à cette dernière doivent être identifiées. 3 types de régressions peuvent être utilisés :

- Régression simple.
- Régression multiple.
- Régression non linéaire.

La prévision via la régression simple est basée sur l'équation 2.9 :

$$Y_i = a_0 + a_1 X_i + \varepsilon_i \quad (2.9)$$

où  $i = 1, 2, \dots, n$  et la valeur moyenne de l'erreur  $\varepsilon_i$  est supposée égale à 0, et sa variance à 1. Les valeurs des paramètres  $a_0$  et  $a_1$  doivent être estimées de façon à ce que

$$\sum_{j=-k}^n (y_i - a_0 - a_1 x_1)^2 \quad (2.10)$$

soit minimisée. Mais dans la majorité des cas pratiques, la régression est utilisée comme une relation mutuelle

$$y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_n X_n \quad (2.11)$$

entre les variables observées  $x_i, i = 1, 2, \dots, n$  et la variable résultat à estimer  $y$ . Le problème dans l'équation 2.11 est d'estimer les coefficients  $a_0, a_2, \dots, a_n$  ce qui garantit la meilleure approximation des données expérimentales par une ligne de régression.

La forme compacte de la régression linéaire multiple est représentée dans l'équation 2.12.

$$y = Ax + \varepsilon \quad (2.12)$$

où

$$y = [y_1, y_2, \dots, y_n]^T, x = [x_1, x_2, \dots, x_n]^T, \varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]^T \quad (2.13)$$

sont les vecteurs associés à l'équation 2.12 et la matrice des paramètres  $A$  est comme suit :

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

Pour obtenir la meilleure estimation de  $\hat{x}$ , on utilise les moindres carrés, on commence par la valeur de l'erreur  $\varepsilon$  avec l'équation 2.14

$$\varepsilon(\hat{x}) = (y - A\hat{x}) \quad (2.14)$$

La généralisation des moindres carrés revient à minimiser la forme linéaire de l'équation 2.15

$$(y - A\hat{x})B(y - A\hat{x})^T \quad (2.15)$$

où  $B = \text{diag}[c_1, c_2, \dots, c_n]$  est la matrice de définition diagonale.

#### 2.4.1.2 La méthode Box-Jenkins

Box et Jenkins ont développé une méthode de prévision généralisée pour les séries temporelles générées par un processus stationnaire.

#### 2.4.1.2.1 Le modèle Auto-Régressif AR(p)

Les modèles Auto-Régressifs (AR) expriment la valeur actuelle d'une série temporelle via une agrégation linéaire finie des observations passées.

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_v x_{t-v} + \mu_t \quad (2.16)$$

où  $a_1$  à  $a_v$  sont les paramètres d'auto-régression,  $\mu_t$  est un bruit blanc et  $v$  est l'ordre du modèle. La validité d'un modèle autorégressif est sous l'hypothèse que la série temporelle étudiée est stationnaire. Aussi, pour certains effets internes cumulatifs, le processus autorégressif est stable quand la valeur du paramètre  $a$  est dans un intervalle bien défini. Le calcul des paramètres du modèle autorégressif peut être réalisé via différentes méthodes, mais celle des moindres carrés (Section 2.4.1.1) est la plus utilisée.

#### 2.4.1.2.2 Le modèle Moyenne Mobile MA(q)

Une autre méthode est très utilisée pour la modélisation uni-variée des séries temporelles et qui est basée sur la méthode de la Moyenne Mobile (MA).

$$\tilde{Z}_t = a_t - \vartheta_1 a_{t-1} - \vartheta_2 a_{t-2} - \dots - \vartheta_q a_{t-q} \quad (2.17)$$

où  $\tilde{Z}_t$  exprime la somme des poids pondérés qui sont finis, la moyenne mobile est écrite sous sa forme compacte comme suit :

$$z_t = \vartheta(B) a_t \quad (2.18)$$

Le modèle  $MA(q)$  contient  $(q + 2)$  paramètres  $\mu, \vartheta_1, \vartheta_2, \dots, \vartheta_q, \sigma_a^2$  qui sont estimés à partir des données observées.

#### 2.4.1.2.3 Le modèle ARMA(p,q)

Le processus Autoregressive Moving Average (ARMA) est une combinaison des modèles AR(p) et MA(q), un tel modèle appelé autorégressif – moyenne mobile ARMA, est caractérisé par le paramètre  $p$  de la partie autorégressive et le paramètre  $q$  de la partie moyenne mobile.

Un processus ARMA(p, q) vérifie l'équation suivante :

$$\tilde{Z}_t = \varphi_1 \tilde{Z}_{t-1} + \varphi_2 \tilde{Z}_{t-2} + \dots + \varphi_p \tilde{Z}_{t-p} + a_t + \vartheta_1 a_{t-1} + \vartheta_2 a_{t-2} + \dots + \vartheta_q a_{t-q} \quad (2.19)$$

Le modèle peut être écrit sous une forme réduite comme suit :

$$\varphi(B)\tilde{Z}_t = \vartheta(B)a_t \quad (2.20)$$

où  $B$  est un opérateur de retard. Le modèle ARMA contient  $(p + q + 2)$  paramètres à estimer, il est généralement suffisant de prendre une valeur pour  $p$  et  $q$  inférieure ou égale à 2. La combinaison des deux méthodes AR et MA nous permet de représenter des séries temporelles avec moins de paramètres qu'un modèle AR.

#### 2.4.1.2.4 Le modèle ARIMA(p,q,d)

Dans la pratique, les séries temporelles ne sont pas stationnaires, contrairement aux modèles cités plus haut. Cependant, un ARMA modèle peut être modifié pour les séries non stationnaires, qu'on appelle ARIMA pour "Autoregressive Integrated Moving Average". Le terme "Integrated" représente le fait que le modèle est produit suite à une somme du processus ARMA.

Une convention a été établie pour définir la structure d'un modèle ARIMA(p,q,d), où  $p$  représente le nombre de paramètres autorégressifs,  $q$  le paramètre de la moyenne mobile et  $d$  est le nombre de passes à réaliser. Pour résumer, le modèle ARIMA revient à appliquer un modèle ARMA sur le processus différentiel. L'équation d'un modèle ARIMA(p,q,d) est donc donnée par l'équation 2.21 :

$$\varphi(B)\Delta_d \tilde{Z}_t = \vartheta(B)a_t \quad (2.21)$$

où  $\varphi$  et  $\vartheta$  sont deux polynômes de degrés respectifs  $p$  et  $q$ .

## 2.4.2 Méthodes de prévision computationnelles

Nous allons voir dans cette section une liste non exhaustive des méthodes de modélisation et prévision de l'intelligence artificielle.

### 2.4.2.1 Réseaux de Neurones artificiels

Les réseaux de neurones artificiels sont massivement parallèles, ils représentent une nouvelle technologie computationnelle construite par analogie avec le système de traitement d'informations humain. Depuis le début de leurs utilisations, les réseaux de neurones ont prouvé leur efficacité pour : l'analyse des signaux, l'extraction de caractéristiques, la classification, la reconnaissance des formes, etc. et cela est dû à leurs capacités d'apprentissage et de généralisation depuis les données d'observations.

L'élément de base d'un réseau de neurones est le neurone, dont l'utilisation n'a pas changé depuis sa création par Widrow et Hoff [17] en 1960. Ces derniers ont considéré le perceptron comme un élément adaptatif ressemblant à un neurone biologique. (Figure 2.1).

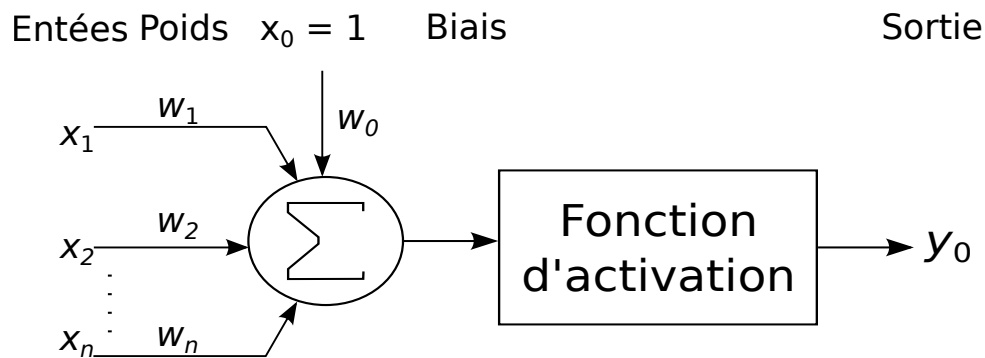


Figure 2.1 – Un perceptron.

Le perceptron est constitué d' :

- Un noyau, nommé perceptron, représenté par une somme et une fonction d'activation.
- Un signal multiple d'entrées  $x_i$  qui sont connectées via ce qu'on appelle des poids  $w_i$ .
- Un signal de sortie  $y_d$ .

Une entrée additionnelle  $w_0$  qui est nommée biais, qui fait office de seuil. Le signal de sortie est défini dans l'équation 2.22

$$y_0 = f\left(\sum_{i=1}^n w_i x_i + w_0\right) \quad (2.22)$$

Le choix de la fonction d'activation est crucial pour le perceptron. Block en 1962 [18] a sélectionné la fonction binaire mais elle fût remplacée par la fonction d'activa-

tion sigmoïdale (Equation 2.23) et cette dernière est considérée comme un approximateur universel.

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (2.23)$$

Le perceptron peut apprendre par le biais d'un algorithme d'apprentissage basé sur un ensemble de données. Pendant la phase d'apprentissage, le perceptron ajuste ses poids en accord avec les données présentées en entrée. Widrow et Hoff [17] en 1960 ont proposé l'algorithme règle delta. Puis, dans les années 90 Rumelhart [19] a popularisé l'algorithme rétro-propagation du gradient qui est le plus utilisé de nos jours.

Même si les algorithmes d'apprentissage sont en majorité efficaces en terme de rapidité de convergence, un perceptron seul ne peut apprendre suffisamment pour résoudre des problèmes complexes (non linéaires) et ceci est dû à sa structure simpliste. Afin de palier à ce problème, une structure multicouche nommée MultiLayer Perceptron (MLP) ou perceptron multicouches (Figure 2.2), qui, en plus des couches d'entrée et de sortie, contient une couche cachée insérée entre les deux.

Plusieurs structures de réseaux de neurones sont utilisées pour la prévision des séries temporelles :

- Les réseaux de neurones à base radiale ou Radial Basis Function (RBF).
- Les réseaux de neurones récurrents.
- Les réseaux de neurones probabilistes.

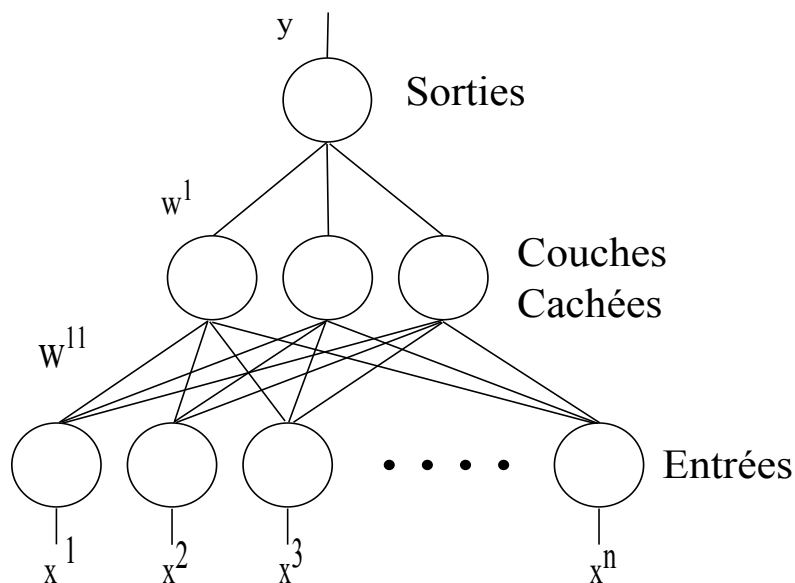


Figure 2.2 – Le Perceptron Multi-couche.

### 2.4.2.2 La logique floue

Le terme « flou » a été introduit par Zadeh en 1965 dans son article sur les ensembles flous. Ainsi, la discipline mathématique nommée « logique floue » qui est basée sur les ensembles flous a été présentée. La logique proposée vise à soutenir la présentation et l'examen de notions inexacts ou imprécises par des ensembles flous. Dans la logique floue, un objet ou une variable définie dans un domaine peut partiellement appartenir à plusieurs ensembles flous simultanément.

La principale idée dans la théorie des ensembles flous est la fonction d'appartenance, qui sert à mesurer le degré auquel plusieurs éléments flous appartiennent à un certain ensemble flou. Par conséquent, un vrai ou faux n'est pas nécessaire pour définir cette propriété, mais peut être défini comme partiellement vrai.

En utilisant la fonction d'appartenance  $\mu$ , on peut définir un ensemble flou  $F$  dans un univers de définition  $U$  comme suit :

$$\mu_F(x) : U \longrightarrow [0, 1] \quad (2.24)$$

Cette fonction est juste un mapping de l'univers de définition  $U$  vers l'intervalle  $[0, 1]$  et  $\mu_F(x)$  représente le degré d'appartenance de  $x$  à l'ensemble flou  $F$ . Le concept de la fonction d'appartenance permet à n'importe quel élément de l'univers d'appartenir partiellement à deux ensembles flous mais avec des degrés différents.

Les ensembles flous, les variables linguistiques et la logique floue ont une relation directe avec les systèmes de logique floue. Ces derniers sont uniques dans le sens où ils peuvent traiter simultanément des données numériques et des connaissances linguistiques. D'un point de vue mathématique, un système de logique floue est un mapping non linéaire d'un vecteur de caractéristiques (données) vers une sortie scalaire.

Le diagramme d'un système flou est représenté dans la figure 2.3 dont on peut déduire que le système de logique floue prend une valeur  $X$ , ensuite, il la convertit en un degré d'appartenance à un ou plusieurs ensembles flous définis. On appelle cette étape la « fuzzification ». Par la suite, un système d'inférence floue va utiliser des règles floues, préalablement définies, pour produire une sortie floue qui sera finalement « défuzzifier » pour obtenir une sortie  $Y$ .

N'importe quel système peut être décrit par une collection de règles « IF-THEN », aussi connues sous le nom de règle floue. Les systèmes de logique floue sont basés sur



ces règles qui sont généralement sous la forme : « IF préposition THEN conséquence », et sur leur système d'inférence. Dans la pratique, les systèmes d'inférence floue suivants ont été le plus souvent utilisés et ont le plus souvent fait l'objet d'études théoriques :

- Le système d'inférence de type Mamdani.
- Le système d'inférence de type Takagi-Sugeno.
- Le système flou relationnel (Pedrycz).

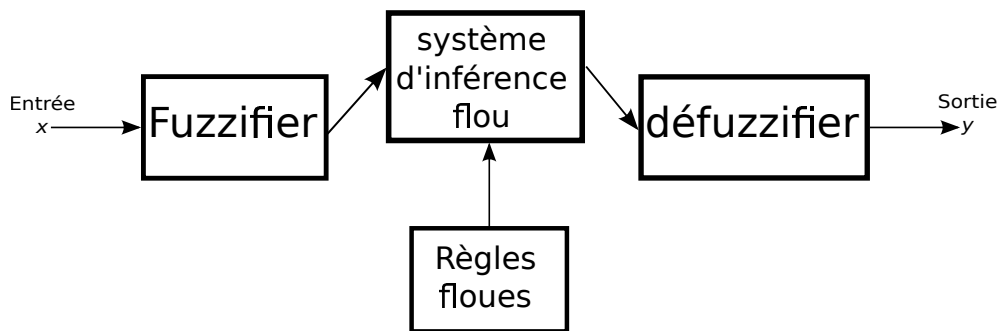


Figure 2.3 – Le diagramme d'un Système Flou.

## 2.5 Horizons de prévision

La prévision de la charge est l'estimation des valeurs futures. Elle aide les industries énergétiques dans la prise de décisions en ce qui concerne le planning, la production électrique, et le développement des infrastructures. La prévision de la charge peut être classée en trois catégories : [20]

- **La prévision à long terme** ; qui est utilisée pour prédire jusqu'à une période de 50 ans cela dans le but de faciliter la planification de l'expansion du réseau énergétique.
- **La prévision à moyen terme** ; est quant à elle utilisée pour prédire les pics hebdomadaires, mensuels et annuels jusqu'à une période d'une année, de sorte qu'une planification opérationnelle efficace puisse être effectuée.
- **La prévision le court terme** ; permet une prévision allant jusqu'à une semaine dans le but de minimiser le fonctionnement quotidien et les frais d'envoi sur le réseau.

En plus de ces trois catégories de prévisions de la charge, une quatrième catégorie a émergé ces dernières années ; le moyen-long terme (Figure 2.4), dans laquelle l'unité de temps considérée est le mois et l'année, elle se place naturellement entre le moyen et le long terme. Le moyen-long terme est très important pour la gestion et planification du ré-

seau, spécialement pour les pays en développement où la demande énergétique augmente avec un taux de croissance dynamique. Aussi, la prévision à moyen-long terme comme celle à long terme est toujours difficile à prévoir et cela est dû aux facteurs affectant la charge électrique.

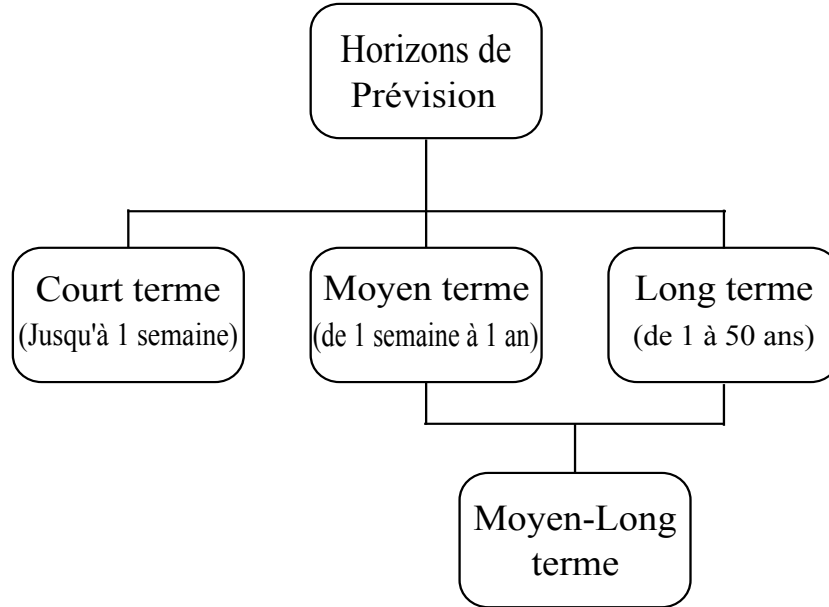


Figure 2.4 – Les différents horizons de prévision.

On considère trois secteurs (Figure 2.5) pour ce qui concerne la prévision de la charge électrique ; le secteur résidentiel qui a un taux de croissance annuel uniforme et des fluctuations saisonnières importantes. Ces dernières sont une conséquence de l'utilisation des appareils tels que : les climatiseurs et le chauffage. Le secteur commercial a aussi des fluctuations saisonnières identiques à ceux du secteur résidentiel. Et pour finir, le secteur industriel qui est un secteur de base qui a peu de variations dépendantes du climat.

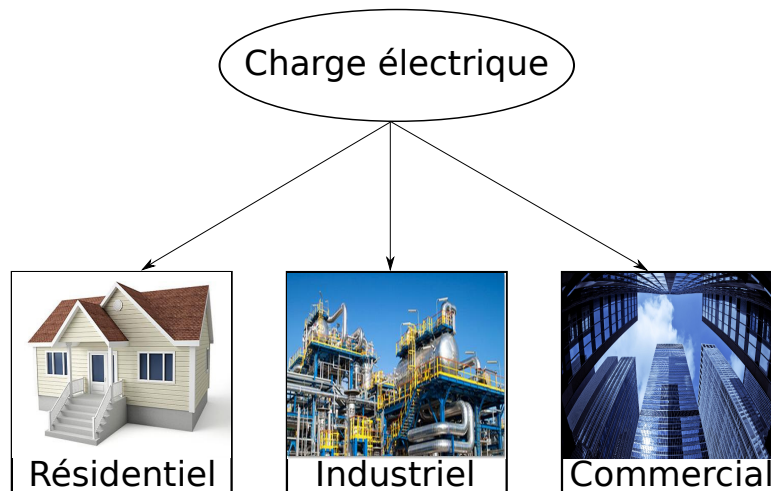


Figure 2.5 – Les secteurs concernés par la prévision de la charge électrique.

Chaque secteur a des besoins énergétiques spécifiques ainsi que des infrastructures

de transport et de distribution adaptées aux besoins du secteur. La prévision de chaque secteur sur les horizons mentionnés plus haut est possible en intégrant dans l'étude les caractéristiques des secteurs respectifs. Nous allons parler dans ce qui suit des travaux réalisés dans la littérature pour la prévision à moyen et à long terme de la charge électrique qu'elle soit par secteur ou globale, par région ou par pays.

### **2.5.1 Moyen terme**

Zhang et Wang [21] ont utilisé dans leur étude un Back Propagation Neural Network (BP-NN) avec une Analyse des Composantes Principales (ACP) afin de réduire le problème de précision et de généraliser les modèles de prévision. L'ACP, ayant la capacité d'éliminer la corrélation existante entre les variables, a non seulement été utilisée pour réduire le nombre de variables mais également pour garantir l'originalité des informations. En effet, elle a permis d'optimiser la structure du réseau de neurones. Ce dernier nommé PCA-BPNN a été utilisé pour prédire la charge électrique annuelle de certaines zones urbaines. Les auteurs ont expérimenté le modèle PCA-BPNN sur la charge électrique de 2003 à 2005 en tenant compte de la demande électrique et des facteurs l'influençant comme : la population, le produit national brut PNB, ainsi que le produit brut des industries primaires, secondaires et tertiaires. Ce modèle a donné une meilleure précision.

Les règles floues, obtenues via les algorithmes génétiques dans le Système de logique floue Takagi-Sugeno, et la création d'un modèle flou résument la méthode proposée par Zhang et Liu [22] pour la prévision de la charge à moyen-long terme. Le choix de cette méthode a été motivé par le fait que : les systèmes flous proposent de meilleures formes pour l'expression de règles, que la logique floue offre la possibilité d'un raisonnement logique équivalent à la pensée humaine, et enfin les algorithmes génétiques proposent une exploitation intelligente d'une recherche aléatoire dans un espace de recherches défini pour la résolution de problèmes. L'expérimentation proposée par Zhang et Liu a été réalisée sur les données du réseau électrique de Hebei en Chine et les résultats confirment que cette nouvelle combinaison est efficace. En effet, elle augmente la vitesse et la précision de la prévision de la charge électrique. Les auteurs sont restés dans la même idée en proposant une autre approche [23], qui est composée de règle floue et de réseau de neurones à ondelette (Wavelet Neural Network), prouvant ainsi que la précision et la vitesse ont été considérablement améliorées. Ces mêmes auteurs proposèrent une nouvelle méthode dans l'étude [22], basée sur les Wavelet Neural Network mais exempte de la logique floue contrairement à l'étude [23], qui a donné des résultats identiques à l'étude précédente.

Différents facteurs peuvent affecter la prévision de la charge électrique. Afin de les

déterminer, Yue et al. [24] proposèrent une méthode de clustering, basée sur les réseaux de neurones exploitant la logique connective et la théorie floue, qui peut être utilisée pour la prévision de la charge électrique à moyen long-terme. La combinaison de ces deux méthodes augmente la vitesse du réseau de neurones, ce dernier fût combiné avec un apprentissage compétitif. En effet, le modèle proposé a été créé et testé sur 20 années consécutives (De 1980 à 2000) dans la province de Guangdong et dont l'algorithme prend en considération les facteurs influant sur la variation de la charge comme : l'industrie lourde, l'industrie légère et la population. La première phase consiste en une classification de l'historique des données d'entrée, de 1980 à 2000, où le nombre de classes doit être connu à l'avance. Les facteurs de 2001 à 2005 sont utilisés, pour la phase de tests, pour prévoir la charge électrique des 5 années. Avec une erreur minimale de 1.568% et maximale de 1.766% en 2005, l'algorithme de classification proposé par les auteurs surpasse en termes de précision les méthodes traditionnelles, à savoir : le lissage exponentiel et les réseaux de neurones artificiels.

Les réseaux de neurones artificiels sont désavantagés pour ce qui est de la prévision de la charge électrique à moyen-long terme. C'est ainsi que Yu et Zhang [25] présentèrent une nouvelle méthode de structures variables pour les réseaux de neurones. Cette méthode, palliant aux défauts intrinsèques de rapidité d'apprentissage et difficultés de détermination d'une structure rationnelle d'un réseau de neurones, propose un algorithme dans lequel la sortie de chaque couche du réseau de neurones est calculée, ensuite l'erreur de chaque neurone de la couche est calculée. Tous les échantillons seront traités de la même manière après avoir obtenu l'erreur globale et vérifier si les paramètres initiaux ont été atteints ; sinon la valeur ajustée de chaque poids est calculée et la modification du réseau est faite en conséquence. Cette méthode, ayant prouvée son efficacité sur les données de charge de la ville de Hebei en Chine, a été comparée aux réseaux de neurones classiques et au Wavelet Neural Network. Les résultats prédisent un avenir prometteur dans le domaine de la prévision de la charge, vue l'augmentation de la rapidité et la précision du réseau de neurones.

Plusieurs études se sont inspirées des avantages de la théorie Grey pour la prévision de la charge électrique. Ainsi, Zhao et Nui [26] ont proposé une approche basée sur la méthode Grey qu'ils ont vérifié en utilisant l'historique des données de la consommation énergétique de 1999 à 2005 de la ville de Pékin. Dans leur étude, ils ont amélioré le model Grey ou en anglais Grey Model (GM) via les méthodes « Smoothly Processing » et « Background Value Processing ». Après que la charge électrique entre dans le modèle, cette dernière passe par 2 étapes : La première consiste à effectuer le lissage des données et à obtenir la meilleure valeur pour la méthode "Background Value". Quant à la seconde, elle se base sur la prévision de la charge électrique. Les résultats obtenus ont démontré que le nouveau modèle présente un avantage clair en termes de précision et atteint une

erreur moyenne relative de 1.19% en comparaison avec le modèle Grey classique qui, lui, arrive à une erreur de 1.39%.

Après avoir analysé les avantages et inconvénients de la prévision avec les modèles gris (Grey Model) et Least Squares Support Vector Machine (LS-SVM), Fuwei et Wei [27] proposèrent un nouveau modèle de prévision : « Grey Least Squares Support Vector Machine » ou « Grey LS-SVM », qui est un multi-modèle basé sur ces deux méthodes. Le modèle « Grey LS-SVM » fonctionne comme suit : dans le but de réduire l'effet de la distribution stochastique des données, la génération de l'accumulation ou Accumulative generating operation (AGO) est conduite via le modèle Grey, ensuite vient l'élaboration d'un modèle de prévision LS-SVM avec la nouvelle séquence. Finalement, les valeurs futures sont obtenues en inversant la fonction d'accumulation. L'expérimentation proposée par les auteurs a prouvé que la précision et la généralisation du modèle proposé est meilleure que les modèle Grey GM(1,1) ou les LS-SVM. C'est grâce à sa grande valeur théorique et son utilité pratique que le modèle « Grey LS-SVM » peut être appliqué à la prévision de la charge électrique.

Sun et Zhao [28] ont quant à eux proposé un modèle basé sur : les Support Vector Machine (SVM) et l'optimisation avec les colonies de fourmis. Ce modèle appelé MG-CACO est utilisé pour optimiser les paramètres du SVM et réduire la cécité de la sélection de paramètres en plus d'éliminer les lacunes de la sélection par test et l'erreur des paramètres. Les auteurs ont utilisé, dans leur étude, les données du National Statistical yearbook et Tianjon Statistical Yearbook et ils ont démontré que le modèle basé sur l'approche MG-CACO est plus performant pour la prévision de la charge à moyen-long terme. En effet, ce modèle a été comparé aux traditionnels SVM et a démontré son efficacité.

Nui et al. [29] ont proposé une méthode basée sur les Least Squares Support Vector Machine (LS-SVM) et Particle Swarm Optimization (PSO) algorithmes. L'algorithme PSO leur a servi à optimiser les paramètres du LS-SVM. L'expérimentation, qui fût réalisée sur les données de la province de Guangxi en Chine de 1992 à 2007, a conclu que le modèle LS-SVM optimisé par la méthode d'inertie dynamique PSO, est plus performant et plus rapide que le modèle optimisé par PSO standard. La méthode et les résultats obtenus ont surpassé tous les modèles concurrents.

Une combinaison nommée SFGM a été proposée par Wang et al. [30] combinant entre le modèle d'équation différentielle à une variable de premier ordre issue de la théorie Grey et la méthode de fluctuations saisonnières des séries temporelles. Cette approche puise ses avantages dans la puissance de prévision des modèles Grey et des méthodes des séries temporelles. Une autre variante du modèle SFGM a été proposée par les auteurs, nommée : APL-SFGM « Adaptive Parameter Learning » ou paramètre d'apprentissage

adaptatif pour l'amélioration de la précision du modèle de base. Les auteurs ont évalué les modèles sur la demande électrique de l'Afrique du sud de l'année 2002 à 2010. Après une comparaison avec plusieurs modèles dont les GM(1,1), il s'est révélé que ceux proposés par les auteurs ont montré de meilleures performances en termes de précision.

Le modèle proposé par Wang et al. [31], pour prédire la consommation d'énergie hydroélectrique de la Chine, est basé sur une décomposition saisonnière ou Seasonal Decomposition (SD) avec l'algorithme X12-ARIMA et Least Squares Support Vector Regression (LS-SVR). Le modèle nommé SD(M)-LSSVR-LSSVR décompose premièrement l'énergie hydroélectrique en 3 composantes, qui sont par la suite prédites via 3 modèles LSSVRs. Ensuite, un dernier LSSVR combine les 3 prédictions des composantes pour obtenir la prévision finale de la consommation de l'énergie hydroélectrique de la Chine. Les auteurs ont prouvé l'efficacité de la méthode X12-ARIMA en la comparant avec d'autres méthodes de décomposition comme Ratio Moving Average (RMA) et Empirical Mode Decomposition (EMD). L'étude montre que les performances du modèle SD(M)-LSSVR-LSSVR diminuent quand l'horizon de prévision est grand, mais le modèle est le plus performant par rapport aux autres modèles de cette étude.

### 2.5.2 Long terme

La prévision à long terme de la charge électrique est nécessaire pour le planning des opérations futures des compagnies de distribution énergétique. Pour cette raison, plusieurs techniques ont été utilisées pour répondre aux besoins de ces compagnies. Un nombre important de variables est pris en considération dans le processus de prévision, c'est pour cela que Ding [32] a utilisé une méthode basée sur les arbres décisionnels pour l'extraction de règles « if-then » et par ce fait, clarifiant les relations existantes entre la charge électrique et les variables exogènes relatives. L'auteur a utilisé, dans la phase d'expérimentation, les données économiques de 1978 à 2004 collectées du « Official Statistical Almanac » et les données de la charge électrique via Energy Management System (EMS). Les données de 1978 à 2001 ont quant à elles été utilisées pour construire l'arbre de décisions avec l'algorithme Iterative Dichotomiser 3 (ID3), 14 facteurs ont été pris en considération par l'auteur. La phase de test a montré que l'arbre décisionnel donne une erreur moyenne de 1.45 qui reste le meilleur résultat en comparaison avec d'autres modèles tel que : le modèle Grey, l'arbre de décisions traditionnel, la régression linéaire multiple, etc.

Delevad et al. [33] proposèrent une étude dans laquelle la charge électrique iranienne a été prédite à long terme grâce à un réseau de neurones artificiels. Les auteurs ont pris en considération les facteurs économiques où la température n'a pas été prise en compte dans

leur étude pour prévenir d'éventuels bruits, 12 facteurs ont été sélectionnés tels que : le Produit Intérieur Brut (P.I.B), le Produit National Brut (PNB), le prix du pétrole iranien, etc. Ces facteurs représentent les entrées du réseau de neurones, avec 19 neurones en couche cachée et une sortie. La phase d'apprentissage s'est faite en utilisant l'algorithme de rétro-propagation du gradient. Les données de 1990 à 2003 ont été utilisées pour la phase d'apprentissage, quant à celles de 2004 à 2006, elles l'ont été pour la phase de test ; suite à laquelle le réseau de neurones arrive à une erreur Mean Squared Error (MSE) de 0.95% pour l'année 2006, ce qui a été jugé satisfaisant par les auteurs.

Du et al. [34] ont proposé un modèle basé sur les Relevance Vector Machine (RVM), elles-mêmes basées sur le framework d'apprentissage : « Probabilistic Bayesian Learning », l'objectif des auteurs était de prévoir les pics de la charge électrique future à long terme. Le choix des RVM comme modèle de prévision a été justifié par les avantages de ce dernier : prévision probabiliste, estimation automatique des paramètres dit de « nuisance » et la facilité d'utilisation de peu de fonctions basiques. Pour la phase d'expérimentation, les auteurs ont utilisé la charge électrique de 1990 à 2000 prise de « Shandong Electric Power Corporation », et des facteurs de contribution ou variables exogènes telle que : le PIB, la population, le prix de l'électricité, etc. comme entrée du modèle RVM. Les mois ont été considérés pour augmenter le nombre d'échantillons des données, l'évaluation du modèle s'est faite sur les années 2001 à 2005 et après comparaison avec les SVM, le RVM s'est révélé plus efficace en terme de précision et d'efficacité. Une prévision allant jusqu'en 2020 a été réalisée par les auteurs.

Shi et al. [35] ont suggéré dans leur étude un modèle très robuste, basé sur les réseaux de neurones génétiques qui sont capables de s'adapter et d'apprendre à partir de données non linéaires. Le modèle proposé combine l'historique de la charge électrique et l'avis des experts ; ce qui lui donne la possibilité de changer en fonction des circonstances. Les auteurs ont utilisé les données nationales de 16 facteurs affectant la charge, le modèle ajuste en fonction des entrées des experts de l'industrie jusqu'à ce que l'erreur voulue soit atteinte. Le programme MATLAB a été utilisé comme outil pour l'expérimentation ; qui s'est faite sur plusieurs années. Ainsi, les pics sont prédits jusqu'à l'année 2012 et les auteurs jugent que les résultats sont satisfaisants en considérant plusieurs types de calculs d'erreurs.

Les deux approches, à savoir : la régression linéaire et les réseaux de neurones artificiels, sont introduites par Daneshi et al. [6] pour prédire les pics mensuels de la charge. Ils ont réalisé une simulation en utilisant les pics mensuels de la charge et la température en Angleterre de 1980 à 1999, les résultats obtenus montrent que les réseaux de neurones sont plus performants que la régression linéaire parce qu'ils ont l'avantage de gérer les systèmes non linéaires, et que le bruit dans les données n'affecte pas les performances du

réseau.

Carpinteiro et al. [36] proposèrent un réseau de neurones hybride hiérarchique pour le problème de la prévision de la charge électrique à long terme. En effet, ils ont utilisé deux réseaux de neurones de type Self-Organizing Map (SOM) l'un sur l'autre et une couche de perceptron. Les auteurs ont comparé le modèle proposé avec un perceptron multicouche ou MLP, l'expérimentation s'est faite sur les données électriques Nord-Américaines de l'année 1985 à 1991. Les résultats de cette étude démontrent que le modèle présenté est supérieur en termes de performance, et cela est justifié par sa topologie hiérarchique. Ce nouveau modèle est considéré comme très prometteur.

Pour prédire la charge électrique annuelle iranienne, Ghanbani et al. [37] ont employé 3 différentes méthodes : les systèmes d'inférence adaptatifs neuro-flou ou Adaptive NeuroFuzzy Inference System (ANFIS), la régression linéaire et enfin la régression log-linéaire. Comme la prévision de la charge peut être affectée par plusieurs facteurs (économiques, météorologiques, sociologiques, etc.), les auteurs ont sélectionné 21 facteurs socio-économiques et énergétiques dont ils ont extrait 5 facteurs influant sur la charge électrique iranienne, tous les modèles prennent en considération les 5 facteurs. Ensuite, un prétraitement des données est fait via le filtre « Savitzky-Golay Smoothing », l'architecture du système ANFIS est composée de 5 couches, elles-mêmes composées de nœuds simples ou adaptatifs (changement de paramètres pendant l'apprentissage). L'outil MATLAB est utilisé pour l'implémentation de la structure proposée. Les 3 modèles sont évalués sur 3 mesures statistiques Root Mean Square Error (RMSE), Mean Absolute Error (MAE) et Mean Absolute Percentage Error (MAPE). Dans la phase d'évaluation ; 7 années (de 2001 à 2007) ont été prédites, et les résultats indiquent que le modèle ANFIS approxime de manière satisfaisante la charge électrique réelle avec une erreur MAPE de 2.788% par rapport aux modèles basés sur la régression linéaire.

Deux méthodes ont été présentées par Aslan et al. [38] pour la prévision à long terme du pic de la charge électrique : la première est une approche statistique représentée par la régression des moindres carrés et la seconde par les Artificial Neural Network (ANN) et plus précisément par un MultiLayer Perceptron (MLP) comportant 3 couches : une couche d'entrée pour les pics de la charge, température mensuelle et population ainsi que 12 neurones pour la couche cachée et un neurone pour la couche de sortie. La fonction sigmoïde et l'algorithme d'apprentissage Levenberg Marquardt (LM) sont utilisés dans cette étude. L'expérimentation prend la population et la température de l'année 2000 à 2007 de Kutahya en Turquie comme variables exogènes. Les résultats montrent que l'approche des réseaux de neurones produit de meilleurs résultats pour l'année 2008 avec une erreur MAPE de 3.53%, et confirme que les méthodes de l'intelligence artificielle sont plus efficaces que celles issues des statistiques pour le problème de la prévision de la



charge à long terme.

Chen [39] a quant à lui présenté une approche neuro-floue collaborative pour la prévision à long terme. Dans le but d'obtenir de meilleures performances, les intersections floues ou Fuzzy Intersection (FI) sont considérées dans le but d'agréger la prévision floue de la consommation d'énergie annuelle de différents experts. Par la suite, le résultat est défuzzifié avec une Radial Basis Function (RBF). La communication s'interrompt quand les performances de la prévision deviennent négligeables. L'approche collaborative neuro-floue proposée par l'auteur est unique et aucune autre étude ne prend en considération ce type de modèles. La consommation énergétique taiwanaise de 1945 à 2008 a été utilisée pour la phase d'expérimentation. Ainsi, trois experts ont été choisis pour donner leur avis sur la prévision énergétique annuelle en répondant à 3 questions. Les résultats montrent que l'approche collaborative augmente la précision de la prévision de la charge électrique à long terme en comparaison avec d'autres modèles : MA, Exponential Smoothing (ES), Back Propagation Network (BPN), ARIMA, Tanaka's Fuzzy Linear Regression (FLR) et Peters's FLR approches.

Une méthode basée sur : les réseaux de neurones floues ou Fuzzy Neural Network (FNN), la rétro-propagation du gradient ou Back Propagation (BP) et l'optimisation par les essaims particuliers ou Particle Swarm Optimization (PSO) est présentée par Wai et al. [40] dans le but de réaliser une prévision intelligente de la charge électrique à long terme d'un campus de la ville de Taïwan. Les auteurs ont réalisé 5 structures de prévision ; une structure neuronale avec rétro-propagation (NN-BP), un réseau neuro floue (FNN-BP), un FNN-BP couplé avec un algorithme d'optimisation : « Varied Learning Rate » la structure est nommé FNN-BP-V, l'optimisation du FNN via les PSO (FNN-PSO) et pour finir un modèle PSO. Les résultats montrent clairement que le modèle PSO est plus performant que tous les autres proposés dans cette étude pour la prévision de 2009 et 2010 avec une moyenne MAPE de 6.45%.

Les Generalized Regression Neural Network (GRNN) ont été développés par Specht [41] démontrant ainsi leur efficacité pour la résolution des problèmes non linéaires, mais peu d'études ont été réalisées dans le cadre de la prévision de la charge électrique. Li et al. [42] ont quant à eux utilisé le modèle GRNN dans le but de tester son efficacité dans la prévision de la charge à long terme. En effet, ils ont utilisé Fruit Fly Optimisation Algorithme (FOA) pour sélectionner automatiquement les paramètres du modèle, ce qui a augmenté la précision de prévision des modèles GRNN. L'algorithme FOA est une technique d'optimisation évolutionnaire, les auteurs ont expérimenté le modèle de cette étude avec la consommation électrique de la ville de Pékin en Chine et une comparaison avec différents modèles a été menée, tels que : le réseau de neurones récurrent ou Recurent Neural Network (RNN), PSOGRNN, SALSSVM et OLS\_LR. Le modèle FOAGRNN

s'est révélé être plus efficace en terme de temps d'apprentissage et a obtenu le score le plus bas en ce qui concerne l'erreur de prévision MSE, ce qui démontre l'efficacité de la combinaison FOA et GRNN.

Un modèle neuro-flou a été proposé par Maraloo et al. [43] en utilisant l'algorithme d'apprentissage Locally Linear Model Tree (LoLiMoT). Le modèle Locally Linear NeuroFuzzy (LLFN) est composé de 3 couches : la première est une couche d'entrée qui a la fonction de diviser l'espace de données en plusieurs sous espaces linéaires avec une fonction de validité floue, une seconde couche cachée, et un neurone dans la couche de sortie. Les auteurs ont utilisé les données journalières du 01/01/1985 au 10/12/1992 issue de la compagnie électrique Nord-Américaine. Suite à la phase d'apprentissage, une prévision journalière et mensuelle a été réalisée dans cette étude, et le modèle LoLiMoT, avec une erreur journalière et mensuelle MAPE de 2.60% et 13.34% respectivement, s'est révélé être le plus performant en comparaison avec Hierarchical Hybrid Neural Model (HHNM) et MLP, cela étant dû à la flexibilité et à l'interprétabilité de la structure du modèle proposé par les auteurs.

Hongyan et al. [44] présentèrent une étude basée sur les modèles prédictifs Grey pour la prévision de la demande électrique à long terme. La consommation annuelle de la Chine de 1983 à 2001 ainsi que 5 autres variables exogènes ont été utilisées comme entrées des différents modèles. La prévision de 2002 à 2006 a été utilisée pour la phase d'expérimentation ; suite à laquelle les auteurs ont conclu que le modèle Grey est insuffisant pour prédire un système électrique volatile. Pour pallier à l'insuffisance de ce dernier, un modèle de réseau de neurones Grey-RBF a été présenté dans cette étude ; l'algorithme proposé étant plus robuste et fiable en comparaison avec les modèles Grey et les réseaux de neurones de type RBF.

## 2.6 Conclusion

Dans ce chapitre nous avons présenté et défini une série chronologique, l'intérêt d'une série quelle qu'elle soit est d'étudier son comportement passé en utilisant des outils graphiques et/ou mathématiques afin de bien prévoir son futur. Nous avons montré les différents composants et caractéristiques de toute série qui soit indexée par le temps, et les outils proposés par les scientifiques pour décomposer et obtenir ces composantes.

La deuxième partie de ce chapitre est consacrée à la présentation des méthodes et concepts utilisés dans la littérature pour prédire une série temporelle, deux grandes familles existent ; les méthodes issues de la statistique et celle de l'intelligence artificielle,

ceci est une liste de méthodes non exhaustive, dans laquelle nous avons essayé de présenter les méthodes les plus importantes et surtout celles touchées dans les études que nous avons proposées dans cette thèse. On a conclu le chapitre par un état de l'art des travaux qui traitent de la prévision de la charge électrique dans le moyen et le long terme.

## **Chapitre 3**

# **Spécificité de la charge électrique algérienne**

## 3.1 Introduction

L'histoire de la science ne change pas même si les ingénieurs et les scientifiques ajoutent des découvertes à un rythme sans cesse croissant. Peut-être que ce phénomène est la raison pour laquelle dans les cours de sciences on enseigne principalement l'histoire de la science, plutôt que la science réelle. Les progrès technologiques incroyables du XX<sup>e</sup> siècle et la série de découvertes scientifiques qui ont eu lieu plus tôt ont été masqués par les différentes colonisations, par les pays européens, qui ont eu lieu parallèlement.

La production d'électricité est le terme générique pour convertir une forme cinétique (géothermique, hydraulique, solaire, éolienne, etc.) ou énergie potentielle (charbon, gaz naturel, uranium enrichi, eau au-dessus d'un barrage) en énergie électrique. L'énergie cinétique est déjà en mouvement tandis que l'énergie potentielle est enfermée dans les combustibles fossiles ou nucléaires, ou dans l'eau d'un réservoir.

Les termes ; électricité et puissance sont souvent utilisés de façon interchangeable, comme l'est aussi courant et même "jus". Un courant électrique est un déplacement d'ensemble de porteurs de charges électriques, généralement des électrons, au sein d'un matériau conducteur. Ces déplacements sont imposés par l'action de la force électromagnétique, dont l'interaction avec la matière est le fondement de l'électricité. La propagation de l'influx électrique peut servir à son utilisation à distance comme source d'énergie, comme vecteur énergétique. Ce fût sa principale utilisation au début de l'ère industrielle.

A l'époque de Thomas Edison le système d'éclairage 110/220 V à courant continu était utilisé. Les clients devaient être dans un rayon d'environ 2.3 Km des dynamos. Cela signifie que les villes ont été électrifiées avant les zones rurales. En France, la Thomson-Houston Compagnie a profité de la proximité des cours d'eau pour que l'énergie hydroélectrique soit développée pour alimenter les villes. Aujourd'hui, les centrales économiques à turbine à gaz à cycle combiné sont construites et exploitées, et les lignes Ultra-Haute Tension (UHT) sont utilisées pour transmettre, efficacement, l'énergie à travers des centaines de kilomètres, cela donne aux usagés un confort notable comparé aux traditionnels systèmes de génération d'électricité.

De nos jours l'électricité est la source d'énergie la plus utilisée dans le monde, toutes les industries et le quotidien de la population y sont strictement dépendants. Dans ce chapitre nous allons aborder l'histoire de l'électricité de Thales à Tesla, tous les inventeurs qui ont contribué à la découverte et au développement de l'électricité tel qu'on la connaît aujourd'hui. En fonction de la source d'énergie utilisée, différentes méthodes ont vu le jour pour transformer cette énergie en électricité : des barrages d'eau et turbines à vapeur

jusqu'aux centrales nucléaires. Une description non exhaustive, de quelques types de générateurs d'électricité, sera présentée dans ce chapitre, ainsi que des notions (facteurs) relatives aux motivations qui poussent vers la construction de nouveaux dispositifs.

L'Algérie, étant un pays riche en hydrocarbures et en gaz, est l'une des puissances économiques Africaines. Dans ce chapitre, une mise en évidence des atouts et des faiblesses de ce pays, sa situation géographique, les ressources disponibles dans le sous-sol algérien, ainsi que la présentation de la spécificité de l'électricité en Algérie, en abordant le parc électrique et la distribution de l'électricité sur le territoire national. Pour conclure le chapitre, nous aborderons les variables économiques (appelées exogènes) qui ont un rapport avec l'étude de l'électricité ; tel que le P.I.B et la clientèle d'une tension donnée, ce qui permettra une meilleure compréhension des facteurs influants sur la production et la consommation électrique dans une région donnée.

## **3.2 Découverte de l'électricité**

L'électricité est la découverte la plus importante du XII siècle, elle est vite devenue indispensable dans la vie quotidienne ; nécessaire pour l'éclairage, les hôpitaux, les aéroports, l'industrie, etc.

Le mot électricité fût mentionné pour la première fois par le Dr William Gilbert en 1600 ap.J.C, il fût le physicien personnel d'Elizabeth I. Passionné par la découverte du magnétisme par Thales en 600 av.J.C en frottant une ambre, le Dr. Gilbert consacra 17 ans de sa vie à l'étude du magnétisme, et publia un livre intitulé « De Magnete » dans lequel il nomma la propriété du magnétisme de l'ambre « électricité ». En grec, ambre est nommé « Elektron ». En examinant les théories du Dr. Gilbert, Von Guericke maire de la ville de Magdeburg (Allemagne) (1602-86) inventa un aspirateur à air en magnétisant une boule de soufre. Il appela son mécanisme « Machine électrostatique » assurant que l'énergie produite est statique.

Stephen Grey (1670-1736) grâce à son ami Wehler a découvert pour la première fois comment déplacer l'électricité d'un point à un autre. En poursuivant ses recherches, Grey découvrit que certains matériaux sont de bons conducteurs électriques alors que d'autres non. Benjamin Franklin (1706-1790) est celui qui a fait une importante contribution dans le domaine de l'électricité, ce dernier a écrit une thèse intitulée : « Les similitudes entre la foudre et l'électricité » où il démontre que les éclairs étaient de simples décharges de nature électrique ; ses recherches conduisirent à l'invention du paratonnerre. Dans le domaine des recherches sur la nature de l'électricité, on lui doit par exemple des termes aussi

courants que « batterie », « positif », « négatif », « charge », « condensateur (condensateur) », etc.

Dans l'évolution de l'électricité, est venu par la suite Volta. En se basant sur les travaux du Dr. Luigi Galvani (1737-98), qui découvrit que le muscle animal se contracte grâce à des décharges électriques, Volta démontra qu'on pouvait créer de l'électricité chimiquement en inventant la pile. Dans les grandes découvertes scientifiques, notamment dans le domaine du magnétisme et de l'électricité, on peut également citer Andre Marie Ampère (1775-1836) qui découvrit l'électromagnétisme qui fût une avancée majeure qu'on utilise jusqu'à maintenant (sonnette électronique, téléphone). C'est cette dernière invention qui a permis à Morse en 1832 de créer le télégraphe et Graham Bell après lui, le téléphone. [45]

Dans l'histoire de l'électricité, l'année 1831 fût significative. En effet, au cours de cette année Michael Faraday (1791-1867) a découvert le processus d'induction électromagnétique ou la dynamo. Il partit de l'idée que si l'électricité produit du magnétisme, le magnétisme peut créer à son tour de l'électricité. Cette découverte révolutionna le domaine de l'industrie qui était à la recherche d'une source électrique plus puissante que la batterie voltaïque. Ce fût le début des turbines à vapeur ou des centrales thermoélectriques. Ensuite est venu, dans l'histoire, le plus grand duel au cœur de New York entre Nikola Tesla et Thomas Edison suite à la création de la lampe par Edison. Le 4 septembre 1882, l'Edison Electric Light Company fondait la première centrale électrique à charbon du monde : « La Pearl Street Station » à base de 6 dynamos Jumbo, pour produire du courant continu dans le quartier de Wall Street à Manhattan, d'une capacité de 1200 lampes pour éclairer 85 maisons, bureaux ou boutiques. Moins d'un an plus tard, d'autres centrales toujours plus puissantes éclaireraient plus de 430 immeubles New-Yorkais avec plus de 10000 ampoules. Plus tard, Londres a pu bénéficier du même progrès.

En 1884, Edison, fervent partisan du courant continu, se sépare de son employé Nikola Tesla, un des pionniers du courant alternatif qui peut être acheminé sur de plus longues distances comparé au courant continu ; grâce à l'utilisation de transformateurs électriques. Les deux hommes ne pouvant s'entendre, Edison a usé de ses relations afin de discréditer Tesla aux yeux de l'opinion publique mais ce dernier, se mettant alors au service de George Westinghouse qui a persuadé les industriels de s'équiper en courant alternatif, a gagné le duel avec son générateur de courant alternatif. Ce type de courant représente plus de 95% de l'électricité produite jusqu'à nos jours. [46]

### 3.3 Production électrique

L'électricité est la seule forme d'énergie simple à produire, transporter, utiliser et à contrôler. C'est également la forme d'énergie terminale pour la transmission et la distribution. La production d'électricité en grande quantité est effectuée dans des centrales électriques, et ces dernières peuvent être de différents types :

- Thermales.
- Nucléaires.
- Hydrauliques.
- Turbines à gaz.
- Géothermales.

Les centrales thermales, nucléaires et géothermales ont des similitudes dans leurs fonctionnements et structures. En effet, elles fonctionnent toutes grâce à la vapeur. Quant aux turbines à gaz, elles sont plus utilisées dans un court laps de temps (à court terme) pour répondre aux pics journaliers.

Contrairement aux autres centrales électriques, les centrales hydrauliques sont à usage multiple. En effet, en plus de la production énergétique, elles répondent aux problèmes ; de l'irrigation, du contrôle des flux, de la pêche, de l'afforestation, de la navigation, etc. Cependant, les centrales hydrauliques sont coûteuses et demandent beaucoup de temps pour leur construction. [1]

Les centrales thermales génèrent plus de 80% de l'électricité globale dans le monde. Les combustibles fossiles, à savoir, charbon, essence et gaz naturel sont leurs sources d'énergie. La vapeur est le fluide qui actionne leurs turbines, elle est également requise dans l'industrie pour le processus de chauffage.

Pour répondre à la double demande, des centrales cogénératrices sont souvent installées. Deux paramètres sont à déterminer lors de la planification de la construction d'une centrale électrique, à savoir :

1. La charge de sortie totale à installer ( $kW_{inst}$ ) qui peut être déterminée par :
  - (a) L'estimation de la demande maximale ( $kW_{max}$ ).
  - (b) L'anticipation de la demande future.
  - (c) La capacité de réserve requise.
2. La taille de l'unité génératrice, qui dépend de :
  - (a) La variation de la charge (Courbe de charge) durant 24 heures (été, hiver, jour de semaine, vacances)



- (b) La capacité totale des unités connectées au réseau.
- (c) La période minimum du démarrage et de l'extinction des unités.
- (d) Le programme de maintenance.
- (e) L'efficacité de la centrale en comparaison à la taille de cette dernière.
- (f) Le prix et l'espace de demande par kW en comparaison avec la taille de la centrale.

La manière la plus utilisée pour estimer la charge maximum est la construction d'une hypothétique courbe de charge. La figure 3.1 montre les courbes de charge par heure de la région. A partir des courbes, on peut estimer la demande maximum. Dans ce qui suit nous allons définir quelques facteurs utilisés dans les industries électriques.

$$\text{Facteur de charge} = m = \frac{kWh_{\text{moy}} \text{ dans une année}}{kW_{\text{max}} * 8760} \quad (3.1)$$

Si une centrale électrique a un pic de 1000 MW et une moyenne annuelle de charge égale à 350 MW, le facteur de charge annuelle de cette centrale sera de 0.35, ce qui signifie qu'une grande partie de la capacité du dispositif n'est pas exploitée durant l'année. (Figure 3.2)

$$\begin{aligned} \text{Facteur de capacité} = n &= \frac{\text{La charge moyenne}}{\text{La capacité de la centrale}} \\ n &= \frac{kWh \text{ generé en une année}}{kW_{\text{inst}} * 24 * 365} \\ &= \frac{kWh_{\text{gen}}(\text{année})}{kW_{\text{inst}} * 8760} \end{aligned} \quad (3.2)$$

Si le pic de la charge et la capacité de la centrale sont égaux alors le facteur de charge et de capacité seront égaux. La différence entre le facteur de charge et de capacité est l'indication de la capacité de réserve.

$$\text{Facteur de réserve} = r = \frac{kW_{\text{inst}}}{kW_{\text{MAX}}} = \frac{m}{n} = \frac{\text{Facteur de charge}}{\text{Facteur de capacité}} \quad (3.3)$$

Chaque consommateur a une « charge connectée » qui est la somme de la consommation de tous les équipements et prises électriques dans son circuit. La demande maximum est la charge maximum qu'un consommateur utilise à n'importe quel moment. Elle est toujours inférieure ou égale à la charge connectée.

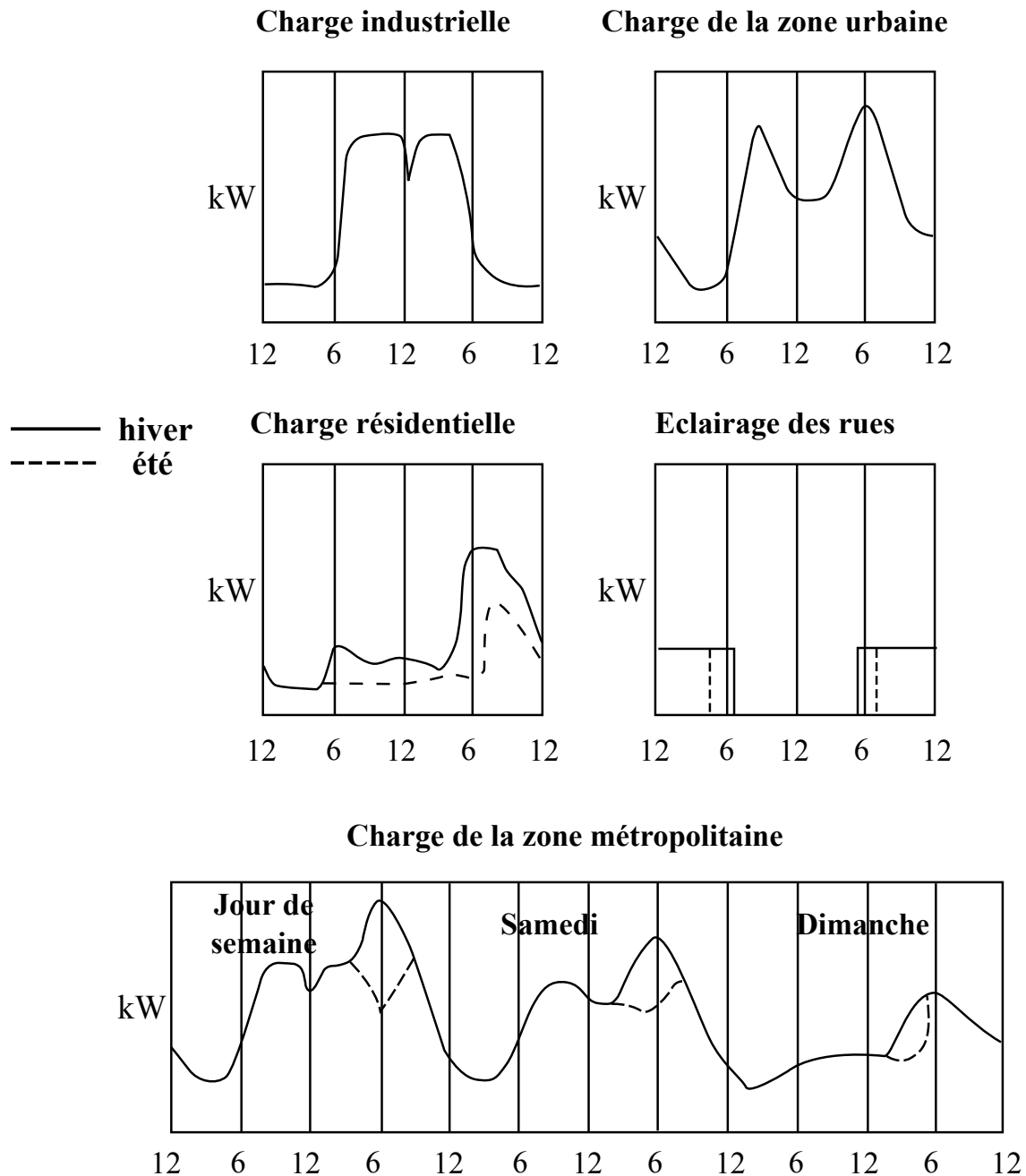


Figure 3.1 – Variation horaire typique de la demande énergétique de différents types de charge électrique et la courbe totale d'une zone métropolitaine.[1]

$$\text{Facteur de demande} = dem = \frac{\text{Demande maximum actuelle}}{\text{La demande totale connectée}} = \frac{kW_{MAX}}{kW_{CONN}} \quad (3.4)$$

Le temps de distribution de la demande maximum, pour le même type de consommateur, est mesuré par le terme « Facteur de diversité » qui est le ratio de la somme de la demande maximum de chaque individu sur la demande maximum de tout le système considéré.

$$\text{Facteur de diversité} = \text{div} = \frac{\text{Somme de la demande individuelle}}{\text{Le pic de charge du système}} \quad (3.5)$$

Le facteur de diversité sert à améliorer le facteur de charge et les opérations économiques d'une centrale électrique. Un facteur de charge élevé est généralement une indication d'une courbe de charge équilibrée avec peu de changements de charge. Ainsi, une grande valeur des trois facteurs : demande, charge et diversité est souhaitable pour les opérations économiques de la centrale et pour de l'électricité à bas prix.

$$\text{Facteur d'utilisation} = u = \frac{kWh_{gen}}{kW_{inst} * \text{total heures de fonctionnement}} \quad (3.6)$$

Le facteur d'utilisation est le ratio de l'énergie produite dans une période donnée sur le maximum d'énergie qui aurait pu être produite sur cette même période. Si le temps de fonctionnement de la centrale est d'une année ou 8760 heures, le facteur d'utilisation sera égal au facteur de capacité. Quand le facteur d'utilisation approche le 1 cela indique le besoin d'augmenter la capacité de la centrale. Cette dernière est toujours conçue pour avoir une capacité plus grande que le pic de charge pour répondre à la demande future. Une grande valeur du facteur d'utilisation indique que le fonctionnement de la centrale est efficace. Dans certains systèmes inter-connectés, le facteur d'utilisation de la centrale dépasse l'unité 1 ; indiquant que la charge produite excède la capacité de l'équipement, puisque ce dernier est conçu pour produire 10% à 20% de la capacité de charge. Les différents facteurs sont illustrés dans la figure 3.2. [2]

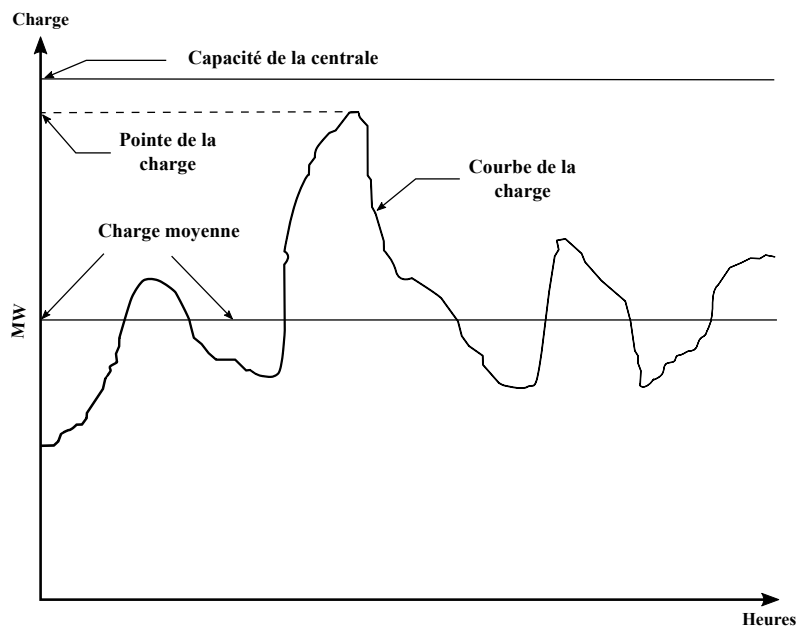


Figure 3.2 – Représentation des différents facteurs utilisés dans les industries énergétiques. [2]

## 3.4 Les différents types de centrales électriques

### 3.4.1 Par chute d'eau

Pour rendre un générateur électrique efficace et avec un coup bas, il faut une source d'énergie qui soit elle-même fiable et pas chère. La roue hydraulique représente donc une source fiable et peu coûteuse avec un mouvement rotatif. Le fonctionnement d'une roue hydraulique peut servir à faire marcher un générateur. Deux changements se produisent dans ce genre de système ; dans la roue hydraulique, l'écoulement spontané d'un haut vers un bas potentiel gravitationnel énergétique permet à une partie de l'énergie de produire un mouvement. Par conséquent, le générateur produit un mouvement non spontané de la charge électrique d'une basse vers une haute tension.

A l'époque médiévale les roues hydrauliques n'étaient pas vraiment efficaces dans la conversion de l'énergie cinétique en mouvement rotatif, pour la simple raison que l'eau possédait toujours son énergie cinétique après avoir quitté la roue. La recherche de plus d'efficacité dans les roues hydrauliques a commencé à l'époque de la renaissance. En effet, Leonardo da Vinci a constaté que le flux de l'eau traversant des lames courbées est plus efficace, ce type de dispositif a été nommé en premier « roue à impulsion » mais de nos jours il est plus connu sous le nom de « turbine ». La turbine est un appareil permettant la conversion de l'énergie cinétique d'un fluide en mouvement en une énergie cinétique rotative. Le liquide utilisé dans une turbine est appelé fluide actif qui est généralement de l'eau.

Dans une turbine à eau, le liquide s'écoule généralement à travers des tuyaux, plus la pression du fluide est élevée au contact de la turbine, plus la vitesse de rotation de la turbine est élevée. La vitesse de rotation d'une roue hydraulique est fortement liée à la vitesse du courant, mais ce n'est pas le cas des turbines qui peuvent atteindre une vitesse de 30 000 tours par minute. Le développement de la turbine au milieu du XIXe siècle était un progrès d'ingénierie majeur pour la production de rotations pour les générateurs électriques. Le processus de génération d'électricité via une turbine fonctionnant grâce à l'énergie cinétique de l'eau est appelé hydroélectrique.

L'installation hydroélectrique à Fully en Suisse, illustre certaines caractéristiques de ce type de centrales. Dans cette installation, l'eau est acheminée du lac via un tuyau de 5 Km en pente. Ainsi, la hauteur entre les deux extrémités du tuyau est pratiquement de deux (02) Km, cette pente raide fait jaillir l'eau à l'extrémité inférieure du tuyau avec une pression de 140 Kilomètre par centimètre carrée, et à une vitesse de 650 Km/h. L'eau à

cette pression et vitesse fait tourner les roues, de 3.5 mètre de diamètre, à une vitesse de 500 tours par minute.

La plus grande centrale hydroélectrique au monde nommée « Three Gorges », est située en Chine et a une capacité de 22 500 MW. Pour mettre cette centrale en perspective ; la puissance générée est approximativement vingt fois plus grande qu'une grande centrale à charbon ou nucléaire. A l'opposé des centrales de la taille de Fully et Three Gorges, il existe beaucoup de centrales hydroélectriques de petite taille : les petites centrales hydroélectriques génèrent moins de 10 MW, les mini centrales moins de 1 MW, quant aux microcentrales, elles génèrent moins de 100 kW. Toutes utilisent le même concept de turbines. Les petits projets ne demandent pas la construction de barrages pour former de grands réservoirs d'eau. Ce type de projets voit une popularité grandissante dans le monde. [47]

L'énergie hydraulique produit 16% de la demande mondiale en électricité, équivalent de 4% du total de la demande énergétique. La production électrique mondiale des centrales hydroélectriques est d'environ 3400 Milliards kW. Elles ont beaucoup d'avantages comparés aux combustibles fossiles :

- L'énergie hydraulique est renouvelable ce qui n'est pas le cas des énergies fossiles qui s'épuisent avec une forte probabilité de pénurie dans le futur.
- 90% de l'énergie cinétique de l'eau peut être convertie en énergie électrique, contrairement aux usines à charbon qui produisent 30-35% d'électricité.
- L'énergie hydraulique est mécaniquement fiable, puisque les seules parties qui bougent sont la turbine et le générateur, ce qui diminue les coûts de maintenance.
- Contrairement aux centrales à combustibles et nucléaires, il n'y a pas de perte de produit et pas de pollution, pas d'émission de dioxyde de carbone, en plus de la gratuité de la matière première (l'eau).
- Le plus grand avantage de l'énergie hydraulique c'est la possibilité d'augmenter ou de diminuer la puissance générée par la centrale et même d'arrêter la production énergétique.

L'hydroélectrique présente beaucoup d'avantages, notamment pour les coûts de production d'électricité et de maintenance très bas. Il faut savoir que 60 pays produisent plus de 50% de leurs besoins énergétiques grâce à l'énergie hydraulique ; la Norvège produit pratiquement 100% en utilisant l'énergie de l'eau. Mais une centrale hydroélectrique nécessite la construction de barrages d'eau, et ces derniers même avec leurs avantages, restent néfastes. Parmi les inconvénients des centrales hydroélectriques :

- L'eau va devenir une source de conflits entre les pays pour sa rareté. Surtout pour les rivières qui coulent dans plusieurs pays.

- Peu de rivières ont le courant nécessaire pour faire tourner une turbine et pouvoir générer l'électricité nécessaire, cela pourrait être corrigé par la construction de barrages.
- Le limon peut s'accumuler derrière les barrages et boucher leurs sorties, par conséquent, ces derniers peuvent déborder et inonder les terres agricoles, et la nature.
- Si le barrage rompt, il y a une grande probabilité que cela provoque une catastrophe, avec beaucoup de pertes humaines et de dommages matériels, comme cela s'est produit en Chine en 1975 ce qui a coûté la vie à 20 000 personnes.
- Les barrages produisent un impact majeur sur l'écosystème ; ils créent un réservoir d'eau à la place d'une rivière, les plantes aquatiques et poissons vivants dans les rivières, en général, ne survivent pas dans une eau stagnante. De nouvelles espèces plus adaptées à ces conditions s'installent dans les barrages ; parmi eux les moustiques. Les barrages peuvent également empêcher la migration des poissons.

### 3.4.2 Par énergie thermique

Vers la fin du 19<sup>ème</sup> siècle, les machines à vapeur se sont tellement développées, fournissant une autre alternative fiable à la génération de l'électricité. La turbine à vapeur tourne à grande vitesse sans générer de vibrations, ce qui a permis à cette invention d'être considérée dans la génération d'électricité dans les temps modernes.

L'ingénieur anglais Charles Parsons est le pionnier dans le développement des turbines à vapeur. Les turbines de Parsons sont constituées d'une séquence de lames fixes et mobiles, la vapeur circule dans le circuit de tuyauterie de telle sorte à rencontrer les lames de la roue dans un angle adéquat, la vapeur est admise au milieu et se propage parallèlement à l'axe dans les deux sens. Au contact de la pression atmosphérique, la vapeur génère une énergie cinétique à la turbine et le générateur électrique est directement connecté à l'axe de cette dernière. Le développement de cette machine pour une utilisation commerciale s'est fait rapidement. Ainsi, vers la fin du 19<sup>ème</sup> siècle la turbine à vapeur s'est répandue dans les centrales électriques.

La Turbine à vapeur est devenue le gold-standard pour ce qui est de la génération électrique dans le monde. Générer de l'électricité à partir d'une turbine à vapeur implique une configuration à haute pression et à température élevée de la vapeur pour la faire fonctionner. Le problème antérieur des centrales électriques était de trouver une source fiable et pas chère pour faire tourner le générateur, ou en d'autres termes ; trouver une technique fiable et pas chère de générer une vapeur à haute pression et à température élevée. [48]

Les centrales thermiques se répartissent en trois grandes catégories, selon la nature de leur source de chaleur :

- Centrales à flamme (charbon, fioul ou gaz).
- Centrales nucléaires.
- Centrales récupérant de la chaleur préexistante (solaire, géothermique...).

#### **3.4.2.1 Centrales à flamme**

Une centrale thermique à flamme utilise l'énergie fournie par la combustion d'un combustible (charbon, pétrole, gaz naturel, gaz issus de hauts-fourneaux). Cette combustion a lieu dans une chaudière (ou générateur de vapeur), et elle dégage une grande quantité de chaleur utilisée pour chauffer de l'eau dans la chaudière. On dispose alors de vapeur d'eau sous pression.

Cette vapeur sous pression fait tourner à grande vitesse une turbine qui entraîne elle-même un alternateur qui produit une tension alternative sinusoïdale. A la sortie de la turbine, la vapeur est refroidie pour se transformer en eau, puis renvoyée dans la chaudière. Le refroidissement de la vapeur issue de la turbine est confié à une réserve d'eau (cours d'eau) ou plus rarement à une tour de refroidissement analogue à celle d'une centrale nucléaire.

La combustion du charbon est l'une des solutions, ce qui fait de lui le combustible dominant pour la production d'électricité de nos jours ; avec plus de 40% de la production électrique mondiale issue de la combustion du charbon, et 60% de la production du charbon qui va vers le secteur électrique.

De nos jours, une centrale électrique à vapeur moyenne génère 400 MW en électricité et consomme environ 10 000 tonnes de charbon, certaines installations peuvent générer jusqu'à 1 000 MW, mais dans ce cas on envisage deux générateurs ou plus en parallèle. Ceci reste mieux que de s'appuyer sur une seule turbine et générateur. Les avantages des centrales thermiques à flamme sont :

- La possibilité d'arrêter la centrale lorsque le courant produit est supérieur au courant consommé ; en stoppant les fourneaux.
- Elles sont souvent plus puissantes que les autres types de centrales.
- L'emplacement de la centrale peut se faire à proximité d'un lieu où la consommation est forte. Il n'y a donc pas de contraintes géographiques.

Parmi les inconvénients de la centrale thermique à flamme :

- Les ressources s'épuisent, elles ne sont pas infinies.
- Les énergies fossiles entraînent une pollution de l'air plus importante.

- Ce type de centrales rend le pays dépendant aux énergies fossiles.
- Les pertes dans la plupart des centrales sont, malgré les avancées technologiques, plus importantes lors de la conversion de l'énergie thermique en énergie électrique par rapport à une centrale hydroélectrique.
- Ces centrales prennent beaucoup de temps lors du démarrage.

#### 3.4.2.2 Centrales nucléaires

L'énergie nucléaire est née dans le secret de guerre et plus tard avancée comme une source d'énergie électrique. Pendant la guerre froide et en raison de la Destruction Mutuelle Assurée (MAD), l'industrie de l'énergie nucléaire civile est restée inextricablement liée aux questions de sécurité nationale et militaire. En 1939, les physiciens Leo Szilard, Eugene Wigner, Edward Teller et le mathématicien Albert Einstein ont écrit une lettre informant le président Franklin Delano Roosevelt du potentiel de l'énergie nucléaire et ont proposé aux États-Unis de développer une arme atomique contre les nazis. Bien que Roosevelt ait ordonné qu'un effort soit fait pour développer une telle arme, des progrès significatifs n'ont été atteints qu'après l'entrée en guerre des États-Unis en 1942 avec la mise en place du projet d'ingénierie à Manhattan.

L'énergie nucléaire a commencé à se développer avec l'envie de construire une bombe atomique. Le premier réacteur utilisé pour maintenir une réaction nucléaire fût construit à l'Université de Chicago en 1942. Ce succès a conduit à la construction de réacteurs surgénérateurs à Hanford, Washington. Pour convertir le  $^{238}\text{U}$  isotope de l'uranium avec un design de fusil à canon simple jugé si intrinsèquement fiable qu'il n'a même pas été testé avant utilisation. La bombe larguée sur Nagasaki utilisait un réacteur de type  $^{239}\text{Pu}$  dont la conception d'implosion avancée a été testée à Alamogordo, Nouveau-Mexique.

Les bombes atomiques de la seconde guerre mondiale et les réacteurs actuels travaillent en divisant des atomes instables dans un processus appelé fission. Les bombes atomiques utilisées à la fin de la seconde guerre mondiale devaient être constituées de plus de 90% de  $^{235}\text{U}$  ou  $^{239}\text{Pu}$ . Par contre, le réacteur l'était d'environ 3% de  $^{235}\text{U}$  ou d'autres matières fissiles comme le plutonium ou du thorium. Dans des bombes atomiques, des charges explosives classiques forcent la masse du matériau fissile ensemble à une grande vitesse qu'une masse peut être formée pour un instant, avec une telle sortie des neutrons, il est immédiatement suivi d'une explosion thermonucléaire.

Une centrale nucléaire est constituée d'un ou plusieurs réacteurs nucléaires dont la puissance électrique varie de quelques mégawatts à plus de 1 500 mégawatts pour le réacteur soviétique de grande puissance High Power Channel-type Reactor RBMK. Selon



les promoteurs du futur réacteur européen Evolutionary Power Reactor (EPR), il devrait atteindre une puissance record de 1 600 mégawatts.

En 2006, 442 réacteurs fonctionnaient dans 31 pays différents dans le monde, soit un total de 370 gigawatts, produisant environ 17% de l'électricité mondiale. La catastrophe de Tchernobyl a conduit à plusieurs moratoires ; la baisse des prix du pétrole durant les années 1990 a renforcé cette tendance, conduisant à construire moins de nouveaux réacteurs dans le monde. Parallèlement, les centrales vieillissent : en 2006, la majorité des réacteurs avaient de 15 à 36 ans, et sept de 37 à 40 ans. [49]

Nous allons explorer les avantages de l'énergie nucléaire :

- L'énergie nucléaire a le moins d'effets sur la nature car elle ne libère pas de gaz comme le méthane et le dioxyde de carbone, qui sont les principaux "gaz à effet de serre". Il n'y a pas d'impact néfaste sur l'eau ou le sol en raison de l'utilisation de l'énergie nucléaire.
- L'énergie nucléaire produit de l'électricité très bon marché. Le coût de l'uranium, qui est utilisé comme combustible dans le processus, est faible. En outre, même si les frais de mise en place des centrales nucléaires sont modérément élevés, la charge de leur exécution est assez faible.
- Fiabilité des centrales nucléaires ; les experts estiment qu'avec le taux actuel de consommation d'uranium, nous avons assez d'uranium pour encore 70-80 ans.
- L'autre point d'intérêt principal de l'utilisation de l'énergie nucléaire est qu'elle est plus convaincante et plus compétente que les autres sources d'énergie. Un certain nombre d'innovations de l'énergie nucléaire ont fait un choix beaucoup plus facile que d'autres. Elles ont une densité d'énergie élevée par rapport aux combustibles fossiles, c'est la raison pour laquelle de nombreuses nations mettent beaucoup de temps et d'argent dans l'énergie nucléaire.
- L'énergie nucléaire n'est pas une ressource renouvelable, l'uranium, le combustible qui est utilisé pour produire de l'énergie nucléaire, est limité et ne peut être reproduit à la demande. Par contre, en utilisant surgénérateurs et réacteurs de fusion, nous pouvons produire un autre élément fissile ; le plutonium qui est produit par les réactions en chaîne. En outre, si nous savons comment contrôler la fusion atomique, nous pouvons avoir une énergie presque illimitée.

D'autre part, les inconvénients de l'énergie nucléaire sont :

- Un des plus grands problèmes est l'impact environnemental par rapport à l'uranium. Le processus d'extraction et le raffinage de l'uranium n'est pas un processus propre. Le transport de combustibles nucléaires à destination et à partir des centrales représente un risque de pollution.
- En règle générale, une centrale nucléaire crée 20 tonnes de combustibles nucléaires par an. Quand vous considérez chaque centrale nucléaire existante, vous

constaterez que ce chiffre passe à environ 2.000 tonnes par an. La plus grande partie de ces déchets transmet des radiations.

- Les déchets radioactifs produits peuvent poser des problèmes de santé graves sur les personnes ainsi que sur l'environnement. L'accident de Tchernobyl qui a eu lieu le 26 Avril 1986 à la centrale nucléaire de Tchernobyl en Ukraine a été le pire accident nucléaire de l'histoire. Plus récemment, le 11 Mars 2011 à 5h 46m 23s UTC, un séisme de magnitude 9 a touché le Japon. Cinquante et une minutes plus tard, un tsunami provoqué par le tremblement de terre aborde la côte orientale. La vague atteint une hauteur estimée à plus de 30 m par endroits, provoquant un accident nucléaire majeur à la centrale de Fukushima, classé au niveau 7, le plus élevé, de l'échelle internationale des événements nucléaires, au même degré de gravité que la catastrophe de Tchernobyl.
- Les centrales nucléaires laissent refroidir les déchets pour une période de temps considérable avant de les mélanger avec du verre et les ranger dans d'énormes structures solides qui sont refroidies. Ces déchets doivent être tenus, observés et regardés pour empêcher les matériaux de tomber entre de mauvaises mains. Ce qui est très onéreux.
- Tout comme d'autres sources de combustibles, l'uranium est également épuisable et n'est pas disponible partout. Il est assez cher à extraire, affiner et transporter.

### **3.4.2.3 Autres centrales**

Comme indiqué précédemment, une source fiable de travail mécanique rotatoire est nécessaire pour faire fonctionner un générateur électrique. Ce travail est fourni par une turbine. Le fluide qui fait tourner la turbine peut être de l'eau dans les centrales hydroélectriques, la vapeur d'eau dans les combustibles fossiles ou les centrales nucléaires, et les gaz de combustion dans les moteurs à combustion. Comme le vent représente un fluide (air) en mouvement, et possède une énergie cinétique, il pourrait également être utilisé dans une turbine. Nous utilisons le terme éolien pour faire référence à une machine qui utilise l'énergie du vent spécifiquement pour fournir du travail à un générateur d'électricité. Le terme moulin à vent est généralement limité à des dispositifs qui utilisent le vent pour d'autres applications de travail mécanique, telles que le pompage de l'eau ou la production de farine.

À la fin du XIXe siècle, la technologie de production électrique a été élaborée. L'utilisation d'éoliennes pour produire de l'électricité date de la fin des années 1880. Les éoliennes faisaient fonctionner un petit générateur pour produire de l'électricité pour recharger les batteries ou pour faire fonctionner les appareils électriques dans les fermes.

La première installation d'une éolienne pour recharger les batteries pour l'éclairage électrique était dans une maison à Marykirk, en Écosse. Un demi-siècle plus tard, il y avait des centaines de milliers d'exemplaires de ce dispositif opérant dans les États-Unis. Ces applications de l'énergie éolienne, pour la production d'électricité locale, a effectivement pris fin après la propagation des réseaux électriques qui ont amené l'électricité à partir de centrales génératrices.

Dans le monde entier, l'industrie éolienne ne cesse de croître. La Chine et les États-Unis sont les leaders en termes de capacité de production installée. Les installations chinoises et américaines sont à la fois sur terre et en mer. Pour la production en mer, les pays européens sont collectivement au premier plan. La Grande Bretagne possède les plus grandes stations d'éoliennes en mer, l'une à 370 MW et une autre à environ 300 MW. La capacité européenne totale est d'environ 2400 MW. Même si environ plus de 20% de l'approvisionnement en électricité des nations se fait grâce au vent, il y a un risque perçu de pénuries occasionnelles d'approvisionnement en électricité en raison de la nature intermittente du vent. Cependant, certains pays, notamment le Danemark, ont dépassé cette limite avec succès.

L'énergie solaire partage beaucoup des avantages de l'énergie hydraulique et éolienne, notamment dans le fait qu'il n'y a aucun coût pour le « carburant », qu'elle soit certainement inépuisable, et qu'il n'y a pratiquement pas de pollution. En outre, l'énergie solaire peut être utilisée dans une zone où l'ensoleillement est abondant, alors que l'hydroélectricité est limitée aux sites avec source d'eau appropriée, et enfin, l'énergie éolienne est mieux dans les zones où le vent souffle souvent. L'énergie solaire peut être utilisée par tout pays ou région, ce qui élimine la plupart des problèmes géopolitiques concernés par les ressources inégalement réparties comme le pétrole.

Chaque méthode, de production d'électricité à grande échelle discutée jusqu'à présent, utilise l'énergie cinétique d'un fluide pour entraîner une turbine. Les énergies fossiles et nucléaires utilisent des grands appareils, technologiquement complexes pour faire bouillir l'eau. Mais il y a un moyen de convertir l'énergie du soleil directement en électricité, cette approche est appelée génération solaire directe, ou l'énergie photovoltaïque.

L'effet photovoltaïque se produit lorsque la lumière touche deux matériaux différents et produit de l'énergie électrique. Certaines combinaisons de matériaux réagissent à l'énergie sous la lumière du soleil. Ensemble, ces matériaux composent une cellule photovoltaïque ou une cellule solaire. Le matériau clé dans la plupart des cellules photovoltaïques est le silicium.

Une installation photovoltaïque typique d'une maison, demande trois composants principaux : le réseau de modules de cellules solaires, un ensemble de batteries pour

fournir à la fois une production régulière d'électricité et un moyen de stockage d'énergie, et un dispositif pour transformer le courant continu en courant alternatif ; qui est le plus utilisé.

L'énergie solaire est disponible partout sur terre et peut être recueillie facilement avec des appareils portables. Cet avantage minimise les problèmes de transport de carburant ou de transmission de l'électricité du point de production au point d'utilisation. L'énergie solaire est non polluante et propre. Aucune exploitation minière ou de forage n'est nécessaire. L'avantage le plus important de l'énergie solaire est qu'elle soit pratiquement inépuisable. La source d'énergie -le soleil- ne peut être détenue et monopolisée comme le sont devenus les combustibles fossiles à cause de la répartition géographique inégale des réserves.

### **3.4.3 Par gaz naturel**

Les centrales à turbine à gaz, sont soit en cycle simple soit en cycle combiné. La turbine du cycle simple fait passer les gaz à travers un compresseur, combustion et enfin la turbine. Par contre les turbines à cycle combiné capturent la chaleur des échappements de gaz et l'utilisent pour faire tourner un autre générateur, chauffer de l'eau ou générer de la vapeur pour une application industrielle.

Les turbines à gaz sont en pleine expansion surtout dans les nouvelles constructions, pour la simple raison qu'elles sont efficaces. Le carburant est la première dépense dans toutes les centrales, par contre les turbines à gaz permettent de gagner de l'énergie potentielle du carburant avec peu de perte. Les centrales à gaz et plus particulièrement les turbines à cycle combiné ne sont pas chères à construire en comparaison avec les centrales à charbon et les centrales nucléaires.

La production électrique est impliquée dans 35% des émissions de carbone dans le monde. Le charbon produit 87% des émissions de carbone, par contre la combustion de gaz qui représente 15% de la génération électrique, produit quant à elle 9% des émissions de carbone. Un carburant propre et efficace implique donc peu d'émission.

Les turbines à gaz ont été placées pour la génération d'électricité dans le milieu des années 1900, cette technologie a vu une expansion pendant la seconde guerre mondiale à travers les programmes de développement des moteurs d'avions des alliés, et spécifiquement en 1943 quand Westinghouse a testé le premier turboréacteur américain. Après la guerre, la première centrale électrique à turbine à gaz, dans le cadre d'une utilisation commerciale, fût installée en 1949 par Oklahoma gas and electric compagnie à l'île Belle.

Des progrès technologiques ont été réalisés en dépit de la loi sur l'utilisation du carburant en 1978, qui a mandaté l'arrêt de l'utilisation du gaz naturel comme un carburant pour la génération électrique à cause de sa pénurie. Après la fin de cette législation, vers la fin des années 1980, les fabricants ont construit de meilleures turbines.

Les turbines à gaz fournissent environ 25 000 MW chaque année, ce qui représente 35% de la capacité, en construction, des centrales à base de combustibles fossiles. Les unités à cycle simple génèrent jusqu'à 150 MW avec un rendement d'environ 35%, tandis que les turbines à cycle combiné produisent plus de 200 MW avec un rendement qui approche les 60%, ce qui représente le meilleur rendement en comparaison avec les autres centrales à combustibles.

La cogénération, principe des turbines à cycle combiné permet la récupération de la chaleur du générateur électrique pour faire tourner un second générateur ou pour faire du travail ; pour faire bouillir l'eau, faire chauffer les immeubles, etc. Elle permet de maximiser l'efficacité et de réduire la quantité de certains polluants atmosphériques. Inversement, les systèmes utilisant les pertes de chaleur des industries pour générer de l'électricité sont appelés cycle creux. Les centrales perdent 2/3 de leur sortie thermique.

Vers la fin de l'année 1998, l'administration de l'information de l'énergie des Etats Unis d'Amérique suggérait que la génération de l'électricité mondiale issue de la cogénération était environ de 200 GW ou environ 6.5% de la production mondiale, 6% en Amérique, 60% en Suède et 2.5% en France. Les installations cogénératrices utilisent le plus souvent des turbines à gaz.

### 3.5 Présentation de l'Algérie

L'Algérie ; grand pays d'Afrique du Nord occupe une place stratégique d'un point de vue géographique. En effet, l'Algérie partage ses frontières terrestres avec sept (07) pays dont : la Tunisie et la Lybie à l'Est, le Maroc, le Sahara Occidental et la Mauritanie à l'Ouest, une large façade maritime de 1622 Km donnant sur la mer méditerranée. Et enfin, le Mali et le Niger au Sud. (Figure 3.3)

L'Algérie est le plus grand pays du continent africain avec 2 380 000  $km^2$  de superficie. Il est divisé dans le sens Nord-Sud en trois (03) grandes zones climatiques, dictant ainsi la répartition des populations :

- Au Nord : climat méditerranéen, étendu sur une centaine de kilomètres, caractérisé par un été chaud et sec et un hiver doux et tempéré permettant ainsi d'y dévelop-



Figure 3.3 – Situation géographique de l’Algérie.[3]

per l’agriculture. Les principales villes appartenant à ce climat sont : Alger (la capitale), Oran, Constantine et Annaba.

- Les zones montagneuses avec l’Atlas Tellien, les Aurès et l’Atlas Saharien : le climat y est semi-aride. Ce sont des zones de reliefs permettant l’élevage des caprins et des ovins.
- Enfin, tout au Sud on retrouve le Sahara : zone totalement aride couvrant 5/6<sup>eme</sup> du territoire algérien. (Figure 3.4)

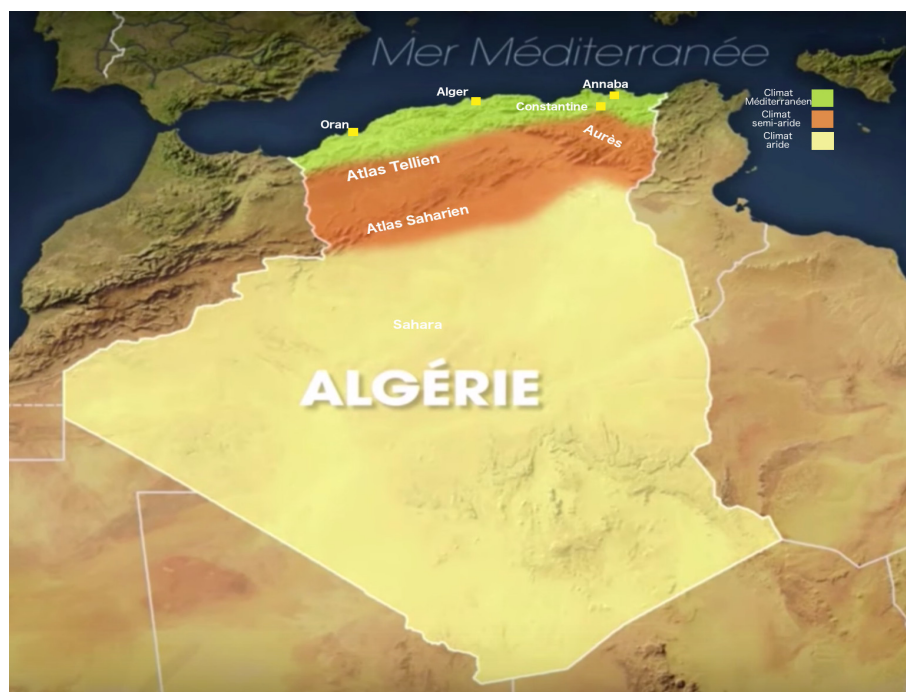


Figure 3.4 – Les zones climatiques algériennes.[3]

L'Algérie comptait plus de 39,5 millions d'habitants en 2015. Si on se confère à la densité des populations, on peut distinguer que la très large majorité (plus de 80% de la population) vit au Nord et se concentre surtout sur les côtes.

L'Algérie était autrefois un pays agricole. Aujourd'hui, la part de l'agriculture dans la formation du Produit Intérieur Brut ou P.I.B est de 10% et ce secteur emploierait environ 11% de la population active. Le pays s'est axé principalement sur la production de céréales, agrumes, olives, sur la culture des vignes et enfin sur l'élevage de bétail.

L'essentiel du P.I.B est dû au secteur secondaire ou industriel à un taux de 47% et qui emploierait 26% de la population active. Quant au secteur tertiaire ; il contribue à un taux de 43% au Produit Intérieur Brut et emploie environ 63% de la population active. Le P.I.B de l'Algérie dépend donc fortement du secteur secondaire et ceci grâce à son sous-sol très riche en phosphate près de la frontière tunisienne, en minerai de fer le long de la frontière marocaine et sur la côte. Des mines d'or, de diamant, d'uranium et d'étain dans le massif montagneux du Hoggar, mais aussi du zinc, du plomb et du charbon. (Figure 3.5)

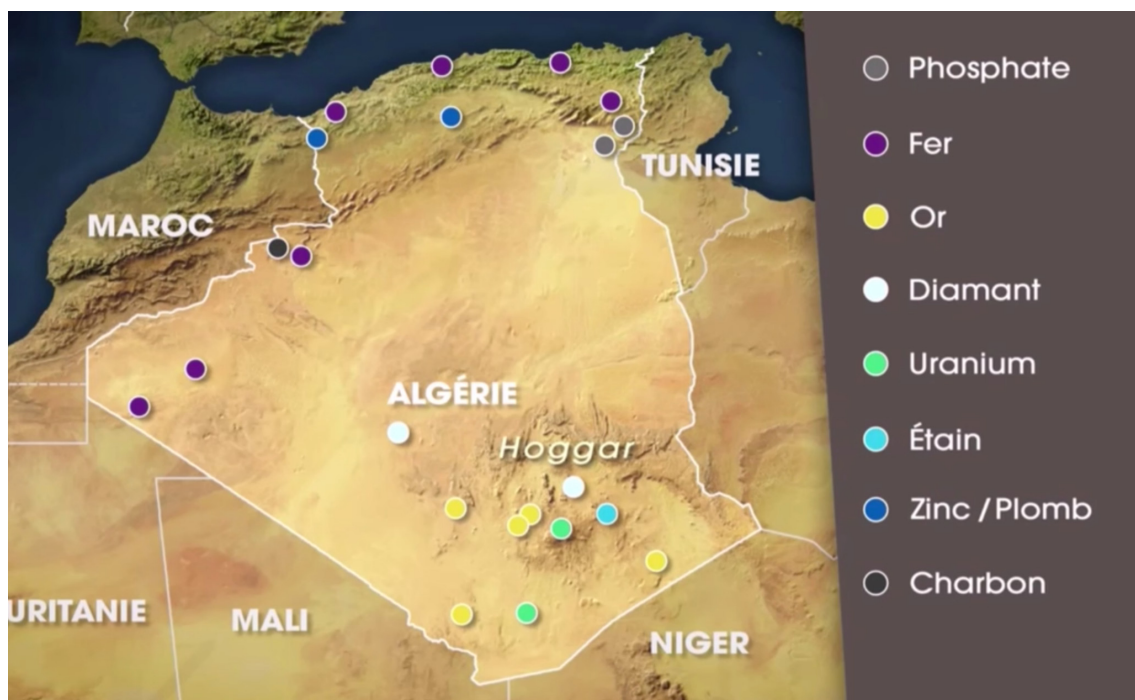


Figure 3.5 – Disposition des gisements du sous-sol algérien.[3]

Toutes ces richesses sont secondaires comparées à la richesse principale du pays et qui est représentée par les hydrocarbures (gaz et pétrole). Les principaux gisements algériens d'hydrocarbures se situent dans le Sud : dans le bassin de Reggane, d'Edjeleh à In Amenas où a eu lieu le premier jaillissement de pétrole en Algérie en 1955 – puis celui de Hassi Berkine et de Hassi R'Mel (Figure 3.6), faisant de l'Algérie le neuvième (9<sup>ème</sup>) producteur de gaz dans le monde en 2014 et le dix-huitième (18<sup>ème</sup>) producteur de pétrole. Du point de vue des réserves, l'Algérie occupe la seizième (16<sup>ème</sup>) place pour les réserves



de pétrole et la dixième (10<sup>ème</sup>) pour le gaz. Au total, 95.4% des recettes de l'exportation algérienne provient des hydrocarbures. L'union européenne achète 34% de la production de pétrole et 60% de la production de gaz algérien ce qui en fait son premier partenaire économique.

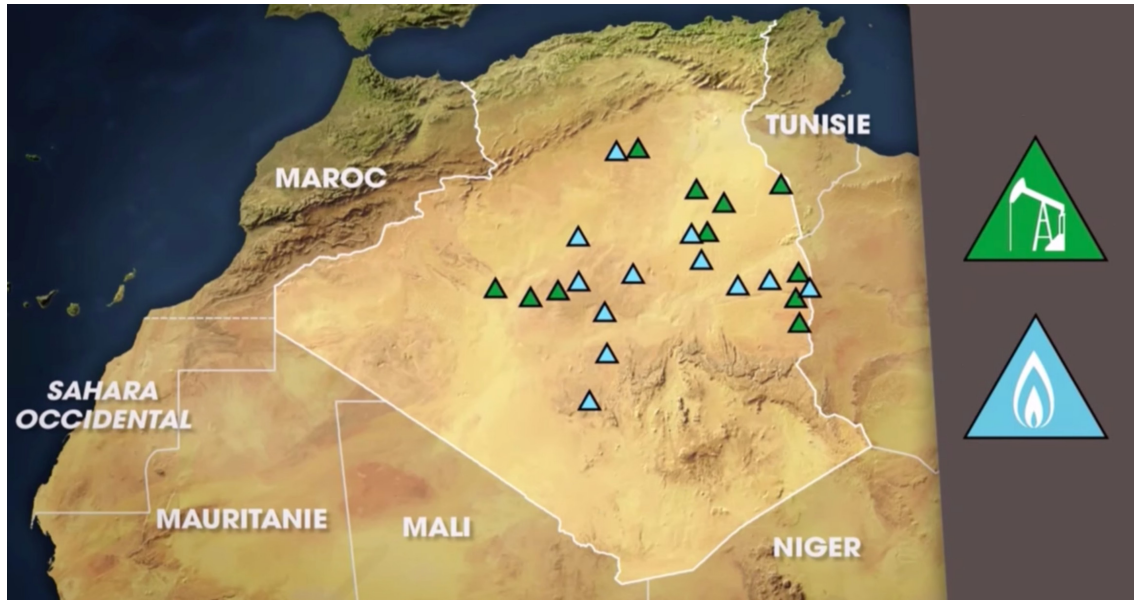


Figure 3.6 – Les principaux gisements algériens d'hydrocarbures.[3]

Depuis 2014 la conjoncture a changé, car on observe trois phénomènes : les exportations sont en baisse suite à la réduction de la demande européenne du fait de la stagnation économique, mais aussi du fait du ralentissement de l'économie chinoise. De ce fait, les recettes algériennes des exportations d'hydrocarbures sont passées de 32 Milliards de dollars dans les six premiers mois de l'année 2014 à 18 Milliards pendant la même période en 2015. L'autre phénomène est que le baril de pétrole est passé de 112 \$ en Juin 2014 à 45 \$ en Novembre 2015, et enfin les grands gisements historiques du pays sont de moins en moins productifs. Depuis, l'Algérie puise dans le fond des régulations des recettes (les réserves de change) qui sont passées de 179 Milliards de dollars en Décembre 2014 à 160 Milliards en Juin 2015 et la balance commerciale est déficitaire depuis la rentrée de l'année 2015 car le pays importe pratiquement tout ce qu'il consomme.

La figure 3.7 montre le planisphère des bassins de gaz et de pétrole de roche-mère (appelé à tort pétrole et gaz de schiste) et on voit que les premières ressources récupérables techniquement sont en Chine 32000 Milliards  $m^3$ , les deuxièmes sont en Argentine avec 23000 Milliards  $m^3$  et finalement l'Algérie détiendrait les troisièmes réserves mondiales soit 20000 Milliards  $m^3$  ce qui représente quatre fois plus que ses réserves de gaz naturel connues. La figure 3.8 montre les principaux bassins de roche-mère en Algérie. Face à une telle disponibilité de ces ressources, le gouvernement a lancé un programme de forage dans le Sahara, ceci peut être retenu comme une option qui viendrait à terme augmenter



la production des hydrocarbures et donc des recettes du pays.

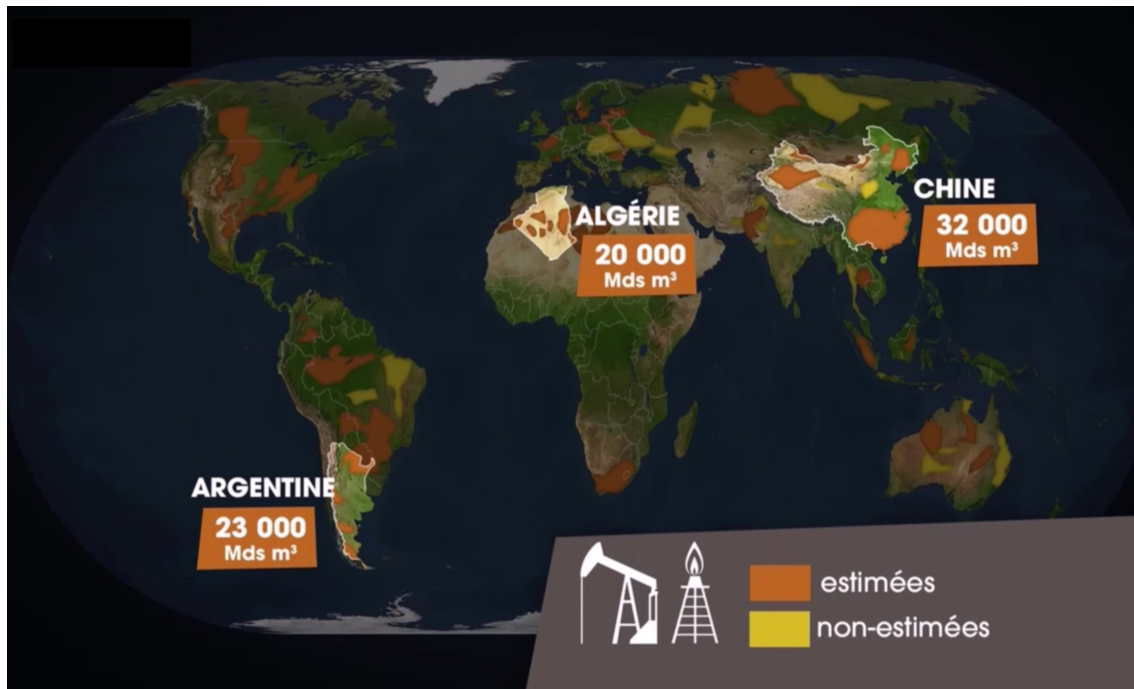


Figure 3.7 – Le planisphère des bassins de gaz et de pétrole de roche-mère.[3]

Il est important de noter que l'Algérie est un pays en pénurie hydrique, c'est-à-dire que l'eau disponible est peu abondante comme le montre la figure 3.9 qui nous indique les risques sur la disponibilité de l'eau. On y remarque qu'une grande partie du territoire algérien est en risque maximal. Selon l'ONU le seuil de pénurie est fixé à  $1000m^3$  / habitant/ an en moyenne, or en moyenne, le volume d'eau disponible en Algérie par habitant et par an est de  $600m^3$ .

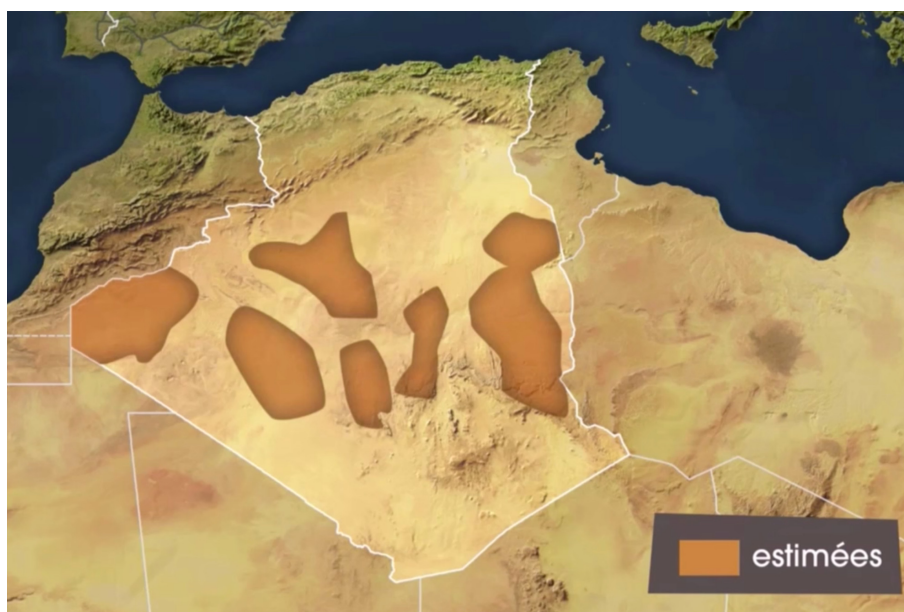


Figure 3.8 – Les principaux bassins de roche-mère en Algérie.[3]

En conclusion, la principale technologie d'extraction de la roche-mère connue de nos jours est la fragmentation hydraulique qui requière donc une énorme quantité d'eau à injecter sous pression avec des adjuvants, qui entrainerait donc un haut risque de pollution des nappes phréatiques, afin de faire éclater la roche.

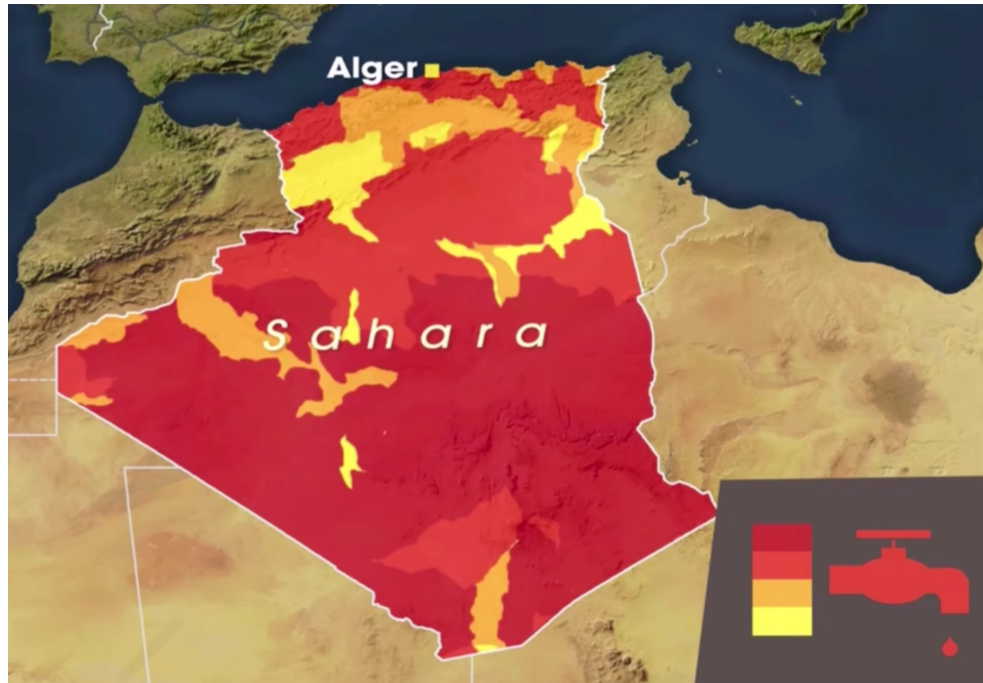


Figure 3.9 – Les risques sur la disponibilité de l'eau.[3]

### 3.6 Spécificité de la charge électrique algérienne

Dès son indépendance en 1962. L'Algérie avait opté pour le développement du secteur de l'énergie dans le cadre d'une politique nationale visant au développement des infrastructures électriques et gazières. Cette politique prévoyait l'accès de la population à l'électricité et au gaz naturel comme une priorité absolue pour l'amélioration de la qualité de vie du citoyen et de la situation économique du pays.

La charte nationale de 1976, annonça la volonté de la généralisation de l'électrification des ménages à travers tout le territoire national. Ces dernières années, la demande en électricité a connu une évolution importante et particulièrement en période estivale, atteignant des pics de consommation importants. Cette forte augmentation de la demande est une conséquence directe du changement des habitudes du consommateur, de l'amélioration de sa qualité de vie, ainsi que la pulsion donnée aux secteurs économiques et industriels.

Conscient de cet enjeu et de l'importance de l'accès à l'électricité pour les citoyens,

le secteur s'est fixé comme priorité de développer tous les axes permettant de garantir la couverture à long terme des besoins en électricité et en gaz du pays, notamment par la diversification des sources d'énergie, le développement du parc de production électrique et des infrastructures de transport et de distribution de l'électricité et du gaz.

L'entreprise nationale Sonelgaz détenait le monopole sur la production, le transport, la distribution de l'électricité ainsi que le transport et la distribution du gaz. Sonelgaz en tant qu'entreprise publique au terme du monopole conféré par l'état, assumait seule la mission de service public dans le domaine de la distribution de l'électricité et du gaz. Suite à la promulgation de la nouvelle loi du 05 février 2002 relative à l'électricité et à la distribution du gaz par canalisations, un coup d'envoi pour la réorganisation du secteur et l'ouverture du marché électrique à la concurrence a été donné.

Cette loi a également permis la réorganisation de Sonelgaz qui s'est achevée en 2011, avec la création de la société holding « Sonelgaz » ainsi que l'ensemble de ses sociétés filiales. Elle a également permis la création de la commission de régulation de l'électricité et du gaz « CREG » installée en 2005, qui encadre et régule les activités du secteur de l'électricité et gaz.

Sonelgaz est aujourd'hui érigée en groupe industriel composé de 39 filiales et 5 sociétés en participation. Des filiales métiers sont chargées des activités de production, de transport et de la distribution de l'électricité et du gaz par canalisations.

### **3.6.1 Parc de Production National**

Les impératifs de préservation de l'environnement imposent l'utilisation du gaz naturel comme énergie primaire pour la production d'électricité, par rapport aux autres énergies fossiles jugées polluantes, sachant que la ressource du gaz est largement disponible en Algérie. Mais la préoccupation environnementale exige aussi de développer d'autres énergies dites propres et renouvelables comme l'énergie solaire et éolienne, dont les gisements sont disponibles et abondants au sud du pays.

Les perspectives de relance de l'économie se traduiront également par un impact certain sur le plan socio-économique et par conséquent, une augmentation notable de la consommation d'électricité, notamment en saison estivale. Aussi, pour répondre à ce besoin en électricité sans cesse croissant, un programme ambitieux a été engagé pour le dédoublement des capacités de production conventionnelle à l'horizon 2018.

La stratégie d'exécution de ce programme de production de l'électricité convention-

nelle consiste en ; la mise en œuvre d'une série de projets industriels en partenariat basés sur la valorisation des ressources locales et des avantages comparatifs dont jouit notre pays. Il s'agit avant tout de développer une industrie nationale chargée d'assurer la fourniture des équipements destinés au secteur de l'énergie électrique. Des programmes indicatifs décennaux des besoins en moyens de production de l'électricité, sont élaborés par la Commission de Régulation de l'Electricité et du Gaz (CREG) chaque deux (02) ans et sont approuvés par Monsieur le Ministre de l'Energie. Ces programmes prévoient, en l'occurrence, les capacités de production à mettre en place sur une période décennale par région (Est, Ouest, Centre et Sud) et par filières selon deux scénarios d'évolution de la demande (moyen et fort).

Le dernier programme approuvé par Monsieur le Ministre est celui de la décennie 2014-2023.

### **3.6.2 Données par énergie**

Il existe plusieurs types d'énergies ; incluant l'énergie nucléaire, électrique, thermique, chimique et enfin l'énergie radiante. En plus de l'énergie gravitationnelle et cinétique qui se combinent pour créer de l'énergie mécanique.

L'énergie nucléaire produit de la chaleur par fission du noyau, ce qui est généré par un moteur thermique. Le nucléaire est la plus grande source mondiale d'énergie non productrice d'émission. Il existe deux processus dans l'énergie nucléaire ; la fission et fusion. Dans la fission, les noyaux d'Uranium ou Plutonium sont fondus ce qui libère de l'énergie. Dans la fusion, de l'énergie est libérée quand des petits noyaux fusionnent. La fission est utilisée dans les centrales nucléaires de nos jours, parce que la fusion ne peut être contrôlée. L'énergie nucléaire est utilisée pour chauffer les moteurs à vapeur.

L'électricité fait fonctionner la plus part des entreprises et maisons dans le monde, elle peut être stockée dans des batteries comme énergie chimique. L'électricité est le transfert de l'énergie d'un point à un autre, autrement dit c'est le flux d'énergies. Les atomes ont des électrons qui gravitent autour d'un noyau central, certains électrons sont libres et ont donc une faible attraction avec le noyau. Le mouvement de l'électron libre d'un atome a le pouvoir de créer un courant électrique.

L'énergie thermique est une énergie cinétique et potentielle, mais elle est associée au mouvement aléatoire des atomes dans un objet, cette association est appelée énergie thermique. Une grande partie de cette énergie est stockée dans les océans de la planète, chaque jour l'équivalent de l'énergie contenue dans 250 Milliards de barils de pétrole est absorbé

par les océans en chaleur provenant du soleil.

L'énergie chimique, est le produit de réactions chimiques dont la batterie automobile est un exemple. La réaction chimique produit donc du courant pour faire démarrer l'automobile. Certaines centrales utilisent le processus de la photosynthèse. L'énergie chimique est utilisée dans les laboratoires scientifiques pour produire des médicaments et de la puissance via le gaz.

L'énergie gravitationnelle potentielle et l'énergie cinétique fonctionnent ensemble, le terme énergie est motivé par le fait que les énergies ; potentielle et cinétique soient les différents aspects de l'énergie mécanique. L'énergie potentielle existe chaque fois qu'un objet a une masse et une position dans un champ de force, dans ce cas l'énergie d'un objet est donnée par la relation :  $PE = mgh$  , où  $PE$  est l'énergie en joule,  $m$  est la masse de l'objet,  $g$  est l'accélération gravitationnelle,  $h$  la hauteur de l'objet.

Par contre, l'énergie cinétique est l'énergie du mouvement d'un objet quelconque qu'il soit vertical ou horizontal. Il existe différentes formes de l'énergie cinétique vibratoire, l'énergie due aux vibrations : qui est une énergie rotationnelle. L'équation pour l'énergie cinétique est :  $EC = 1/2mv^2$ , où  $m$  représente la masse et  $v$  la vitesse. Cette équation montre que l'énergie cinétique d'un objet est proportionnelle au carré de la vitesse de l'objet.

Un réseau électrique est composé de quatre grandes parties : la production, le transport et interconnexion, la distribution et répartition et enfin la consommation. Les niveaux de tension y sont différents.

L'étape de la « production » concerne la production d'électricité par le biais de centrales. Une fois l'électricité produite, elle emprunte un réseau de lignes aériennes et souterraines avec ses voies principales (lignes à très haute et haute tension du réseau de transport d'électricité), ses voies secondaires (lignes à moyenne et basse tension gérées par les réseaux de distribution) et ses échangeurs (postes électriques).

Au moment de la production, la tension électrique (circulation du champ électrique le long d'un circuit) est moyenne, aux alentours de 20 000 volts. La figure 3.10 montre les différentes tensions dans un système de transmission et de distribution, et le passage d'une tension à une autre selon le secteur visé.

Chaque niveau de tension est divisé en deux catégories ; la première par sociétés de distribution, et la seconde par secteurs d'activité économique.

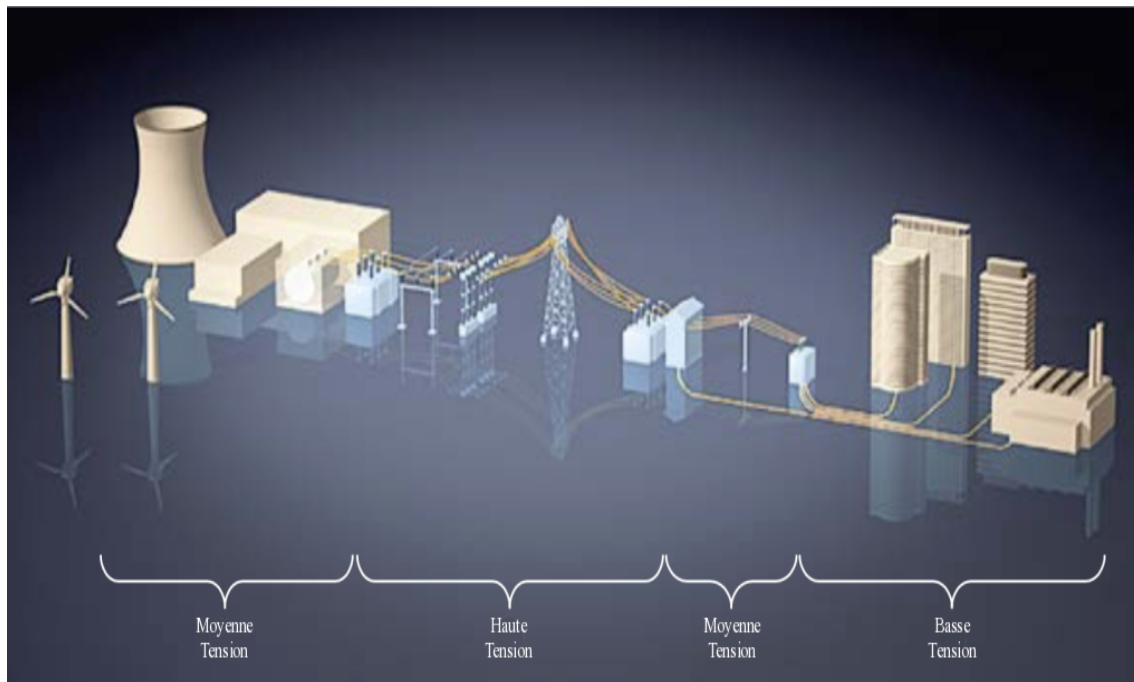


Figure 3.10 – Les différentes tensions dans un système de transmission et de distribution.[4]

### 3.6.2.1 Haute Tension (HT)

Lorsqu'elle intègre le réseau de transport et d'interconnexion, l'électricité est portée à une tension de 400 000 volts, soit une très haute tension, ce qui permet de transporter l'électricité en limitant les pertes. Le réseau de transport est interconnecté aux réseaux des pays voisins. La définition numérique de la Haute Tension (HT) dépend de deux facteurs, qui sont considérés dans la classification d'une tension comme haute, et qui sont : la possibilité de produire une étincelle dans l'air et le danger d'électrocution par contact ou par proximité. Les définitions peuvent indiquer la tension entre deux conducteurs d'un système, ou entre un conducteur et la terre. Dans l'ingénierie électrique de transmission de puissance, la haute tension est généralement considérée comme toute tension comprise entre environ 35 000 volts et 230 000 volts. Ceci est une classification basée sur la conception de l'appareil et l'isolation.

La Commission électrotechnique internationale et ses homologues nationaux (Institution of Engineering and Technology (IET) , Institute of Electrical and Electronics Engineers (IEEE) , Vanguard Energy (VDE) , etc. ) définissent comme haute tension ; toute tension supérieure à 1000 V pour le courant alternatif , et au moins 1500 V pour le courant continu. Les lignes de haute tension forment un réseau qui permet le transport et l'alimentation en électricité à une échelle régionale ou locale. Elles alimentent également en électricité les industries lourdes et les gros consommateurs. Leur tension est de 63.000 ou

90.000 volts. Pour désigner les tensions supérieures à celles que l'on trouve dans les lignes à haute tension, on parle de très hautes tensions. Les lignes très haute tension (225.000 et 400.000 volts) permettent de transporter l'électricité, depuis les centrales électriques, sur de longues distances et avec des pertes minimales.

La Haute tension est divisée en quatre sociétés de distribution :

- Société de distribution d'Alger (SDA).
- Société de distribution Centre (SDC).
- Société de distribution Est (SDE).
- Société de distribution Ouest (SDO).

Et elle est composée des secteurs économiques suivants :

- Eau et Energie EeE.
- Transport TRA.
- Grands Barrages.
- Autres Industries IND.
- Secteur d'Activité des Industries Grosses Consommatrices d'Energie.

### **3.6.2.2 Basse Tension (BT)**

L'électricité intègre ensuite des réseaux de distribution devant l'acheminer vers les particuliers, les petits commerçants, les collectivités locales et les petites et moyennes entreprises. Des « postes sources » permettent de faire passer le courant du réseau de transport aux réseaux de distribution. Ces échangeurs sont chargés de réduire la haute et très haute tension en basse tension.

La basse tension est un terme relatif, la définition varie selon le contexte. Plusieurs définitions sont utilisées dans la transmission électrique et distribution. Le code de sécurité électrique définit la basse tension comme tout circuit exempté de la protection requise à des tensions supérieures, cette définition varie entre les pays et certains codes spécifiques. Dans les systèmes d'alimentation électrique, la basse tension fait référence le plus souvent à la tension secteur utilisée par l'éclairage et les appareils portables. La « basse tension » dans ce contexte présente toujours un risque de choc électrique, mais seulement un risque mineur d'arcs électriques à travers l'air.

La Commission Internationale électrotechnique (IEC) définit le système d'alimentation basse tension que pour la tension dans la gamme 50-1000 V pour le courant continu ou 120-1500 V pour courant discontinu. Les lignes de distribution basse tension (entre

230 et 380 volts) sont les plus petites lignes du réseau électrique. Elles servent à alimenter les ménages, les artisans et les très petites industries en électricité.

La basse tension est elle aussi divisée en secteurs économiques et sociétés de distribution mais dans le cadre de cette thèse seule la seconde catégorie est prise en compte. Les sociétés de distribution pour la BT sont divisées en :

- Le Réseau Interconnecté du Nord (RIN) : le réseau interconnecté du Nord.
- Le Réseau Isolé du Sud (RIS) : le réseau isolé du Sud.

### **3.6.2.3 Moyenne Tension (MT)**

La Moyenne Tension (MT) de 10 à 35 kV est utilisée pour la distribution régionale du courant. Des réseaux locaux de distribution approvisionnent certains quartiers des villes ou des villages ainsi que des petites et moyennes exploitations industrielles.

La Moyenne Tension est divisée en quatre sociétés de distribution :

- Société de distribution d'Alger (SDA).
- Société de distribution Centre (SDC).
- Société de distribution Est (SDE).
- Société de distribution Ouest (SDO).

Et elle équipe les secteurs économiques suivants :

- Eau et Energie EeE.
- Secteur Tertiaire TER.
- Secteur Résidentiel RES.
- Petites et Moyennes Entreprises (P.M.E).
- Agriculture AGR.

### **3.6.2.4 La production énergétique**

Chaque équipement de production électrique présenté est dépendant de l'énergie utilisée. La figure 3.11 montre la production électrique par énergie en Algérie et le tableau ?? montre la production en MW par énergie utilisée.



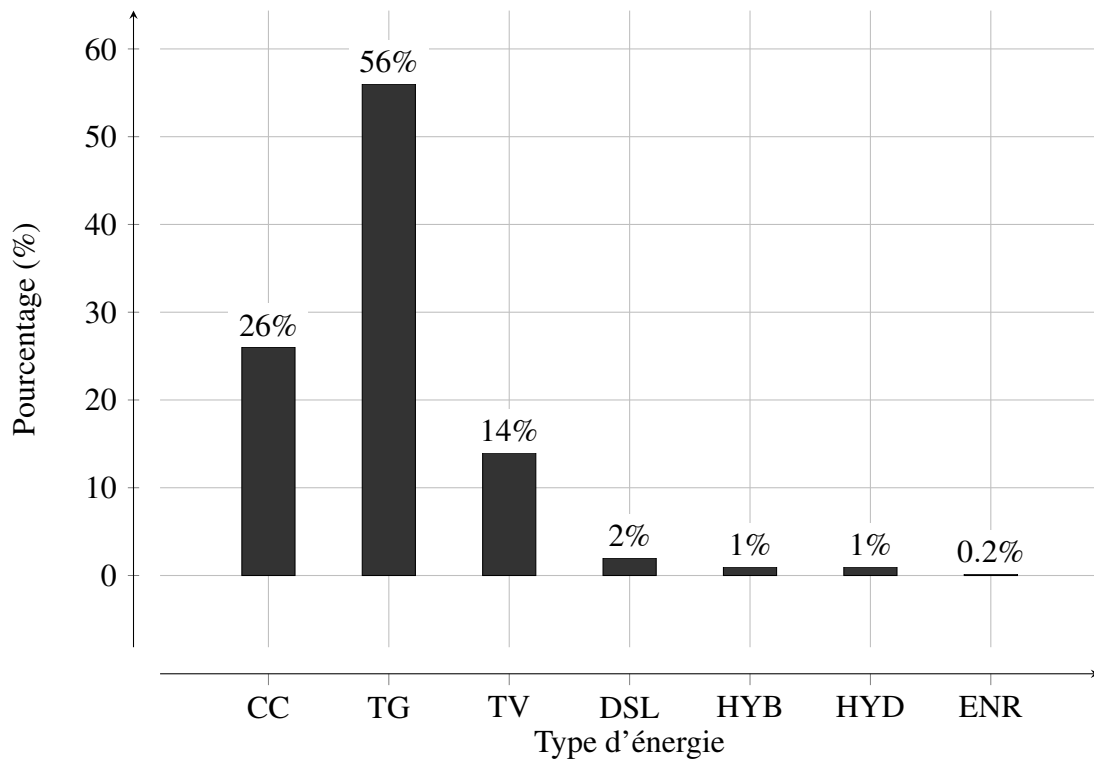


Figure 3.11 – Production électrique par types d'énergie en Algérie.

Type d'équipement	1980	1990	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Thermique Vapeur (TV)	3621	8397	16624	14558	14142	13384	11857	9692	9654	9422	9582	10221	10227
Thermique Gaz (TG)	2223	6704	15679	16463	17011	20339	19940	19564	22055	24075	17400	20211	26970
Cycle Combiné (CC)	-	-	386	3419	5321	5704	10318	15341	15701	18623	27685	28444	26122
Hydraulique (HYD)	251	135	555	218	226	277	342	173	378	389	98	193	145
Diesel	125	216	281	264	250	283	313	403	464	416	227	248	276
Hybride	-	-	-	-	-	-	-	-	1619	1159	1155	1181	889
Eolien	-	-	-	-	-	-	-	-	-	-	-	1	19
Photovoltaïque	-	-	-	-	-	-	-	-	-	-	-	1	14
Total	6220	15452	33525	34922	36950	39987	42770	45173	49871	54084	56147	60500	64662

Tableau 3.1 – La production en MW par énergie utilisée.

### 3.6.2.5 Infrastructures et distribution de l'électricité

Le programme de développement en moyens de production et transport d'électricité doit s'accompagner du renforcement du réseau de distribution (lignes MT/BT et postes), pour assurer la fiabilité de l'approvisionnement et de la distribution de l'énergie électrique et garantir une meilleure qualité de service pour les citoyens. A la fin de l'année 2011, la longueur totale du réseau national de distribution de moyenne et basse tension de l'électricité, a atteint les 263 585 km. Le tableau 3.2 montre la clientèle et la consommation par ville.

Wilaya	NOMBRE DE CLIENTS				CONSOMMATION (GWh)			
	HT	MT	BT	TOTAL	HT	MT	BT	TOTAL
ALGER	9	4791	891269	896069	385.4	2039.9	2663.6	5088.9
TIAPAZA	2	907	121087	121996	96.6	206.3	291.6	594.5
CHLEF	2	1177	174337	175516	211.5	228.3	469.4	909.2
TIARET	4	685	142927	143616	18.6	148.3	355.9	522.8
MOSTAGANEM	3	564	143909	144476	89	133.8	338.2	561
AIN DEFLA	2	1031	134387	135420	3.6	177.8	360.6	542
RELIZANE	-	687	128756	129443	-	102.3	317.3	419.6
TISSEMSILT	-	207	49443	49650	-	27.5	109.6	137.1
ADRAR	-	820	103132	103952	-	109.5	326.9	436.4
SAIDA	2	393	64614	65009	70.8	68.9	162.6	302.3
EL BAYADH	-	246	47869	48115	-	23.2	114.4	137.6
NAAMA	-	357	39205	39562	-	41.3	89.9	131.2
BECHAR	-	543	62791	63334	-	117.2	260.6	377.8
TINDOUF	-	259	11553	11812	-	37.1	49.2	86.3
ORAN	10	1862	310588	312460	685.4	681.9	835.5	2202.8
TLEMCEN	4	1294	235412	236710	340.1	350.9	592.3	1283.3
SIDI BEL ABBES	-	805	134509	135314	-	144.6	325.5	470.1
MASCARA	2	941	160994	161937	411.5	172.6	415.4	999.5
AIN TEMOUCHENT	3	379	91880	92262	400.2	62.9	209.4	672.5
CONSTANTINE	3	1114	212076	213193	178.8	305.9	566.5	1051.2
BATNA	1	1506	227063	228570	96.8	308.5	679	1084.3
OUMEL BOUAGHI	-	735	125343	126078	-	123.7	321.5	445.2
KHENCHELA	-	452	82087	82539	-	71.6	217.5	289.1
TEBESSA	4	850	119167	120021	105.9	144.8	314.3	565
MILA	1	716	146348	147065	264.7	142.6	370.7	778
ANNABA	6	902	141602	142510	697.5	205.6	408.7	1311.8
GUELMA	-	571	106172	106743	-	174.6	236.6	411.2
SKIKDA	11	909	153029	153949	759.5	183	386.8	1329.3
SOUK AHRAS	-	369	83855	84224	-	95.6	190.2	285.8
AL TAREF	-	554	78876	79430	-	130.8	210.1	340.9
SETIF	2	1911	312083	313996	184.2	548.1	921.2	1653.5
BEJAIA	1	1413	251256	252670	82	459.4	542.5	1083.9
JIJEL	-	625	125557	126182	-	89.9	262.5	352.4
M'SILA	4	1307	173516	174827	415.7	223.7	594.6	1234
BORDJ BOUARRERIDJ	-	895	134156	135051	-	226.2	362.9	589.1
BLIDA	4	1941	208138	210083	177.5	486.5	672.2	1336.2
BOUIRA	2	878	137438	138318	134.9	172.2	338.5	645.6
MEDEA	2	701	137318	138021	45.5	129.5	358.3	533.3
TIZI OUZOU	3	1106	300427	301536	32	329.4	695.7	1057.1
DJELFA	-	966	131539	132505	-	145.4	354.7	500.1
OUARGLA	7	1845	121111	122963	1607.8	438	549.3	2595.1
BISKRA	2	1584	142214	143800	33.7	299.8	824.4	1157.9
LAGHOUAT	6	768	69088	69862	287.1	160.4	244.4	691.9
TAMANRASSET	-	423	39935	40358	-	60.4	134.2	194.6
EL OUED	-	1125	137695	138820	-	152.2	784.6	936.8
GHARDAIA	-	753	98828	99581	-	154.9	324.6	479.5
ILLIZI	-	251	12667	12918	-	41.9	51.3	93.2
Total	102	45118	7057246	7102466	7816.3	10878.9	20205.7	38900.9

Tableau 3.2 – La clientèle et la consommation (GWh) par ville.

### 3.6.3 Données par puissance

La puissance est définie comme le taux du travail généré par l'énergie, ce qui équivaut à l'énergie par unité de temps. L'énergie reste indispensable pour générer de la puissance dans les centrales électriques pour produire de l'électricité. Les unités de la puissance sont le watt, le joule par seconde, le cheval-vapeur. Où :

$$\begin{aligned}1 \text{ Watt} &= 1 \text{ joule/seconde} \\1 \text{ kiloWatt} &= 1000 \text{ Watt} \\1 \text{ MegaWatt} &= 1000 \text{ KiloWatt} = 1 \text{ Cheval vapeur}\end{aligned}\tag{3.7}$$

L'électricité est la forme la plus pratique et polyvalente de l'énergie. Sa demande s'est vue croître d'une manière considérable en comparaison avec d'autres énergies. Les centrales électriques ont connu aussi une croissance phénoménale que ce soit du côté du volume ou de l'avancement technologique. L'électricité reste une variable indispensable dans le secteur industriel et de l'agriculture, par conséquent la consommation électrique est un indicateur de croissance et de productivité. Beaucoup d'efforts ont été fournis par Sonelgaz et ses sociétés filiales pour le renforcement des capacités de production, qui a connu ces dernières années une évolution conséquente de la puissance de production d'électricité installée. Celle-ci est passée de 7 492 MW en 2005 à 17 238,6 MW en 2015, soit près de 10 000 MW additionnelle mise en service en dix ans. La répartition de la puissance installée par filière et par producteur pour l'année 2015 est illustrée dans la figure 3.12, dont le parc de production national est constitué des centrales électriques de la Société Algérienne de Production de l'Électricité (SPE) et de Shariket Kahraba wa Taket Moutadjadida (SKTM), ainsi que des sociétés en partenariat, à savoir :

- Kahrama Arzew mise en service en 2005 ;
- Shariket Kahraba Skikda (SKS) mise en service en 2006 ;
- Shariket Kahraba Berrouaghia (SKB) (Médéa) mise en service en 2007 ;
- Shariket Kahraba Hadjret Ennouss (SKH) mise en service en 2009 ;
- SPP1 mise en service en 2010 ;
- Shariket Kahraba Terga (SKT) mise en service en 2012 ;
- Shariket Kahraba de Koudiet Edraouch (SKD) mise en service en 2013.

Le Tableau 3.3 montre l'évolution de la puissance produite en MW de chaque société de production du parc national algérien de l'année 1980 à 2015.

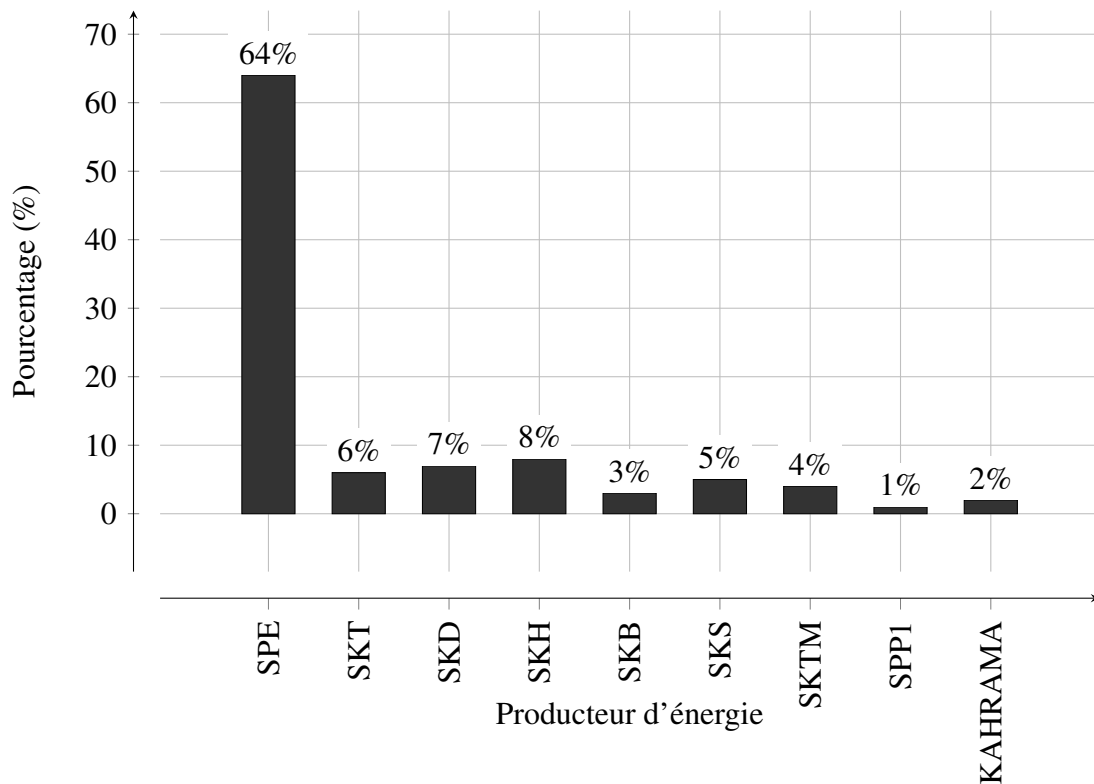


Figure 3.12 – La puissance installée par filière et par producteur pour l'année 2015.

	1980	1990	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
<b>SPE</b>	1852	4567	6762	6736	6752	6844	8439	8446	8506	8845	9351	10131	11258.7
<b>SKTM</b>	-	-	-	-	-	-	-	-	-	-	448	528	681.8
<b>KAHRAMA</b>	-	-	230	345	345	345	345	345	345	345	345	345	345
<b>SKS</b>	-	-	500	825	825	825	825	825	825	825	825	825	825
<b>SKB</b>	-	-	-	-	489	489	489	489	489	489	489	489	489
<b>SKH</b>	-	-	-	-	-	-	1227	1227	1227	1227	1227	1227	1227
<b>SPP1</b>	-	-	-	-	-	-	-	150	150	150	150	150	150
<b>SKT</b>	-	-	-	-	-	-	-	-	-	1122	1122	1122	1122.1
<b>SKD</b>	-	-	-	-	-	-	-	-	-	-	1140	1140	1140
<b>Total</b>	<b>1852</b>	<b>4567</b>	<b>7492</b>	<b>7906</b>	<b>8411</b>	<b>8503</b>	<b>11325</b>	<b>11482</b>	<b>11542</b>	<b>13003</b>	<b>15097</b>	<b>15957</b>	<b>17238.6</b>

Tableau 3.3 – Evolution de la puissance produite en MW de chaque société de production du parc national algérien.

### 3.6.3.1 Le réseau interconnecté du Nord (RIN)

Le RIN, étalé sur le Nord du pays et couvrant les régions de Béchar, Hassi Messaoud, Hassi R'Mel et Ghardaïa, est alimenté par une quarantaine de centrales de production d'électricité, reliées entre elles à travers un réseau de transport en 220 kV et 400 kV, permettant le transfert d'énergie des sites de production vers les centres de consommation. Durant les dix dernières années, la puissance maximale appelée du réseau interconnecté du Nord est passée de 5 921 MW en 2005 à 12 380 MW en 2015, soit une évolution moyenne annuelle de 7.2% comme le montre le tableau 3.4

	1980	1990	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
<b>PMA enregistrée (MW)</b>	1088	2742	5921	6057	6411	6925	7280	7718	8606	10363	10464	1927	12380
<b>Taux d'évolution annuel (%)</b>	-	-	6.9	2.3	5.8	8	5.1	6	11.5	16.9	1	4.2	11.7

Tableau 3.4 – La puissance maximale appelée du RIN.

### 3.6.3.2 Le pôle In Salah-Adrar-Timimoune (PIAT)

Le Pôle In Salah-Adrar-Timimoune (PIAT) est alimenté par les centrales turbines à gaz d'Adrar et d'In Salah, interconnectées à travers un réseau 220 kV s'étalant d'In Salah à Timimoun via Aoulef et Adrar. Le tableau 3.5 montre l'évolution de la Puissance Maximale Appelée (PMA) de 2008 à 2015.

	2008	2009	2010	2011	2012	2013	2014	2015
<b>PMA enregistrée (MW)</b>	106	121	149	172	199	232	261	279
<b>Taux d'évolution annuel (%)</b>	-	14	23	16	15	17	12.5	6.5

Tableau 3.5 – La puissance maximale appelée du PIAT.

### 3.6.3.3 Les réseaux isolés du Sud (RIS)

Il s'agit de 26 sites du grand Sud, alimentés par des réseaux locaux à travers des groupes diesels ou des turbines à gaz compte tenu des distances mises en jeu et des niveaux de consommation relativement faibles.

Pour faire face à l'évolution importante de la demande électrique des localités isolées du Sud, plusieurs centrales turbines à gaz et groupes diesel ont été installées entre 2012 et 2015.

Ces capacités de production ont été implantées au niveau des différents sites isolés du Sud et contribuent fortement à la sécurisation de leur alimentation en énergie électrique, au développement local et à l'amélioration des conditions de vie de ces régions désenclavées (Tableau 3.6).

	2011	2012	2013	2014	2015
<b>PMA enregistrée (MW)</b>	139.7	175.7	208.4	214	226
<b>Taux d'évolution annuel (%)</b>	9%	26%	19%	2.60%	5.60%

Tableau 3.6 – La puissance maximale appelée du RIS.

### 3.6.4 Variables exogènes

Une variable est exogène dans un modèle économique ou mathématique lorsque ses valeurs sont le résultat d'une observation, ou bien découlent d'hypothèses spéciales. Une variable exogène est explicative, mais non expliquée par le modèle. Ce sont des variables déterminées dans le modèle et qui déterminent les valeurs des autres variables. Plus précisément, on utilise des variables exogènes qui peuvent expliquer les variations de la variable endogène (la variable étudiée). Les variables étudiées dans la partie pratique sont définies ci-dessous.

#### 3.6.4.1 Produit Intérieur Brut

Le Produit Intérieur Brut (P.I.B) est un indicateur économique qui permet de mesurer la production économique intérieure réalisée par un pays. Il a pour objet de quantifier la production de richesses réalisées par un pays sur une période donnée, généralement un an ou un trimestre, grâce aux agents économiques résidant dans le pays concerné. Il s'agit donc d'un indicateur qui reflète l'activité économique interne d'un pays. La variation du P.I.B d'une année à l'autre permet de mesurer le taux de croissance économique d'un pays. Le P.I.B mesure la valeur de tous les biens et services produits par des agents économiques dans un pays sur une année, calculée selon le prix du marché. Il convient d'ajouter la valeur ajoutée récupérée par l'état. Une augmentation du PIB signifie qu'un pays connaît une croissance économique. A l'inverse, une diminution du P.I.B est l'indicateur d'une décroissance. Il se calcule par l'équation suivante :

$$PIB = \text{somme des valeurs ajoutées} + TVA + \text{droits et taxes sur les importations} - \text{subventions sur les produits} \quad (3.8)$$

#### 3.6.4.2 Taux d'Occupation par Logement et clientèle

Le Taux d'Occupation par Logement (TOL) caractérise le taux d'occupation de la population non sédentaire pour chaque logement, il est calculé par la formule suivante :

$$TOL = \frac{\text{Nombre d'habitants}}{\text{Nombre de logements}} \quad (3.9)$$

Le TOL est un bon indicateur de la moyenne du nombre de personnes dans un logement, il est généralement utilisé dans la basse tension qui est essentiellement constituée du secteur résidentiel. Chaque logement est considéré comme un client pour les fournisseurs d'électricité.

Le TOL est lié au nombre de logements, et les constructions de logements sont généralement planifiées sur plusieurs années, étant donné que c'est un processus long. Les gouvernements ont des plans de construction qui sont des objectifs de logements à réaliser sur une période donnée. Ce qui revient à dire que l'estimation des logements prévus à la construction n'est pas une tâche difficile, par conséquent cela nous donne une estimation de la clientèle BT dans le futur.

### **3.7 Conclusion**

Dans ce chapitre, nous avons parlé de l'histoire de l'électricité en essayant de souligner les plus grandes avancées dans le chemin de la découverte de l'électricité, ceci est pour des raisons historiques et de culture générale mais nous permet aussi de comprendre le principe même de l'électricité et les facteurs liés à cette dernière. Ensuite, nous avons discuté de l'énergie sous toutes ses formes et comment transformer cette dernière en électricité prête à être consommée par les utilisateurs du réseau. Ces parties sont non techniques ou du moins elles le sont le moins possible. L'objectif était d'avoir une idée sur le processus de chaque type de transformation, les énergies utilisées, ainsi que les avantages et inconvénients de chaque transformation.

La deuxième partie de ce chapitre se consacre au sujet de notre thèse c'est-à-dire l'Algérie, après une brève présentation de notre pays, nous avons exposé les spécificités géographiques et les richesses et faiblesses du pays surtout sur le point des énergies disponibles, ceci pour le mettre en parallèle avec la partie 1 du chapitre pour comprendre sur quoi l'Algérie se base pour ce qui est de la production électrique. Par la suite, nous avons exposé plus en détails les spécificités de la charge électrique algérienne et sa production en termes d'énergie et de puissance, et surtout en expliquant la structure du réseau électrique algérien qui reflète l'ensemble des données proposées dans cette thèse. Et pour finir nous avons discuté des variables ou données nécessaires à l'étude de la charge électrique algérienne.

## **Chapitre 4**

# **Approches régressives pour la prévision à moyen long terme**



## 4.1 Introduction

Les techniques de prévision quantitatives font un usage formel des données historiques et du modèle de prévision. Le modèle résume formellement les tendances dans les données et exprime une relation statistique entre les valeurs antérieures et actuelles de la variable. Ensuite, le modèle est utilisé pour projeter les tendances (Pattern) dans les données du futur. En d'autres termes : le modèle de prévision est utilisé pour extrapoler le comportement passé et actuel dans l'avenir. Il existe plusieurs types de modèles de prévision à usage général ; les plus utilisés sont des modèles de régression et les modèles de lissage. Les modèles de régression font usage des relations entre la variable d'intérêt et une ou plusieurs variables prédictives. Parfois, les modèles de régression sont appelés modèles de prévision de cause à effet, car les variables prédictives sont supposées décrire les forces qui causent ou conduisent les valeurs observées de la variable d'intérêt.

L'analyse de régression est l'une des techniques les plus populaires pour la modélisation prédictive. La popularité des modèles de régression peut être attribuée à l'interprétation des paramètres du modèle et la facilité d'utilisation. Cependant, la principale limite conceptuelle de toutes les techniques de régression est que l'on peut observer la relation, mais on ne peut jamais être sûr du mécanisme causal sous-jacent [33].

Ce chapitre décrit l'analyse préliminaire des données de la charge électrique algérienne, qui est une étape importante pour la construction de modèles prévisionnels. La charge électrique est une série temporelle complexe qui, entre autres choses change au fil du temps et intègre des cycles ou saisonnalités. Nous allons présenter par la suite le processus de sélection d'un modèle.

L'une des contributions de notre thèse, en plus de la réalisation d'une prévision, est l'implémentation de l'outil PREVELEC-DZ qui utilise les modèles de régression multiple et plus précisément la régression linéaire multiple et la régression exponentielle multiple. Ce modèle permet également de visualiser la prévision, et manager tous les aspects relatifs à son exercice.

## 4.2 Disponibilité des données

Les données utilisées dans le cadre de cette thèse ont été fournies par l'entreprise Sonelgaz. Plusieurs types de données ont été fournis ; les données macro-économiques comme le Produit Intérieur Brut (P.I.B), ainsi que les données concernant les plans de

construction d'habitations d'ici 2030. Sonelgaz a également fourni les données de la charge électrique ainsi que sa clientèle.

La structure du réseau national algérien expliqué dans le chapitre 3, ainsi que les différents types de données sont l'entrée principale de notre étude. Pour une prévision de la charge électrique, il est nécessaire d'avoir l'historique de cette dernière. Nous allons récapituler les données fournies par Sonalgaz :

- Basse Tension (BT) : par Société de Distribution (SD).
- Moyenne Tension (MT) : par Société de Distribution (SD) et Secteur d'Activité Economique (S.A.E).
- Haute Tension (HT) : par Société de Distribution (SD) et Secteur d'Activité Economique (S.A.E).
- Matrice de courbe de charge RIN.
- Matrice de courbe de charge RIS.
- Matrice de courbe de charge du Pôle In Salah-Adrar-Timimoune (PIAT).
- Le Produit Intérieur Brut (P.I.B).
- La consommation Moyenne BT.
- La clientèle BT et le Taux d'Occupation par Logement (TOL).
- La clientèle BT par SD .

Le tableau 4.1 résume l'échelle de temps, la rangée des données ainsi que le nombre de points disponibles par données.

Par énergie (GWh)	BT	RIN	De 2000 à 2012	13 (par an) points
		RIS		13 (par an) points
		Total		13 (par an) points
	HT/MT	SD	De 2000 à 2012	13 (par an) x 5 points
		S.A.E	De 2000 à 2011	12 x 6 points
Par puissance (MW)	RIN		De 2000 à 2012	8760 (heure) x 13 (Année)
	RIS		2012	8760 (heure) par région
	PIAT		De 2008 à 2012	8760 (heure) x 5 (Année)

Tableau 4.1 – L'échelle de temps et la rangée des données de la charge électrique algérienne.

### 4.3 Étude de la charge électrique algérienne

Comme expliquées dans les sections précédentes, les données fournies par Sonelgaz sont divisées en deux parties ; celles en énergies (GWh) représentant la consommation annuelle des différents niveaux de tension, et les données par puissance par horaire, qui

représentent la puissance (MW) produite par heure. Chaque section fait l'objet d'une étude approfondie. Bien que les données des deux sections sont différentes, elles restent néanmoins complémentaires pour la finalité de notre proposition.

La complexité de l'étude est de comprendre tout le processus de production et de distribution de l'électricité, ainsi que l'identification et la prévision des facteurs influant sur la consommation de cette dernière et leur incorporation dans des modèles régressifs linéaires et/ou exponentiels afin d'obtenir une projection sur le long terme allant jusqu'à 30 ans. Une première partie de l'étude est basée sur une prévision en énergie (GWh) annuelle de la production électrique, laquelle peut être convertie en puissance. La seconde partie est une prévision en puissance (MW) en utilisant les matrices de courbe de charge qui représentent la production par heure sur toute une année et la puissance préalablement prédite, on obtient donc une prévision sur une échelle de temps plus petite (heure, semaine, mois).

La consommation énergétique algérienne globale de l'année 2000 à 2012 montrée dans la figure 4.1 est le résultat de la somme des consommations des trois tensions ; BT, MT et HT ce qui revient à dire que pour comprendre un phénomène donné sur la consommation globale, il est obligatoire de découvrir la ou les tensions qui sont à l'origine du phénomène étudié. Dans la figure 4.1 la consommation de l'année 2003 et 2005 montre une croissance inhabituelle par rapport à l'année qui les précède. Le tableau 4.2 représente le taux de croissance de chaque année, où on remarque deux phénomènes : le premier en 2003 où le taux de croissance est de 7.86% ce qui fait 3.16% de plus que la croissance de 2002 qui est de 4.70%, le pourcentage de la croissance diminue radicalement de -4.10% l'année qui suit (2003) pour passer à 3.76%. Le même scénario se répète entre 2008 et 2009 avec une croissance de -3.32%.

Année	Taux de Croissance
2001	5.47%
2002	4.70%
2003	7.86%
2004	3.76%
2005	5.14%
2006	4.54%
2007	5.63%
2008	6.96%
2009	3.64%
2010	5.55%
2011	8.50%
2012	9.31%

Tableau 4.2 – Le taux de croissance annuel de la charge électrique algérienne.

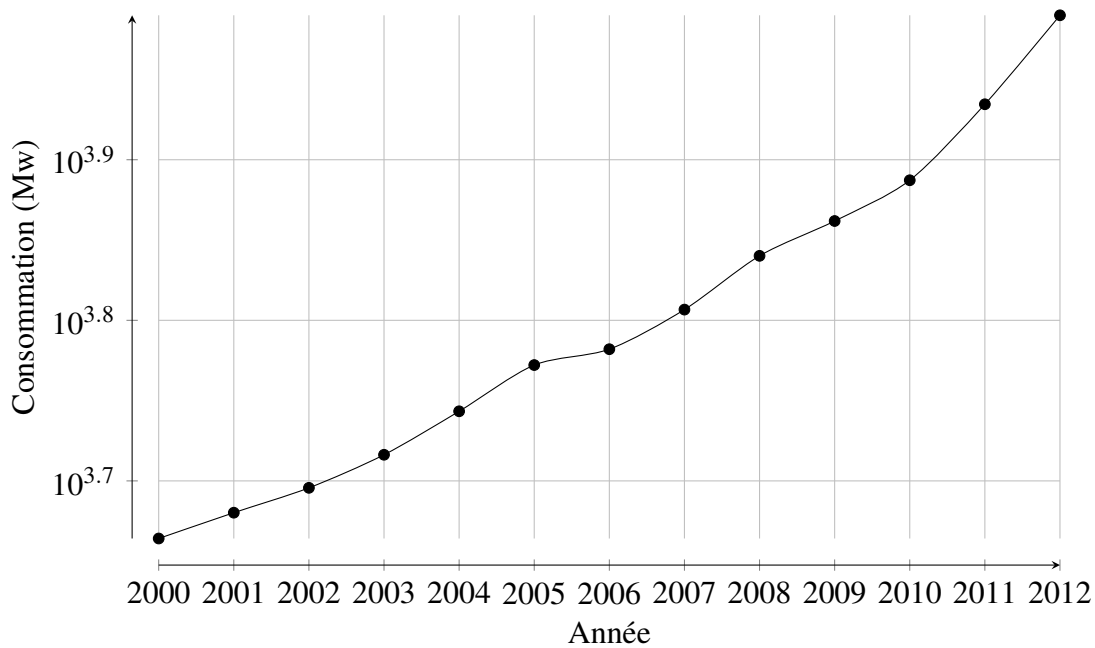


Figure 4.1 – La consommation énergétique algérienne globale de l'année 2000 à 2012.

Les acteurs qui peuvent influencer la consommation globale sont premièrement les trois tensions, ensuite chaque tension est quant à elle influencée par d'autres facteurs appelés variables exogènes ; qui ont été étudiées séparément en utilisant une analyse régressive pour déterminer l'impacte de la variable exogène sur la tension électrique. Le tableau 4.3 montre le taux de croissance des différentes tensions, on remarque que le même schéma se reproduit dans la BT et MT en 2003 et 2008. Ce phénomène est clairement expliqué par une consommation moyenne qui diminue après 2003 et 2008 comme le montre la figure 4.2 où on constate visuellement une baisse du taux de croissance de la consommation moyenne en 2004 et 2009. Visuellement on constate que la variance de la consommation moyenne par client peut expliquer la variance de la production électrique, ce qui fait de cette variable une variable explicative pour les modèles linéaires et exponentiels de la BT et MT. le choix d'une variable exogène ou explicative ne se fait pas que visuellement, une erreur de corrélation est calculée pour chaque variable pour appuyer le constat visuel effectué en premier lieu.

Les trois tensions sont étudiées séparément, chaque tension est différente par sa nature, les secteurs qu'elle cible et le nombre de ses clients. La BT, MT et HT ont une croissance de 1038.9, 407.55, 283.27 GWh par an respectivement. Le  $R^2$  ou le coefficient de détermination démontre que la variabilité de la charge électrique qui peut s'expliquer par l'année avec un pourcentage qui avoisine les 95.6%, 97.6% et 96.2% pour la BT, MT et HT respectivement, et ce qui reste est dû à des causes mal comprises et facteurs aléatoires et s'explique donc très mal, on serait incapable de dresser la liste de ces causes et, même si c'était faisable, il serait difficile d'expliquer les mécanismes par lesquels les causes se

conjuguent. Cette procédure a été faite pour toutes les séries temporelles en énergie, fournies par notre partenaire, pour déterminer si des variables macro-économiques et/ou le décalage nécessaire dans l'historique de la même série doivent être intégrés pour avoir un bon modèle de prévision.

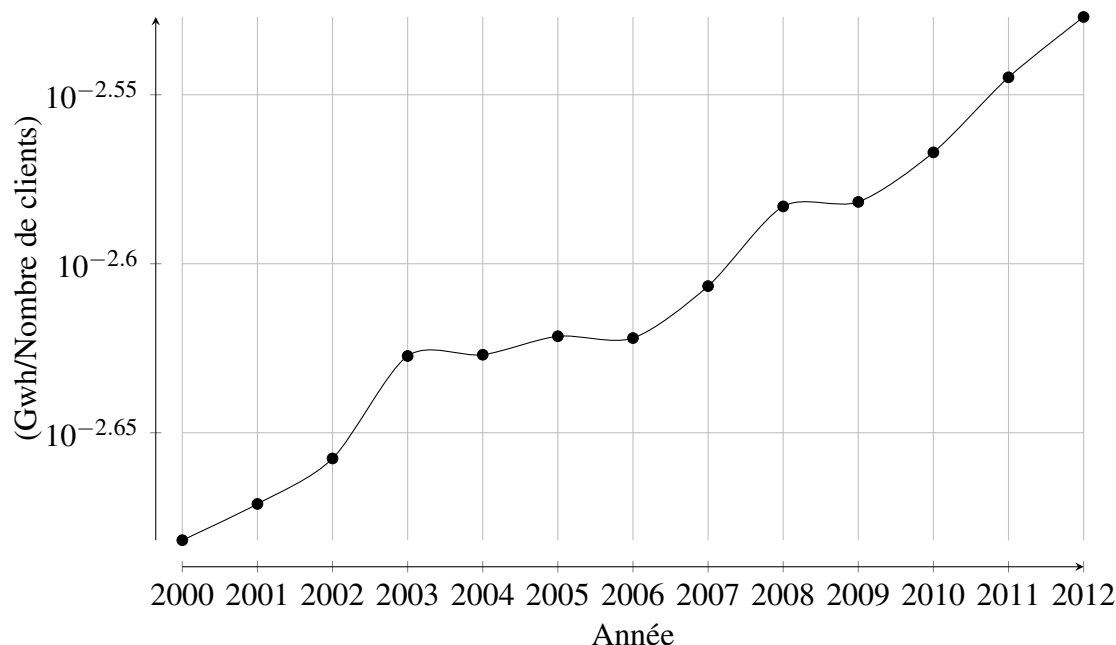


Figure 4.2 – La consommation moyenne de la clientèle BT de 2000 à 2012.

Année	BT (GWh)	MT (GWh)	HT (GWh)
2001	6.10%	4.39%	5.72%
2002	7.13%	3.79%	2.17%
2003	11.93%	10.50%	-1.06%
2004	4.90%	2.15%	4.27%
2005	5.08%	4.64%	7.35%
2006	5.08%	2.95%	6.62%
2007	7.29%	3.91%	5.85%
2008	9.84%	6.37%	3.67%
2009	4.45%	2.67%	3.72%
2010	8.10%	4.36%	2.62%
2011	9.91%	6.65%	11.47%
2012	14.41%	9.06%	1.51%

Tableau 4.3 – Le taux de croissance annuel des trois tensions BT, MT et HT.

La régression linéaire n'est pas utilisée pour déterminer un phénomène de cause à effet, et notre analyse n'a pas pour but de déterminer une causalité entre le temps et la production de la charge électrique annuelle, mais démontre que la variation des valeurs de la charge électrique peut s'expliquer par les variations des valeurs du temps et plus particulièrement l'année. Aussi, la prévision de la charge électrique est faite sur trois

scénarios ; faible, moyen et fort qui sont obtenus en prévoyant trois scénarios pour les variables exogènes de chaque modèle.

### 4.3.1 Prévision de la Société de distribution Est (SDE)

Pour illustrer la partie prévision des modèles multiples ; linéaires et exponentiels, nous allons étudier la Société de distribution Est (SDE) de la Basse Tension. La prévision est faite sur trois scénarios ; faible, moyen et fort. Chaque scénario est défini par le taux de croissance des variables impliquées dans la tension étudiée (variables exogènes), comme mentionné plus haut une sélection de ces variables est préalablement réalisée. Plusieurs modèles sont générés pour chaque série étudiée et le meilleur d'entre eux, en termes d'erreur, sera pris comme modèle définitif pour la série temporelle.

Nous allons présenter les résultats des modèles régressifs à la fois linéaires et exponentiels de la SDE en interprétant et expliquant les critères de chaque étape de l'analyse de la série temporelle étudiée. Il est important de rappeler que nos données représentent la consommation annuelle en (GWh). Les variables explicatives utilisées dans le modèle sont :

- La clientèle reliée au réseau de la société de distribution de l'Est.
- La consommation moyenne par habitation.
- La consommation au moment  $t - 1$ .

Les deux premières statistiques qu'on analyse sont le coefficient de détermination  $R^2$  et l'erreur-type du modèle (Tableau 4.4). Le coefficient de détermination qui avoisine les 99% signifie que la variabilité de la charge électrique peut s'expliquer par les variables exogènes, et les 1% restant s'expliquent très mal et ceci est dû à des causes mal comprises et des facteurs purement aléatoires. On serait incapable de dresser la liste de ces causes et même si c'était faisable, il serait difficile d'expliquer les mécanismes par lesquels les causes se conjuguent. Par conséquent, ce taux de 99%, dénote une forte corrélation entre les variables explicatives (exogènes) et la variable expliquée (consommation de la SDE).

Coefficient de détermination multiple	0.99
Coefficient de détermination $R^2$	0.99
Erreur-type	57.31

Tableau 4.4 – Statistiques de la régression.

L'erreur-type représente la distance standard entre les valeurs des données et la droite de régression. Elle est utilisée pour évaluer la capacité de l'équation de régression à pré-

voir la réponse. Plus l'erreur-type est petite, plus l'équation prévoit correctement la réponse. Avec une erreur de 57.31, ce modèle de la SDE est le meilleur en comparaisons avec les autres.

ANOVA ou l'analyse de la variance, montrée dans le tableau 4.5, permet d'étudier le comportement des variables impliquées dans l'équation de régression. L'interprétation des résultats consiste en la vérification des hypothèses émises avant chaque régression. Le F de Fisher nous indique le rapport entre la variance de la variable dépendante expliquée et non expliquée par le modèle de régression. L'hypothèse que ce test tente d'évaluer (hypothèse nulle) (Annexe A) est que le rapport entre la variance expliquée par le modèle est (approximativement) égale à la variance qui reste inexpliquée (auquel cas  $F = 1$ ). Dans notre cas le F de Fisher est égale à 60363.98, ce qui signifie que la part de la variance de la variable dépendante (consommation de la SDE) expliquée par le modèle est 60363.98 fois plus importante que la part de la variance de la même variable qui reste inexpliquée. La valeur critique de  $F = 5.69E-38$  signifie qu'il y a 5.69E-38 de chances sur 100 que l'on observe, sur un échantillon donné, un F supérieur ou égal au F calculé, nous en déduisons donc que globalement notre modèle est satisfaisant.

	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	3	594967442.6	198322480.9	60363.98	5.69E-38
Résidus	19	62423.42	3285.44		
Total	22	595029866			

Tableau 4.5 – Analyse de variance.

Le tableau 4.6, récapitulatif du modèle, confirme que le modèle est significatif. La signification de t nous permet de répondre à la question ; si chaque variable contribue significativement au modèle. Plus la valeur de t est élevée, plus la valeur de probabilité est petite et plus le prédicteur contribue au modèle. Dans notre cas, nous avons pris un intervalle de confiance à 95%, ce qui implique que la probabilité de la variable explicative doit être  $> 0.05$ . Nous pouvons en déduire que seules les variables consommation moyenne et la consommation de la SDE au moment t-1 sont significatives pour le modèle. Cependant, même avec une probabilité de la clientèle à 0.89, cette variable réduit considérablement l'erreur du modèle en comparaison avec les modèles sans clientèle.

SDE	Coefficients	Erreur-type	Statistique t	Probabilité	Seuil de confiance inférieure = 95%	Seuil de confiance supérieure = 95%
Constante	-1440.93	252.29	-5.71	1.66E-05	-1969	-912.87
Clients	1.23E-05	9.45E-05	0.13	0.89	-0.0001	0.0002
Conso Moy	957108.76	218792.04	4.37	0.0003	499171.75	1415045.78
T-1	0.89	0.04	19.87	3.58E-14	0.79	0.98

Tableau 4.6 – Récapitulatif du modèle linéaire de la SDE.

Le modèle linéaire multiple présenté plus haut permet d'avoir une prévision satisfaisante jusqu'à l'année 2023. Le modèle est modélisé en utilisant les données de 2000 à 2010 et le test du modèle est fait sur l'année 2011 et 2012. Tous les modèles proposés dans

cette étude ont suivi le même schéma. Ceci dit, quelques-uns ont été testés sur d'autres années à cause du manque de données disponibles. La figure 4.3 montre la comparaison entre la prévision 2011, 2012 du modèle proposé pour la SDE et les valeurs réelles. Avec un MAPE de 2.93%, le modèle est le meilleur en terme de performance par rapport aux autres modèles. Une projection des trois scénarios est réalisée jusqu'en 2023 en utilisant le modèle présenté. (Figure 4.4)

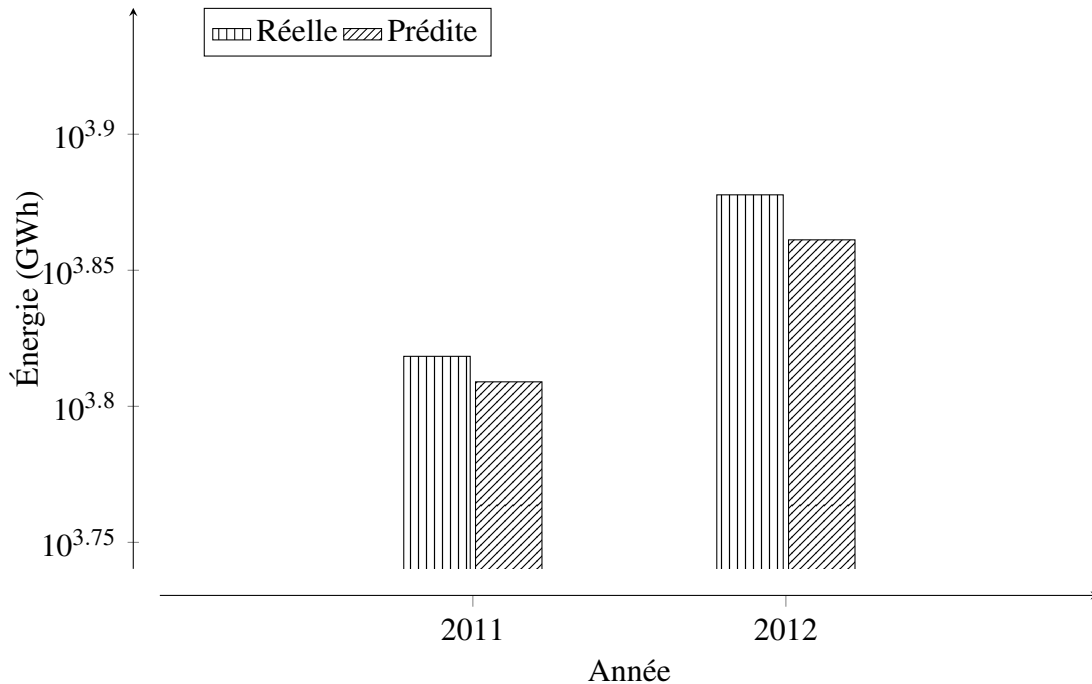


Figure 4.3 – Comparaison des valeurs prédites et réelles de 2011 et 2012 du modèle linéaire SDE.

Pour une projection plus réaliste de la charge, et pour refléter la grande croissance de la clientèle qu'elle soit globale ou par société de distribution prévue jusqu'à l'année 2015, une projection exponentielle est nécessaire ; en vue du nombre d'habitations (clients) délivrées jusque-là. Cependant, dans la figure 4.4 la croissance sera exponentielle si on suit le même schéma de croissance de la clientèle, ce qui est considéré comme un problème. La consommation électrique ne peut pas suivre une croissance exponentielle indéfiniment, elle doit avoir un taux de croissance bas dans le long terme.

Pour faire face aux problèmes, nous avons proposé une combinaison entre un modèle exponentiel et un modèle linéaire, le processus de sélection de chacun des deux modèles reste identique. L'équation 4.1 montre le modèle exponentiel sélectionné pour la société de distribution Est (SDE), et les résultats de ce dernier montrent une erreur MAPE de 2.5% sur l'année 2011 et 2012, ce qui est meilleur en terme de performance ; par rapport au modèle linéaire. La figure 4.5 montre une prévision jusqu'à l'année 2030 en mode mixte, c'est à dire une prévision exponentielle sur les premières années suivie d'un basculement vers un modèle linéaire pour le long terme. L'année de basculement est modifiable selon



l'envie de l'expert, cependant, une limite à l'année de basculement est obligatoire et fixée à l'année 2023 pour la raison suivante : une croissance exponentielle de la consommation au long terme n'est pas une hypothèse envisageable.

$$SDE = 657.5 * 1^{cli} * 1^{consomoy} * 0.9^{t_1} \quad (4.1)$$

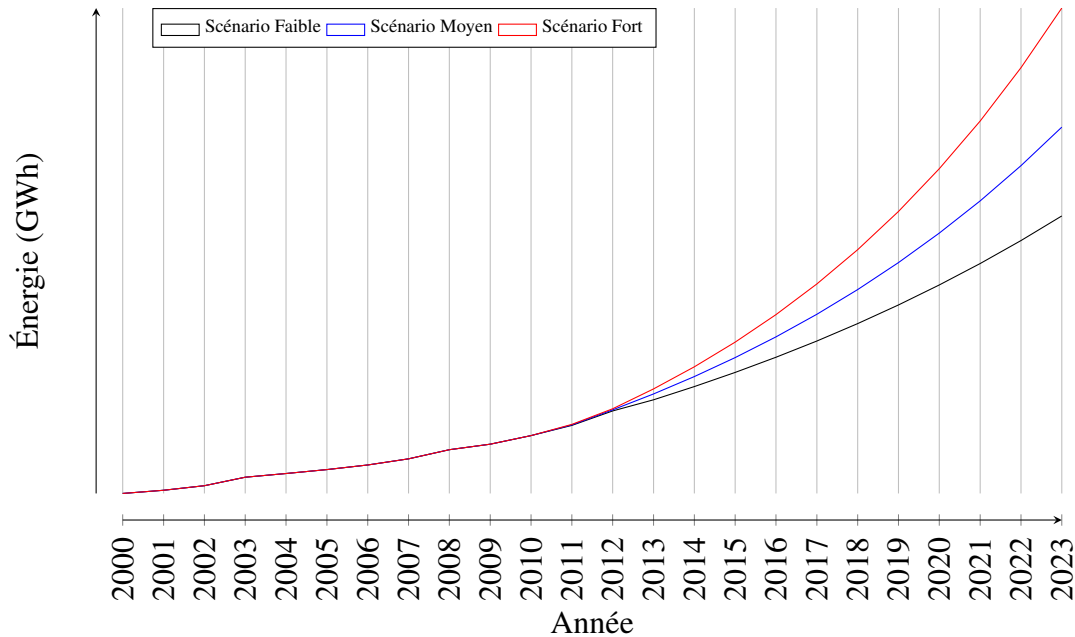


Figure 4.4 – Prévision du modèle linéaire SDE à l'horizon 2023.

Tous les modèles dans cette étude utilisent au choix ; une prévision linéaire multiple impliquant des variables exogènes pertinentes pour la tension étudiée et/ou une prévision mixte. Le choix d'une méthode mixte entre les deux régressions est motivé par deux raisons :

1. **L'insuffisance des données :** L'étude d'une série temporelle et sa bonne prévision sont conditionnées par deux facteurs importants ; le nombre de données disponibles et le modèle prévisionnel créé à partir des données. Plus l'échantillon de données est grand plus le modèle généré est performant et permet une bonne extrapolation vers le futur. Dans notre cas, treize années de données ont été fournies, ce qui réduit à onze années pour la génération des modèles régressifs et deux années pour la phase de test des modèles ; ce qui est insuffisant pour avoir de bonnes performances.
2. **La nature des séries temporelles :** Chaque série temporelle est différente par ; la variance, la moyenne, la distribution gaussienne, la corrélation et la saisonnalité de la série qui est fortement dépendante de la saison (temps) et des facteurs (variables exogènes) affectant la série étudiée. Les séries fournies dans cette étude sont an-

nuelles, ce qui rend l'impact des saisons sur la série insignifiant en comparaison avec les variables exogènes qui elles, ont un impact important sur la variance de la série. Dans notre cas, les indices économiques de l'Algérie au milieu des années 2000 étaient bons, ce qui a permis au gouvernement de réaliser un grand nombre de logements qui ont été branchés sur le réseau électrique, engendrant une augmentation de la consommation annuelle. Cependant, vers la fin des années 2000, l'incapacité de respecter le calendrier de livraison des habitations proposé, pour des raisons liées à la fragilité de l'économie, a rendu la consommation et par conséquent la production stable sur les dernières années.

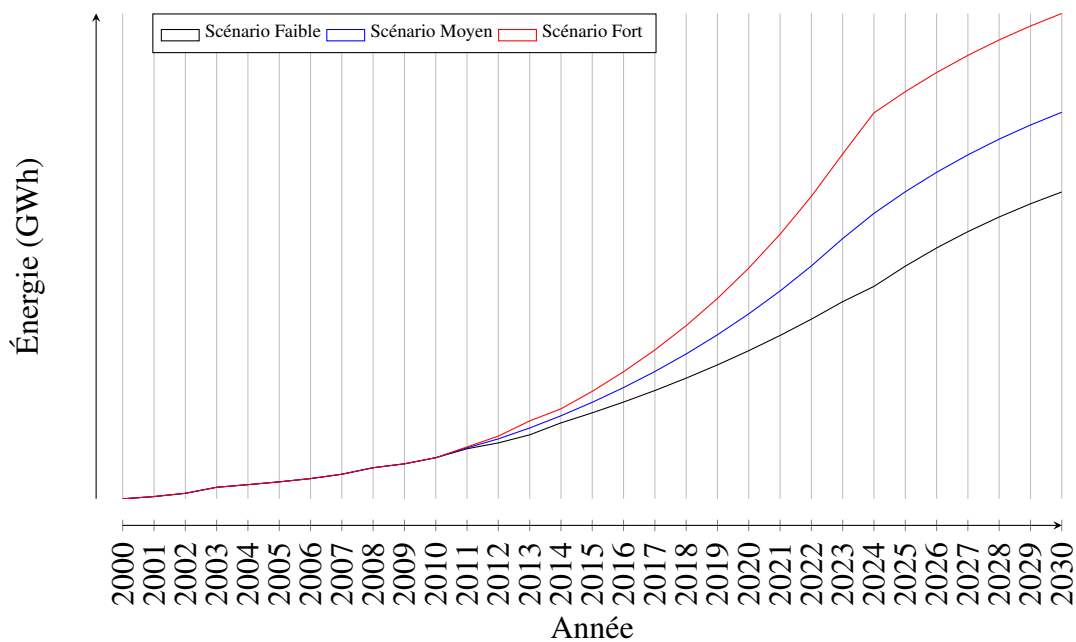


Figure 4.5 – Prévision du modèle SDE en mode mixte à l'horizon 2030.

Les raisons citées plus haut, se répercutent sur la consommation électrique d'une manière exponentielle du début jusqu'au milieu des années 2000 et linéaire vers la fin, ce qui a motivé le choix d'une méthode mixte. Le processus proposé dans cette section est complexe par le nombre de tensions et les secteurs d'activités économiques de chaque une d'entre elles, les 44 modèles régressifs (linéaires et exponentiels) retenus et implémentés dans ce chapitre sont montrés dans le tableau 4.7.

		Modèle linéaire	Modèle exponentiel
BT	BTALL	$-6239.4 + (cli * 4E^{-4}) + ((consomoy/100) * 630.0) + (t_2 * 0.3) + (t_1 * 0.3)$	$1583.8 * 1^{cli} * 1^{consomoy} * 0.9^{t_2} * 0.9^{t_1}$
	SDABT	$-971.8 + (cli * 0.0004) + ((consomoy/100) * 40.5) + (t_1 * 0.8)$	$438.1 * 1^{cli} * 1^{consomoy} * 0.99^{t_1}$
	SDCBT	$-963.2 - (cli * 3E^{-4}) + ((consomoy/100) * 91.2) + (t_1 * 0.7)$	$278.8 * 1^{cli} * 1^{consomoy} * 0.99^{t_1}$
	SDEBT	$-1440.93 + (cli * 1.23E^{-5}) + ((consomoy/100) * 95.7108) + (t_1 * 0.89)$	$657.5 * 1^{cli} * 1^{consomoy} * 0.9^{t_1}$
	SDOBT	$-1537.0 + (cli * 1E^{-4}) + ((consomoy/100) * 86.8) + (t_1 * 0.8)$	$488.0 * 1^{cli} * 1^{consomoy} * 0.99^{t_1}$
	RIS	$-134.9 + (cli * 0.001) + (t_2 * 1) + (t_1 * -0.04)$	$33.7 * 1^{cli}$
	PIAT	$-31.77 + (cli * 1E^{-4}) + (t_2 * 0.8) + (t_1 * 0.5)$	N/A
MT	MTALL	$-1553.3 + (0.2 * pib) + (-0.3 * t_2) + (1.4 * t_1)$	$1355.4 * 1^{pib}$
	RIS	$36.8 + (0.06 * cli) + (0.2 * t_2) + (0.4 * t_1)$	N/A
	AGR	$-50.3 + 0.01 * P.I.B + 0.8 * t_1$	$33.1 * 1^{pib} * 1^{t_1}$
	EeE	$-646.8 + 0.8 * pib - 0.6 * t_1$	$659.8 * 1^{pib} * 0.9^{t_1}$
	P.M.E	$-243.4 + 0.2 * pib + 0.6 * t_1$	$364.6 * 0.9^{t_1} * 1^{pib}$
	TER	$-972.9 + 0.7 * pib - 0.07 * t_1$	$990.2 * 1^{pib} * 1^{t_1}$
	RES	$38.8 + 0.03 * pib - 0.1 * t_1$	$89.2 * 0.9^{t_1} * 1^{pib}$
	PIAT	$57.4 + (0.1 * cli) - (0.01 * t_2) + (0.6 * t_1)$	N/A
HT	HTALL	$-887.68 + (t_2 * -0.06) + (t_1 * 1) + (0.1 * pib)$	N/A
	SDAHT	$8276.4 + (t_2 * 0.8) + (t_1 * 0.2)$	N/A
	SDCHT	$296153.3 + (t_2 * -0.1) + (t_1 * 1)$	N/A
	SDEHT	$241386.4 + (t_2 * -0.2) + (t_1 * 1.1)$	N/A
	SDOHT	$11.7 + (t_2 * 0.05) + (t_1 * 0.9)$	N/A
	EeE	$13000.0 + (t_2 * 0.3) + (t_1 * 0.6)$	N/A
	TRA	$3781.7 + (t_2 * 0.7) + (t_1 * 0.2)$	N/A
	IGCE	$338176.8 + (t_2 * -0.2) + (t_1 * 1.1)$	N/A
	AUT	$-(1 + (t_2 * 0.5) - (t_1 * 1.5))$	N/A
S.A.E	AGR	$134.9 + (cli * 0.001) + (t_2 * 1.04) + (t_1 * -0.04)$	N/A
	EeE	$-1227.0 + (pib * 0.2) + (t_2 * 0.3) + (t_1 * 0.7)$	$129.75 * 1^{pib}$
	IND	$-240.7 + (pib * 0.6) + (t_2 * 0.1) + (t_1 * 0.5)$	$3086.9 * 1^{pib}$
	RES	$-12267.2 + (cli * 0.001) + ((consomoy) * 6488941.85955412) + (t_2 * 0.2) + (t_1 * -0.04)$	$1475.1 * 1.0^{cli} * 1.05^{(consomoy * 10000)}$
	TER	$-4873.2 + (pib * 1.94) + (t_2 * -0.09) + (t_1 * 0.2)$	N/A

Tableau 4.7 – Tableau récapitulatif des différents modèles régressifs développés.

## 4.4 Le système de profils

A la fin du processus de sélection des modèles régressifs et de la prévision, une seconde phase commence par une conversion de l'énergie vers la puissance pour les objectifs suivants :

- Avoir une prévision en puissance pour déterminer la production nécessaire durant la période de prévision.
- La ventilation de la charge annuelle prédite précédemment via les profils de charge dans une échelle de temps plus courte ; heures, jours, semaines et mois pour obtenir les Puissance Maximale Appelée (PMA) de chacune d'entre elles.

Le système de profils représente l'usage de l'électricité dans une période de temps bien spécifique. L'extraction des profils se fait à partir de ce qu'on appelle les matrices de courbe de charge qui sont des matrices de taille 365 x 24 ou 366 x 24 dans le cas d'une année bissextile, ces matrices représentent la Puissance Maximale Appelée (PMA) de chaque heure de l'année.

Dans cette étude, nous proposons un nouveau type de profils ; composé du taux de production de chaque heure. Ce dernier est le pourcentage de chaque heure comparé au pic de la charge annuelle. Il existe trois types de profils :

- Le profil journalier : il est composé des 24 heures d'une journée, il faut noter qu'il y a trois types de jours :
  - Jours Ouvrables (JO) ;
  - Jours Fériés (JF) ;
  - Veilles Jours Ouvrables (VJO).
- Le profil hebdomadaire : il est composé du taux des 7 jours de la semaine.
- Le profil annuel : composé des 52 semaines de l'année.

Deux granularités ont été utilisées dans les études proposées dans cette thèse ; une par saison pour la partie de prévision en utilisant les modèles régressifs (première contribution), et une décomposition par mois pour les méthodes basées sur les réseaux neuronaux (deuxième contribution), le principe de décomposition étant le même pour les deux approches.

Afin d'extraire les profils, la matrice de courbe de charge par heure d'une année donnée passe par un processus dans lequel on obtient les trois profils cités plus haut. Pour illustrer ce dernier, nous allons prendre une décomposition par saison pour des raisons de simplicité mais le processus reste identique pour la décomposition par mois. Les étapes

de l'extraction des différents profils sont les suivants :

1. **Profils Journaliers** : 12 profils (3 par saison), l'année 2012 est donnée comme exemple dans le tableau 4.8. Le processus à suivre pour créer ces profils est le suivant :
  - (a) Identifier les trois types de jours ; Jours Fériés (JF), Veilles Jours Ouvrables (VJO) et Jours Ouvrables (JO) de l'année sélectionnée.
  - (b) Décomposer la matrice de courbe de charge en quatre saisons.
  - (c) Prendre le maximum de chaque heure des différents types de jours pour une saison, et refaire cette étape pour toutes les saisons.
  - (d) Diviser le maximum de chaque heure de chaque type de jour par la valeur maximale des 24 heures de ce dernier, et multiplier le résultat par 100 pour obtenir un pourcentage.

	Saison 1			Saison 2			Saison 3			Saison 4		
	VJO	JO	JF	VJO	JO	JF	VJO	JO	JF	VJO	JO	JF
00h00	69.40	70.48	69.88	76.59	77.04	77.27	89.02	88.49	89.21	68.09	69.67	69.14
01h00	65.56	66.57	65.48	71.50	71.98	72.33	84.63	83.63	84.64	64.71	66.30	65.75
02h00	64.05	64.87	63.68	69.10	69.40	69.24	80.37	80.00	80.44	63.23	64.59	64.10
03h00	63.06	64.07	62.59	67.83	67.90	67.63	78.53	77.78	78.49	62.36	63.80	63.24
04h00	62.71	63.68	62.17	67.77	67.81	67.38	76.67	76.17	76.55	61.98	63.36	62.77
05h00	63.51	64.33	62.82	67.69	67.69	67.26	74.12	73.16	73.21	62.88	63.99	63.41
06h00	68.04	68.21	66.13	65.56	64.86	64.27	71.28	69.80	69.57	67.28	67.86	66.74
07h00	75.35	75.11	70.25	68.55	66.97	64.68	69.44	67.92	67.26	75.22	74.25	69.69
08h00	77.40	77.08	73.07	73.65	72.47	70.51	72.48	70.85	69.91	76.93	75.90	72.55
09h00	81.68	81.57	78.44	81.67	79.84	77.83	78.82	77.16	74.67	81.64	80.42	77.78
10h00	81.34	81.19	79.51	84.17	82.65	81.50	82.78	81.07	78.84	82.23	80.89	79.58
11h00	80.55	79.91	78.64	86.34	84.39	83.36	86.43	84.80	82.99	81.89	80.51	79.71
12h00	79.53	78.36	77.37	87.28	85.26	84.54	89.85	88.19	86.90	80.61	79.35	79.04
13h00	79.21	77.93	76.81	88.43	86.35	85.42	93.09	91.72	90.57	80.15	78.93	78.28
14h00	78.99	77.51	77.10	89.34	86.79	86.22	95.14	94.58	93.88	79.68	78.65	78.06
15h00	77.97	76.69	75.91	88.34	86.55	85.19	95.42	94.84	93.80	79.04	78.06	77.24
16h00	77.21	76.14	76.09	86.67	85.11	84.25	93.25	92.93	92.18	78.48	77.72	77.32
17h00	76.44	76.00	76.61	82.66	81.73	81.91	88.93	88.83	88.83	78.63	78.44	78.84
18h00	82.39	82.17	82.56	80.32	79.47	79.16	86.09	85.92	86.41	91.95	91.27	91.68
19h00	95.33	95.62	96.08	82.06	81.09	81.56	88.85	88.20	88.89	100.00	100.00	100.00
20h00	100.00	100.00	100.00	93.11	92.76	93.37	96.96	96.42	96.32	98.31	98.36	98.85
21h00	96.02	95.75	96.12	100.00	100.00	100.00	100.00	100.00	100.00	91.74	92.52	92.78
22h00	86.17	87.27	86.97	95.57	95.82	95.60	96.97	97.94	98.13	83.37	83.45	83.70
23h00	76.30	77.52	77.44	85.18	86.64	86.46	91.76	93.37	94.15	74.21	75.41	75.08

Tableau 4.8 – Profils Journaliers des quatre saisons de l'année 2012.

2. **Profils Hebdomadaires** : 4 Profils (1 par saison), le tableau 4.9 montre les profils hebdomadaires de l'année 2012. Le processus à suivre pour créer ces profils est le suivant :
  - (a) Décomposer la matrice de courbe de charge en quatre saisons.
  - (b) Prendre le maximum de chaque heure des 7 jours de la semaine pour une saison, et refaire cette étape pour toutes les saisons.

- (c) Prendre le maximum des 24 heures de chaque jour.
- (d) Trouver le maximum des 7 jours.
- (e) Diviser le maximum de chaque jour par le maximum de la semaine et multiplier par 100 pour obtenir un pourcentage.

	<b>Dimanche</b>	<b>Lundi</b>	<b>Mardi</b>	<b>Mercredi</b>	<b>Jeudi</b>	<b>Vendredi</b>	<b>Samedi</b>
<b>Saison 1</b>	99.47	100	98.15	95.59	92.03	96.96	97.87
<b>Saison 2</b>	97.16	97.76	100	99.49	99.57	99.31	98.11
<b>Saison 3</b>	97.37	98.2	99.11	100	99.43	99.56	98.52
<b>Saison 4</b>	99.22	99.55	100	96.87	94.10	97.20	98.44

Tableau 4.9 – Profils hebdomadaires des quatre saisons de l’année 2012.

3. **Profils Annuels** : 1 Profil comme montré dans le tableau 4.10. Le processus à suivre pour créer ces profils est le suivant :

- (a) Diviser la matrice de courbe de charge en semaines.
- (b) Prendre le maximum de chaque semaine.
- (c) Trouver le maximum des 52 semaines.
- (d) Diviser le maximum de chaque semaine par le maximum des 52 semaines et multiplier par 100 pour obtenir un pourcentage.

Les profils sont liés entre eux, comme démontré dans la figure 4.6. Chaque semaine du profil annuel est connectée à un profil hebdomadaire, et le jour de chaque profil hebdomadaire est connecté à son tour au profil journalier associé.

L’intérêt des profils est d’offrir la possibilité du passage d’une puissance annuelle, considérée comme une valeur moyen long terme, vers un terme très court notamment heures ou jours et aussi pour avoir les Puissance Maximale Appelée (PMA) de chaque jour, semaine et mois pour prévoir la production de puissance nécessaire pour satisfaire la demande dans un futur proche.

Pour obtenir la charge par heure de chaque mois de l’année, on multiplie la charge prédite annuelle par le pourcentage des trois profils ; cela permet une détection facile des pics de la charge. Le système de profils permet un aspect dynamique en utilisant différents profils pour une tension spécifique ou un mois prédit. En plus, le système permet une modification du profil pour concorder avec le temps de prévision.

<b>Semaine</b>	<b>Valeur (%)</b>	<b>Semaine</b>	<b>Valeur (%)</b>
<b>Semaine N 1</b>	85.79	<b>Semaine N 27</b>	90.72
<b>Semaine N 2</b>	92.71	<b>Semaine N 28</b>	96.26
<b>Semaine N 3</b>	86.14	<b>Semaine N 29</b>	97.83
<b>Semaine N 4</b>	88.50	<b>Semaine N 30</b>	92.54
<b>Semaine N 5</b>	90.42	<b>Semaine N 31</b>	95.98
<b>Semaine N 6</b>	88.62	<b>Semaine N 32</b>	96.35
<b>Semaine N 7</b>	91.21	<b>Semaine N 33</b>	93.62
<b>Semaine N 8</b>	87.74	<b>Semaine N 34</b>	100.00
<b>Semaine N 9</b>	78.61	<b>Semaine N 35</b>	96.89
<b>Semaine N 10</b>	89.20	<b>Semaine N 36</b>	88.83
<b>Semaine N 11</b>	88.63	<b>Semaine N 37</b>	85.23
<b>Semaine N 12</b>	77.53	<b>Semaine N 38</b>	87.84
<b>Semaine N 13</b>	78.05	<b>Semaine N 39</b>	81.24
<b>Semaine N 14</b>	80.24	<b>Semaine N 40</b>	86.59
<b>Semaine N 15</b>	81.86	<b>Semaine N 41</b>	85.31
<b>Semaine N 16</b>	79.42	<b>Semaine N 42</b>	80.19
<b>Semaine N 17</b>	77.23	<b>Semaine N 43</b>	82.02
<b>Semaine N 18</b>	77.32	<b>Semaine N 44</b>	83.59
<b>Semaine N 19</b>	78.78	<b>Semaine N 45</b>	83.86
<b>Semaine N 20</b>	77.80	<b>Semaine N 46</b>	84.42
<b>Semaine N 21</b>	78.26	<b>Semaine N 47</b>	88.28
<b>Semaine N 22</b>	80.65	<b>Semaine N 48</b>	87.84
<b>Semaine N 23</b>	86.40	<b>Semaine N 49</b>	90.42
<b>Semaine N 24</b>	81.86	<b>Semaine N 50</b>	92.97
<b>Semaine N 25</b>	81.64	<b>Semaine N 51</b>	90.47
<b>Semaine N 26</b>	85.97	<b>Semaine N 52</b>	93.17

Tableau 4.10 – Profil annuel de l'année 2012.

## 4.5 PREVELEC-DZ

Suite à l'appel d'offre proposé par la société Sonelgaz pour répondre au besoin imminent et crucial d'un outil permettant la visualisation, l'analyse et la prévision de la charge électrique à moyen long terme, le Laboratoire de Gestion Electronique de Documents LabGED a été sélectionné pour mettre en œuvre :

1. Des modèles régressifs pour la prévision de tous les niveaux de tension.
2. Une application permettant la visualisation de l'historique des données, gestion des profils de courbes de charge, réalisation de prévisions, visualisation des résultats et sauvegarde de ces derniers.

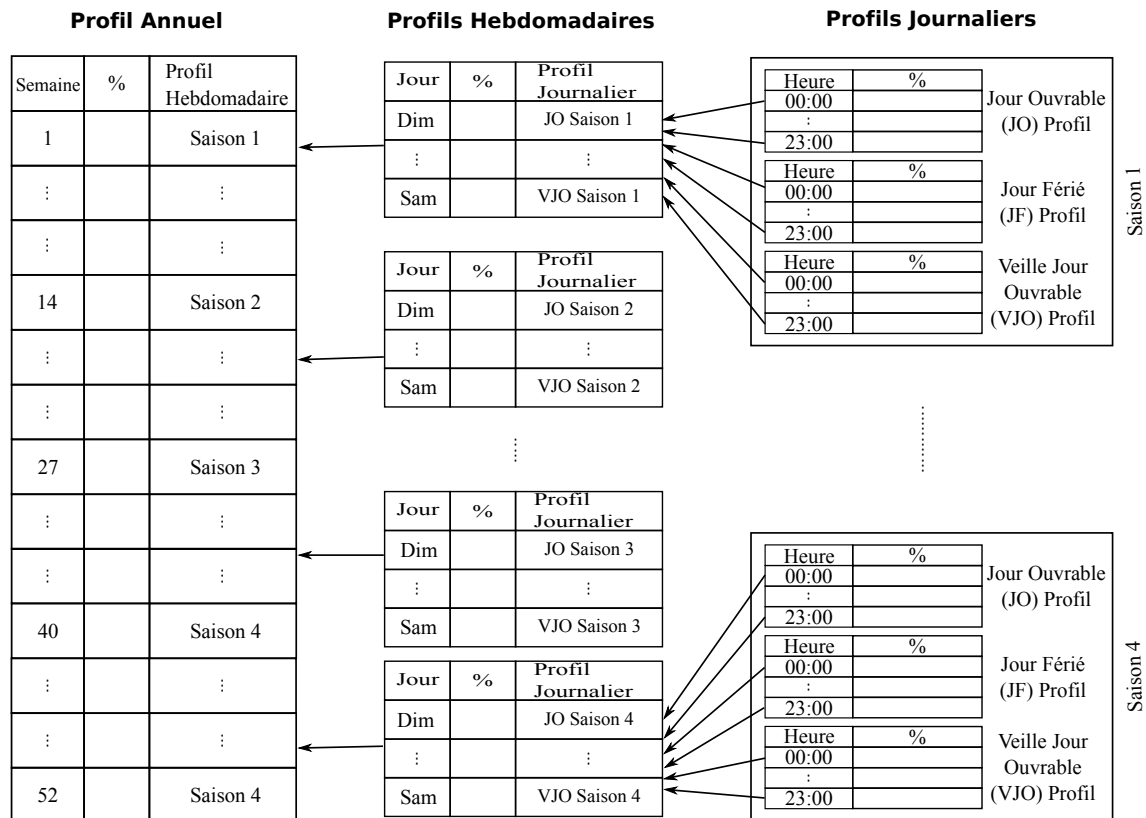


Figure 4.6 – Le système de profils par saison.

#### 4.5.1 Technologies utilisées et environnement de développement

Le choix de l'environnement est fortement lié à la plateforme utilisée et au type de l'application (Application serveur, Application lourde). La demande présentée par Sonelgaz consistait en la création d'une application monoposte, pour leurs terminaux, fonctionnant avec la plateforme Microsoft Windows.

Le choix du langage de programmation Csharp s'imposait par lui-même, ce dernier étant un langage orienté. Objet proposé par Microsoft pour concurrencer Sun Java, le CSharp combine la puissance du C++ et la facilité de programmation proposée par Visual Basic. Il est conçu pour fonctionner avec la plate-forme .Net de Microsoft ce qui facilite son intégration dans un environnement Microsoft Windows.

Le langage de développement étant choisi, l'environnement Visual Studio s'imposait comme l'IDE le plus approprié pour la mise en œuvre du programme PREVELEC-DZ. La volonté de réaliser une solution moderne et complète pour l'entreprise de production d'électricité nationale à travers l'utilisation de bibliothèques graphiques très poussées, notamment Krypton Suite de la compagnie ComponentSource permettant l'intégration du Ruban style Microsoft Office dans PREVELEC-DZ. Une autre librairie connue, très



utilisée dans le développement d'applications professionnelles pour Microsoft Windows qui est DevExpress a été utilisée pour la génération de graphiques modernes et complets.

Dans l'article présenté par Malcolm Atkinson et al. dans le manifeste des systèmes de base de données orientées objet en 1995 [50], ils définissent un Système de Gestion de Base de données Objet ou OODBMS comme devant satisfaire deux critères : d'être un Gestionnaire de Base de Données DBMS et être un système orienté objet. Donc, un OODBMS supporte la modélisation et la création de données objet, ceci inclut le support des classes des objets, et l'héritage des propriétés et méthodes d'une classe. La librairie db4o est un projet open source distribué par la société Actian, cette librairie a été choisie pour la gestion des données sous PREVELEC-DZ.

## 4.5.2 Fonctionnalités de PREVELEC-DZ

L'outil PREVELEC-DZ a été conçu pour répondre à plusieurs besoins parmi lesquels :

1. Gestion des utilisateurs et de la Base de Données.
2. Visualisation des données ; que ce soit la charge électrique ou les données exogènes.
3. Création et gestion des profils.
4. Prévision par niveau de tension
5. Prévision par secteur d'activité économique.
6. Visualisation graphique des différents scénarios de prévision.
7. Conversion entre l'énergie et la puissance.

Pour des raisons de sécurité, l'application PREVELEC-DZ gère les utilisateurs autorisés à se connecter, cela, à cause des données sensibles présentes dans la base de données que seul un expert avec les droits nécessaires peut y accéder. A l'ouverture de l'application seuls les boutons Connexion et Quitter sont actifs contrairement aux autres qui sont grisés comme indiquer dans la figure 4.7. L'utilisateur est prié d'introduire un Nom d'utilisateur et un mot de passe valide pour accéder aux fonctionnalités de PREVELEC-DZ, les utilisateurs sont gérés par un administrateur.

Après la connexion à l'application avec une combinaison de Nom d'utilisateur et mot de passe correct, l'utilisateur voit apparaître de nouveaux menus dans le ruban, ainsi que le dégrisement des boutons de l'accueil. Nous allons voir dans ce qui suit les différents menus :

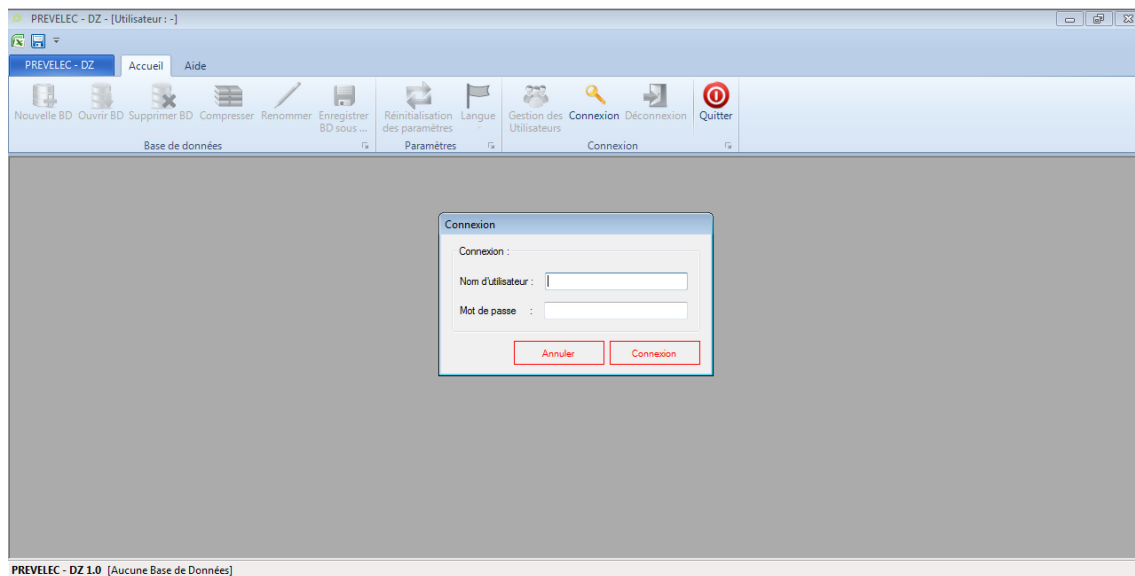


Figure 4.7 – Fenêtre de connexion des utilisateurs.

#### 4.5.2.1 L'Accueil

Le menu de l'accueil (Figure 4.8) est composé de 3 sous-sections qui permettent respectivement, la gestion des bases de données BD (créer une nouvelle BD, l'ouvrir, la supprimer, la compresser, la renommer ou l'enregistrer), définir les paramètres (Réinitialisation des paramètres, Langue) ou la gestion de connexion (gestion des utilisateurs, connexion, déconnexion et quitter l'application)

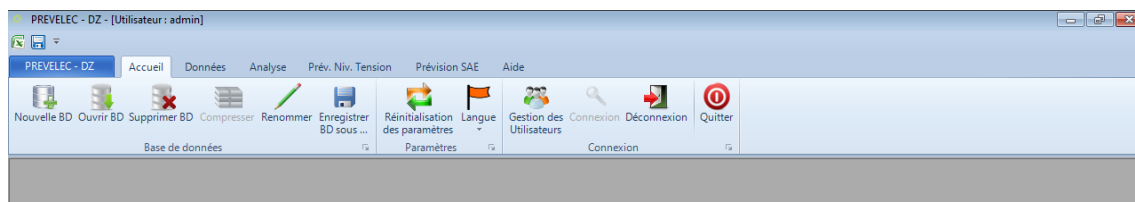


Figure 4.8 – Le menu Accueil.

Le tableau 4.11 résume brièvement la description des fonctionnalités du menu principal.

Lors du lancement de PREVELEC-DZ aucune donnée n'est pré-chargée, pour se faire il faut impérativement créer une nouvelle base de données qui initialement contient l'historique des données de la charge électrique ainsi que des données exogènes, ou ouvrir une base de données existante. Le bouton Nouvelle BD comme son nom l'indique permet la création d'une nouvelle base de données, une fenêtre dialogue apparaît (Figure 4.9) dans laquelle l'utilisateur peut entrer le nouveau nom de la base de donnée (Exemple NouvelEtude1). Entrer le Chemin de sauvegarde en cliquant sur Parcourir. Le bouton Créer

permet de valider la création de la nouvelle étude avec le nom de la BD choisie.

	Fonction	Action
Base De Données	Nouvelle BD	Création d'une nouvelle BD
	Ouvrir BD	Ouverture d'une BD existante
	Supprimer BD	Suppression d'une BD
	Renommer BD	Renommer une BD
	Enregistrer BD sous	Exporter une BD
Paramètres	Réinitialisation des paramètres	Réinitialisation des paramètres
	Langue	Changement la langue de l'interface
Connexion	Gestion des utilisateurs	Gestion des utilisateurs
	Connexion	Connexion à l'application
	Déconnexion	Déconnexion de l'application
	Quitter	Quitter l'application

Tableau 4.11 – Tableau récapitulatif des fonctions du menu Accueil.

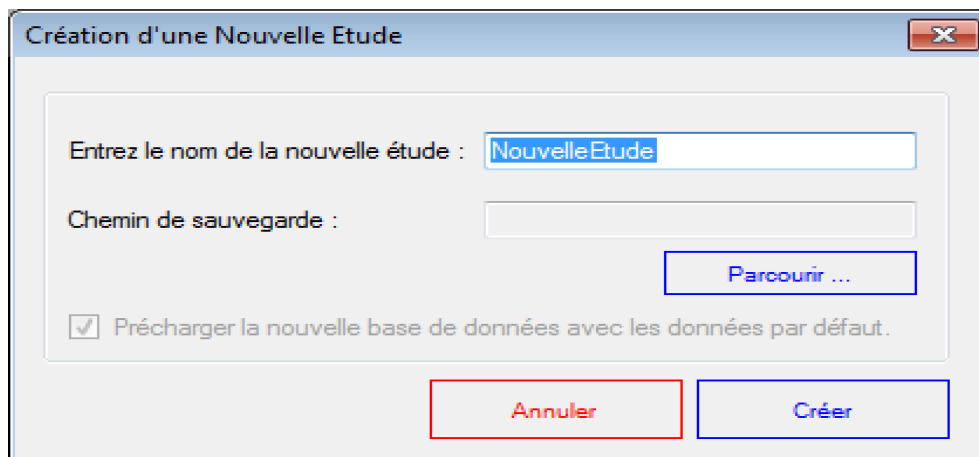


Figure 4.9 – Dialogue de création d'une nouvelle étude.

Dans le cas où l'utilisateur veut ouvrir une base de données déjà existante, il suffit de sélectionner dans le menu la commande Accueil ! Ouvrir BD... Ceci ouvre une fenêtre (Figure 4.10) où l'utilisateur peut sélectionner la base de données avec l'option Parcourir pour choisir le chemin où se trouve le fichier. Ensuite cliquer sur Ouvrir pour ouvrir cette dernière.

La gestion des utilisateurs est un volet important dans ce menu, il permet la connexion et l'utilisation de la solution proposée par un utilisateur. PREVELEC-DZ est par défaut livré avec un nom d'utilisateur et mot de passe qui permettent une première connexion. Nous recommandons fortement le changement de l'utilisateur créer par défaut, pour se faire, il suffit de sélectionner dans le menu la commande Accueil => Gestion des utilisateurs, l'action ouvre une nouvelle fenêtre, comme montré dans la figure 4.11, permettant l'ajout, la suppression et la modification d'un utilisateur.

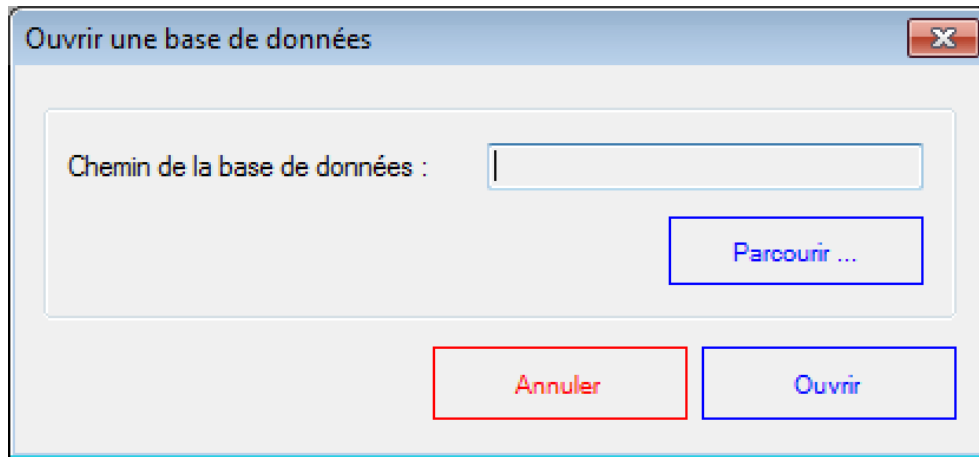


Figure 4.10 – Dialogue pour ouvrir une base de données existante.

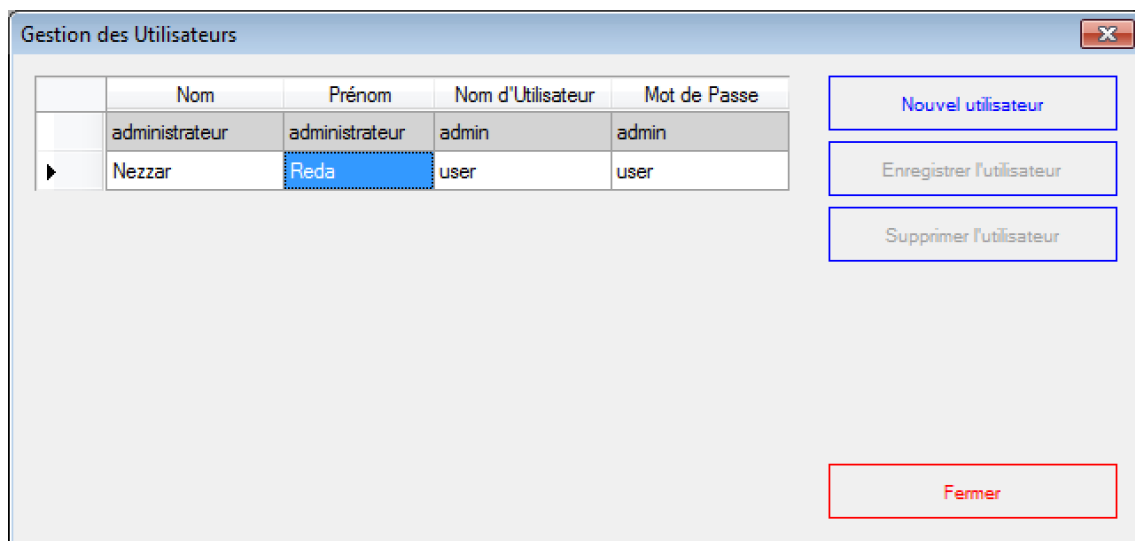


Figure 4.11 – Fenêtre de gestion des utilisateurs.

#### 4.5.2.2 Menu données

Le menu données est le menu dans lequel est centralisée la totalité des données historiques (données réelles) et des variables exogènes nécessaires à l'utilisation des modèles mathématiques en vue d'une prévision. Ce menu se décompose en 4 sections (Figure 4.12) :

1. Historique : Permet de manipuler les données historiques (ajout, modification, suppression d'une année).
2. Energie : Affiche un historique des données en énergie de la Basse Tension (BT), Moyenne Tension (MT) ou Haute Tension (HT).
3. Puissance : permet la visualisation des matrices de courbe de charge des différents réseaux ; RIN, RIS et PIAT pour les années disponibles.
4. Variables exogènes : Permet la gestion de l'ensemble des variables exogènes, qui

représentent les entrées des modèles de prévision.

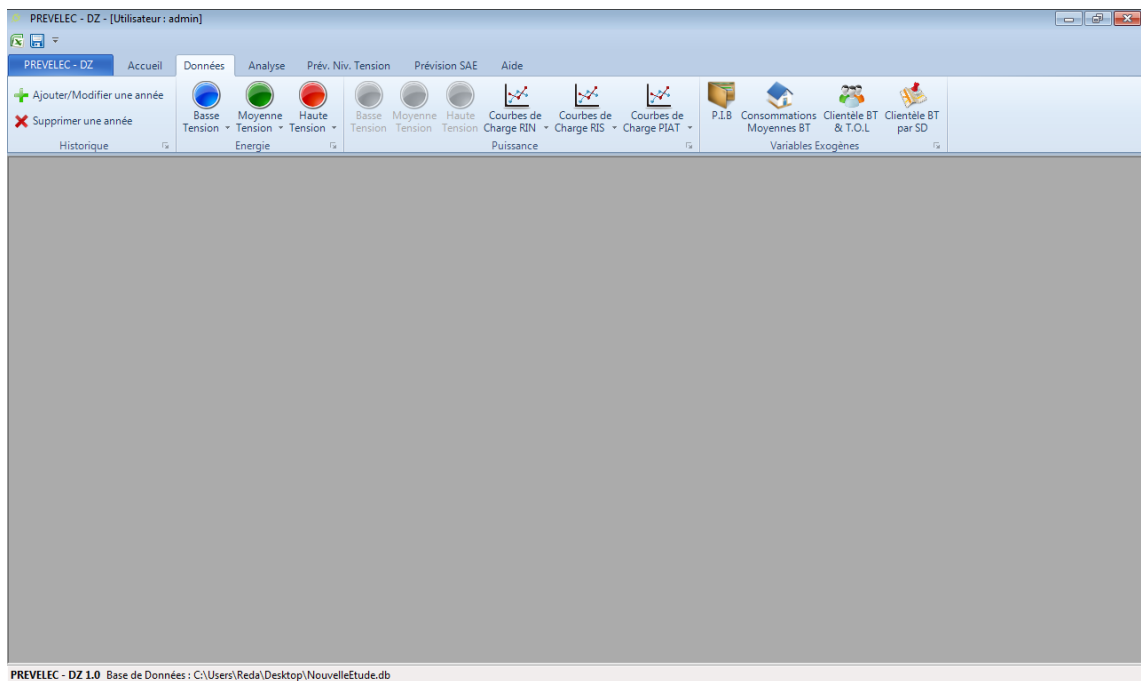


Figure 4.12 – Le menu données.

Le tableau 4.12 montre les actions des boutons du menu données.

	Fonction	Action
Historique	Ajouter/Modifier une année	Permet d'ajouter ou de modifier une année.
	Supprimer une année	Suppression d'une année
Energie	Basse Tension	Affiche données en Energie de la BT
	Moyenne Tension	Affiche données en Energie de la MT
	Haute Tension	Affiche données en Energie de la HT
Puissance	Courbes de charges RIN	Affiche les matrices de courbe de charge RIN
	Courbes de charges RIS	Affiche les matrices de courbe de charge RIS
	Courbes de charge PIAT.	Affiche les matrices de courbe de charge PIAT
Variables exogènes	P.I.B	Visualisation du Produit Intérieur Brut
	Consommation moyenne BT	Affiche les données historiques la consommation moyenne pour la BT
	Clientèle BT & TOL	Calcule la clientèle BT globale et le Taux d'occupation logement
	Clientèle BT par SD	Affiche la répartition de clientèle BT par Société de Distribution

Tableau 4.12 – Tableau récapitulatif des fonctions du menu Données.

L'une des parties les plus importantes de cette section est l'ajout ou modification d'une année. Des données historiques cohérentes permettent l'obtention de résultats satisfaisants lors de la prévision. Une fenêtre s'ouvre (Figure 4.13) lors de l'action du bouton Ajouter/Modifier une année, permettant d'ajouter toutes les informations relatives à une année :

- La consommation Basse Tension par Société de Distribution, ainsi que la consommation des réseaux RIN et RIS.
- La clientèle Basse Tension par société de distribution, ainsi que la clientèle des réseaux RIN et RIS.

- La consommation Moyenne de la BT par société se calcule automatiquement après l'ajout de la consommation et de la clientèle relatives à cette tension.
- La consommation de la Moyenne Tension par secteur d'activité économique.
- La consommation par secteur d'activité économique et par société de distribution de la Haute Tension.
- L'importation de fichiers contenant les matrices de courbe de charge RIN, RIS et PIAT.
- L'ajout des variables exogènes nécessaires à la prévision.

Figure 4.13 – Fenêtre d'ajout et de modification de l'historique.

L'une des fonctionnalités les plus importantes du menu données, est la visualisation de la Matrice de courbe de charge (Figure 4.14). Cette section a pour objectif de montrer les données sans qu'aucune modification ne puisse se faire ; sauf dans le cas d'un ajout d'une année.

La fenêtre clientèle de BT et TOL donne l'historique des données du TOL (Taux d'occupation des logements, population, accroissement, parc logement, accroissement, livraison annuelle et clientèle BT). A noter qu'un client BT est équivalent à un logement électrique. La fenêtre permet de définir les prévisions selon deux méthodes de calculs (TOL Objectif fixé ou Nombre de logements fixé) et 3 scénarios différents (Faible, Moyen

ou Fort). Un exemple avec TOL objectif fixé, scénario moyen est donné par la figure 4.15.

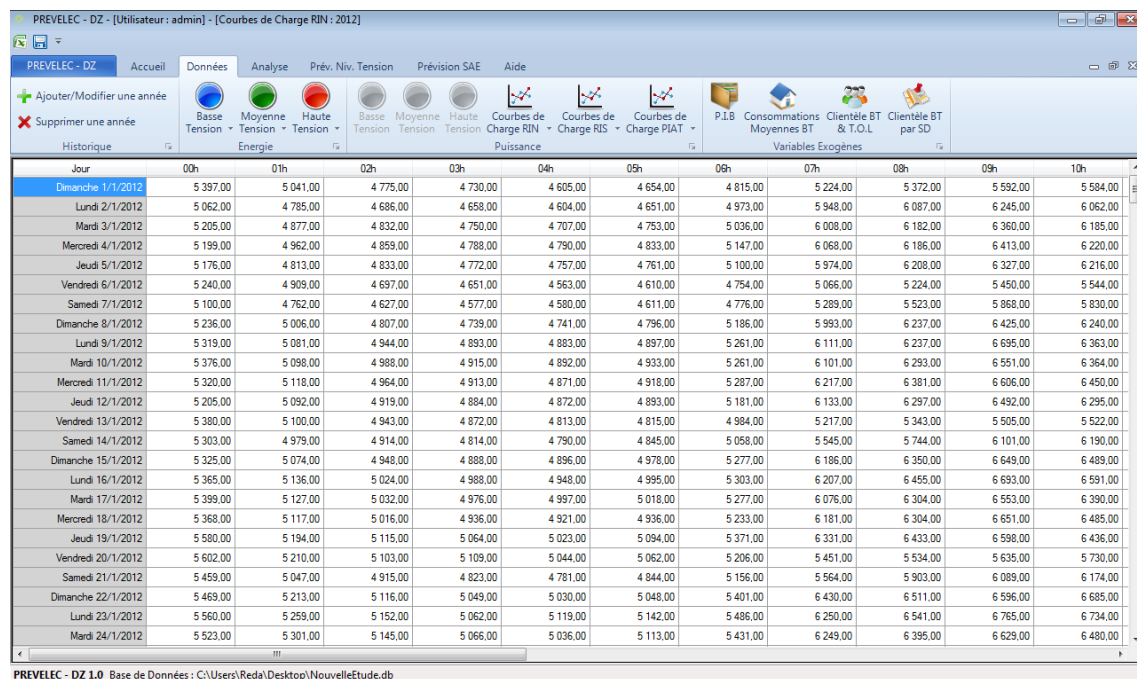


Figure 4.14 – Affichage d’une matrice de courbe de charge.

Dans ce cas, le TOL est fixé par palier de 5 années. En lançant le calcul, tous les autres paramètres seront calculés sur cette base. Par la deuxième méthode de calcul, le nombre de logements est fixé, on pourra alors changer, par palier de 5 ans, la livraison annuelle de logements. En lançant le calcul, tous les autres paramètres TOL inclus seront alors calculés.

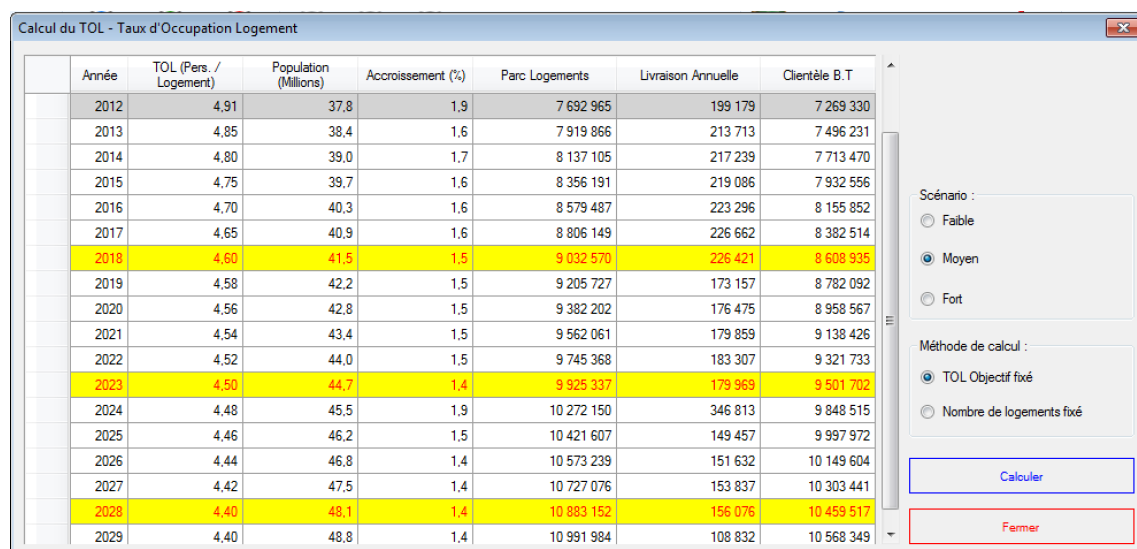


Figure 4.15 – Calcul du taux d’occupation logement.

#### 4.5.2.3 Menu Analyse

Le menu Analyse, est totalement consacré à la gestion des profils. Le menu permet de procéder aux actions suivantes :

- La création de profils (Annuel, hebdomadaire, journalier) manuellement.
- Le calcul des profils appropriés pour une courbe de charge donnée.
- La visualisation des courbes du profil sélectionné.

Il existe deux manières de créer un profil, la première manuellement en cliquant sur le bouton Profils Utilisateur du menu Analyse, la fenêtre montrée dans la figure 4.16 permet de créer les trois types de profils ; journalier, hebdomadaire et annuel en cliquant respectivement sur les boutons Nouveau Profil Journalier, Nouveau Profil Hebdomadaire, Nouveau Profil Annuel. Le profil rempli, peut-être renommé à la guise de l'utilisateur et enregistré pour une utilisation ultérieure.

The screenshot shows a software window titled "Profils de [admin]". On the left, there are three buttons: "Nouveau Profil Journalier", "Nouveau Profil Hebdomadaire", and "Nouveau Profil Annuel". Below these is a text field "Nom du Profil :" containing "NouveauProfilJournalier", followed by a red "Enregistrer" button. A "Liste des Profils :" section contains a scrollable list of profile names like "RIN\_S1VJO\_2012". At the bottom left is a red "Fermer" button. The main area is a table with two columns: "Heure" and "Valeur (%)". The "Heure" column lists times from 0:00 to 18:00 in one-hour increments. The "Valeur (%)" column is empty for all rows.

Heure	Valeur (%)
0:00	
1:00	
2:00	
3:00	
4:00	
5:00	
6:00	
7:00	
8:00	
9:00	
10:00	
11:00	
12:00	
13:00	
14:00	
15:00	
16:00	
17:00	
18:00	

Figure 4.16 – Fenêtre de création et de gestion des profils.

La seconde façon de créer un profil est d'utiliser une matrice de courbe de charge déjà présente dans la base de données. Il suffit juste de choisir le type de réseau, et de spécifier l'année. Ensuite, appuyer sur le bouton Sauvegarde Profil du réseau choisi, PREVELEC-DZ se chargera de nommer les profils générés, à la fin de cette action une confirmation de la réussite de la création des profils s'affiche comme montré dans la figure 4.17



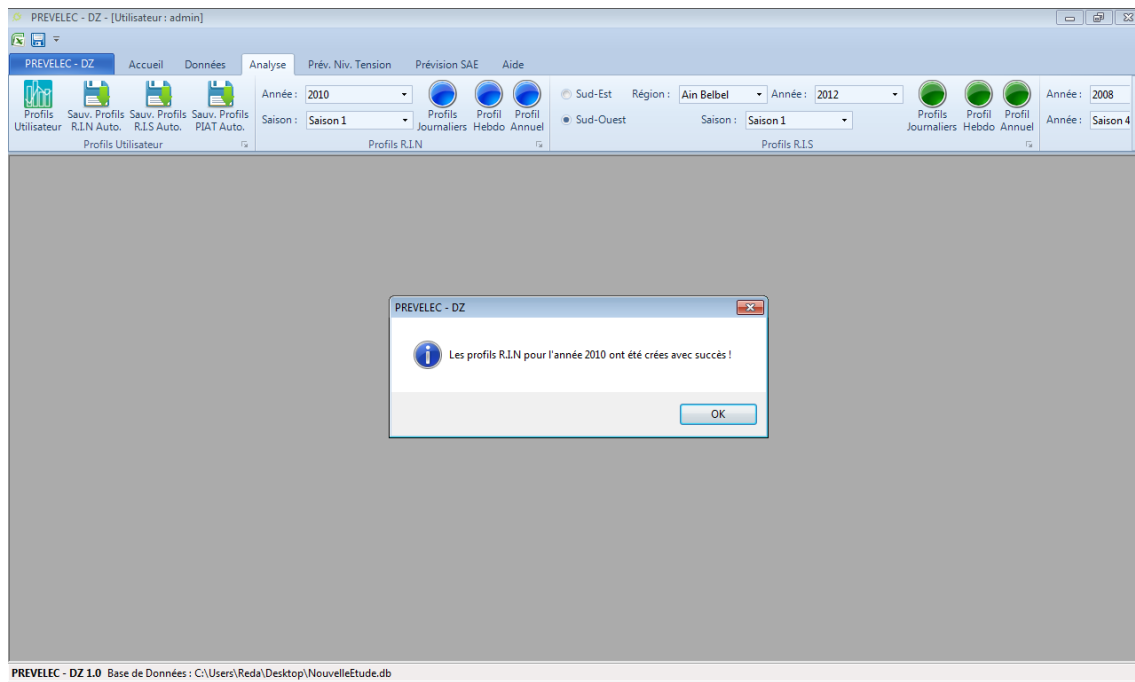
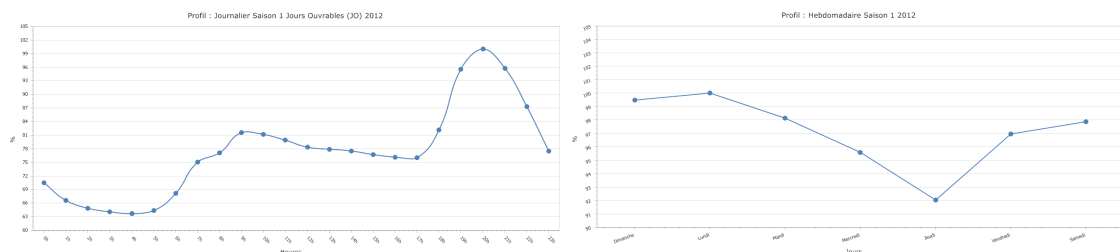


Figure 4.17 – Le menu Analyse.

L'autre aspect de cette section de PREVELEC-DZ est la visualisation des courbes des trois profils, une fenêtre s'affiche avec la/les courbe(s) de la saison de l'année du profil choisi. Les figures 4.18c, 4.18b, 4.18a montrent respectivement le profil annuel, hebdomadaire et journalier de l'année 2012.



(a) Profil journalier du jour ouvrable de saison 1 (b) Profil hebdomadaire de saison 1 de l'année 2012.



(c) Profil annuel de l'année 2012.

Figure 4.18 – Profils de l'année 2012.

#### 4.5.2.4 Menu Pré.Niv.Tension

Cette section centralise tous les outils nécessaires pour mener à bien une prévision pour le RIN et le RIS (Figure 4.19). Elle est subdivisée en :

- Période d’Historique Affichée : Permet de modifier la période d’historique affichée lors du lancement d’une prévision.
- Horizon de prévision : Permet de sélectionner l’année de début et l’année de fin d’un exercice de prévision.
- Basse Tension (BT) RIN : Lance une prévision basse tension pour le RIN.
- Moyenne Tension (MT) : Lance une prévision moyenne tension pour RIN.
- Haute Tension (HT) : Lance une prévision haute tension pour RIN.
- Total RIN en énergie et en Puissance.
- Réseau Isolé du Sud (RIS).
- Pôle In Salah-Adrar-Timimoune (PIAT).

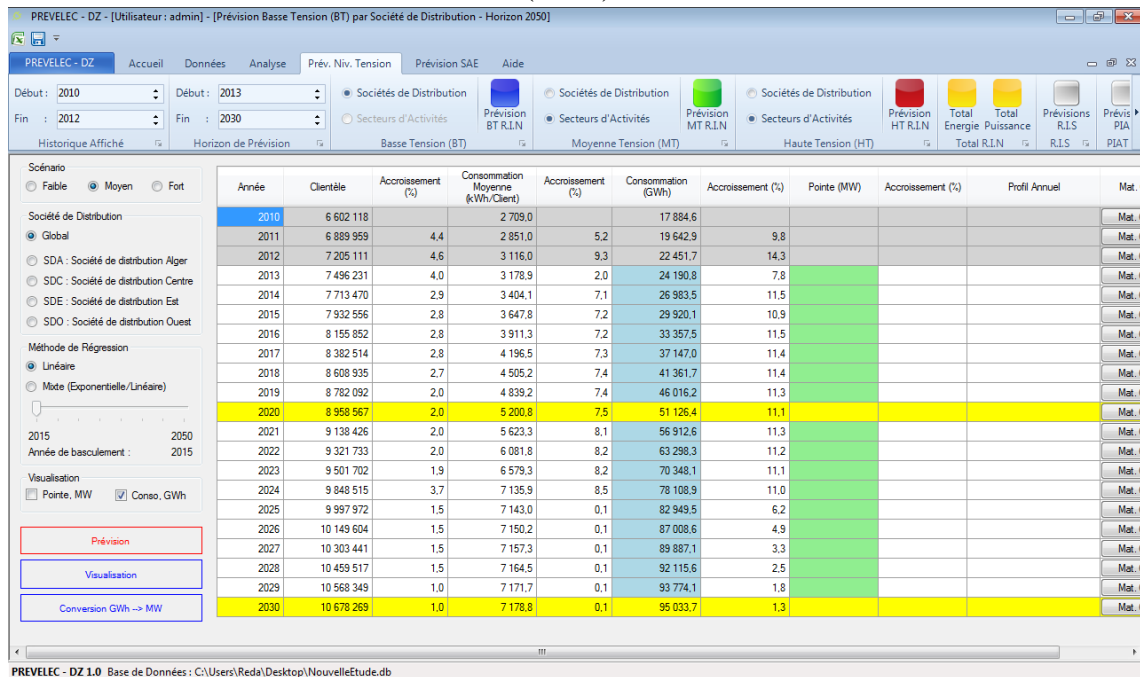


Figure 4.19 – Menu prévision par niveau de tension.

Avant de commencer une prévision, il faut choisir le début et la fin de l’historique que l’utilisateur veut utiliser dans son analyse, ainsi que, le début et la fin de l’horizon de prévision. En cliquant sur Prévision BT RIN, la prévision se lance avec une fenêtre permettant à l’utilisateur de choisir à gauche le scénario de prévision (Faible, Moyen ou Fort), la société de Distribution (Global, SDA : Société de distribution d’Alger, SDC : Société de distribution Centre, SDE : Société de distribution Est ou SDO : Société de distribution Ouest), la méthode de régression (Linéaire ou Mixte) ainsi que le type d’éléments à visualiser (Pointes MWh ou Consommation GWh ou les deux en même temps).

Pour faire une conversion de l'énergie GWh vers la puissance MW, il est impératif de remplir les champs de la colonne Profil Annuel en sélectionnant un profil sauvegardé, et faire une conversion grâce au bouton Conversion GWh -> MWh. Une visualisation de la consommation (en énergie ou puissance ou les deux) est possible en cliquant sur le bouton visualisation. La figure 4.20 montre la prévision du scénario moyen de la BT jusqu'à l'année 2030.

Chaque Tension permet de faire, en fonction des données disponibles, une prévision globale et par société de distribution et/ou par secteur d'activité. Le principe reste le même pour toutes les tensions. Les prévisions sont réalisées sur trois scénarios possible : faible, moyen et fort pour toutes les tensions. Le même menu permet de faire une prévision totale en Energie et en Puissance du RIN respectivement avec les boutons Total Energie et Total puissance, les deux fenêtres vous invitent à charger les prévisions par tension, préalablement faite, afin d'y ajouter les pertes de transport et distribution pour avoir la puissance à produire. Evidemment l'aspect graphique est présent dans les deux phases.

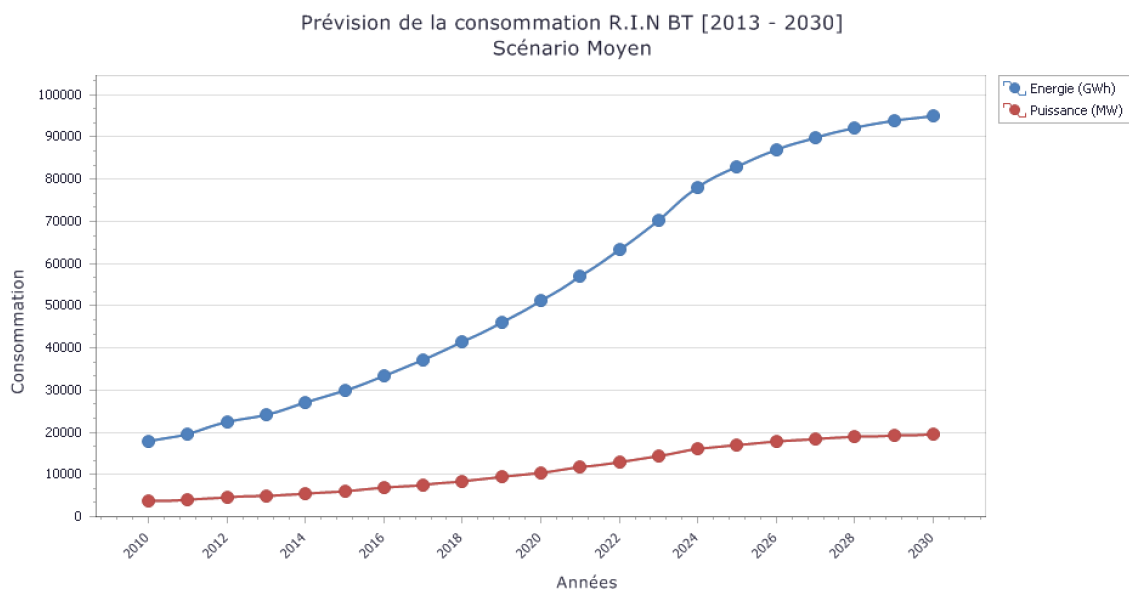


Figure 4.20 – Exemple de visualisation de la prévision R.I.N de la BT jusqu'à l'horizon 2030.

## 4.6 Conclusion

Dans ce chapitre, une analyse de la charge électrique algérienne a été réalisée, en étudiant les données soumises par la société SONELGAS et en produisant plus d'une quarantaine de modèles linéaires et exponentiels qui ont été proposés dans cette partie l'étude. Pour répondre aux attentes de notre partenaire, un mode mixte entre régression exponen-

tielle et régression linéaire a été proposé afin de répondre aux problèmes de manque de données ainsi qu'au taux de croissance élevé des habitations en cours de construction ces dernières années et par conséquent une clientèle plus importante.

Le projet réalisé permet de faire une prévision globale de la charge électrique algérienne sur le long terme (jusqu'à l'horizon 2030) par énergie et/ou par puissance. Ainsi, qu'une prévision par secteur d'économie ou par société de distribution. L'incorporation du system des profils, qui permet la ventilation d'une charge électrique annuelle en une charge horaire, permet aussi d'obtenir la monotonie de charge. Les profils sont paramétrables par saison ou par mois selon les besoins, ils permettent d'avoir une meilleure perspective sur les pics de tension pendant l'année prédite.

L'étude proposée dans ce chapitre a été concrétisée par le développement d'un outil nommé PREVELEC-DZ, qui propose différentes fonctionnalités : gestion des utilisateurs, sauvegarde des prévisions réalisées, visualisation et ajout des données historiques de la charge électrique et des données macro-économiques, ainsi que l'implémentation des modèles régressifs et profils. L'outil est développé avec C sharp pour une meilleure intégration à la Platform Windows

## **Chapitre 5**

# **Approches multi-modèle pour la prévision de la charge électrique à moyen long terme**

## 5.1 Introduction

Les pays en croissance et les industries modernes dépendent entièrement de la disponibilité et de la capacité des centrales électriques ; qui influent directement sur l'économie d'un pays. Par conséquent, avoir une prévision à moyen long terme pour les industries énergétiques est une donnée d'entrée cruciale pour toutes les études sur le développement du marché de la production et de la consommation d'énergie. La prévision de la charge est une étape, dans le processus de planification des travaux d'énergie, qui détermine les investissements impliqués, ainsi que les choix irréversibles sur le nombre de centrales à développer ou à activer et enfin le nombre de turbines et d'autres sources de production d'électricité à installer afin de satisfaire la demande nationale et / ou régionale. En effet, faire une prévision précise est une question de la première importance [51].

La prévision de la charge peut être effectuée par une approximation de fonction non linéaire, en utilisant diverses variables telles que la météo, le PIB, la croissance de la population, etc. ainsi que plusieurs autres entrées exogènes, qui ont une influence directe sur la consommation ou la production électrique. Ce genre de modèles sont appelés multi-variée, ils impliquent la prévision de chaque variable incluse dans le modèle, pour projeter la variable étudiée qui est dans notre cas la charge électrique, cela reste une tâche complexe et cette complexité s'accroît avec l'augmentation des variables exogènes. C'est pour cette raison qu'un autre type de modèles a vu le jour, et qui est basé sur l'historique de la charge électrique afin de réduire la complexité de l'élaboration du modèle, ce genre de modèles sont appelés uni-variés.

Dans cette partie de la thèse, l'accent est mis sur la prévision de la charge à moyen long terme, car elle influence directement les décisions d'investissements énergétiques majeurs, tels que le développement de nouveaux plans d'alimentation. Traditionnellement, les modèles statistiques tels que la régression linéaire [52], [53] et les modèles autorégressifs [54] ont été largement utilisés dans la pratique pour la prévision de la charge, principalement en raison de leur simplicité et des bonnes performances de ce type de modèles [55]. Toutefois, ces dernières années, de nombreux autres modèles ont été proposés. Parmi ces méthodes ; l'intelligence artificielle (AI) qui offre des solutions intéressantes. Les techniques de l'IA les plus couramment utilisées sont l'ANNs [56], [25], [57] , la logique floue [24], [58], [59], l'intelligence en essaim [60], [28], [29], la théorie Grey [61], [62], les réseaux d'ondelette [63], [64], les machines à vecteurs de support, plus communément connues sous le nom de SVM [65], etc.

Dans cette partie de l'étude proposée, un modèle univarié a été conçu, de manière à réaliser une prévision de la charge à moyen long terme pour l'industrie de l'énergie

algérienne qui est en croissance mais d'une manière irrégulière. Le modèle est basée sur différents types de réseaux de neurones ainsi que sur la décomposition saisonnière. La combinaison mentionnée ci-dessus se trouve être de meilleure performance que d'autres approches qui ont déjà prouvé leur efficacité.

Dans ce chapitre, nous allons parler de la problématique et de la contribution proposée pour répondre aux problèmes liés à la complexité de la prévision d'une série temporelle fortement corrélée, ainsi que l'obtention de résultats satisfaisants avec un ensemble de données qui est petit et inadéquat pour une prévision à moyen- long terme, en utilisant particulièrement les méthodes computationnelles qui sont généralement gourmandes en données d'apprentissage.

## 5.2 Problématique et contribution

La prévision est nécessaire à la prise de décisions et à la planification des tâches opérationnelles. La difficulté de ces tâches varie en fonction de l'horizon de prévision, et dépend fortement de la prévision des variables exogènes affectant la future charge électrique. Les modèles développés dans ce sens sont appelés multivariés. Les méthodes multivariées tiennent compte des variables exogènes affectant la charge électrique ; la difficulté de la prévision augmente avec le nombre de variables, en particulier pour l'horizon moyen et long terme. Pour cette raison, d'autres modèles dits uni-variés ont été proposés ; ils utilisent exclusivement des données passées de la charge pour la tâche de prévision. Les méthodes univariées ne dépendent pas des variables exogènes qui ne sont pas toujours disponibles, et doivent être parfaitement prédites à leurs tours, ce qui rend la tâche de prévision plus difficile.

Les séries temporelles de la charge sont fortement corrélées. Trouver des facteurs externes qui influent sur la charge est généralement une tâche difficile, d'autant plus que pour obtenir des résultats satisfaisants, en utilisant les méthodes de l'intelligence artificielle et particulièrement les réseaux de neurones artificielles, un grand échantillon de données sera nécessaire pour la partie d'apprentissage et la partie de test pour valider les performances d'un modèle computationnel, ce qui représente un handicap préalable à notre étude vu la taille des données fournies par notre partenaire Sonelgaz qui représentent 13 années et donc 13 points ; ce qui est extrêmement peu pour obtenir des résultats satisfaisants.

la prévision présentée dans cette partie de l'étude, est une prévision de la puissance produite sur le moyen long terme, cet horizon implique généralement l'utilisation de mois

ou d'années comme unité de temps. Comme le nombre d'années est insuffisant, l'utilisation des mois reste une alternative en re-échantillonnant les matrices de courbe de charge en mois pour obtenir un total de 156 observations à diviser entre apprentissage et test pour le modèle, ce qui reste insuffisant pour les réseaux de neurones qui demandent plusieurs centaines de données pour la phase d'apprentissage, en plus des problèmes liés au nombre de données, le problème du pré-traitement de ces derniers qu'on appelle "Whitening", qui demande des connaissances supplémentaires en ce qui concerne les modèles utilisés, ce qui reste une tâche difficile pour des utilisateurs non experts.

Par conséquent, l'objectif principal de l'étude proposée, est d'effectuer une prévision de la charge qui soit précise et fiable avec un minimum de modifications des données originales pour concevoir un modèle qui peut répondre à une prévision avec un petit échantillon de données. Cet objectif est atteint en utilisant une combinaison de la décomposition saisonnière et des réseaux neuronaux, le modèle est appelé X12-NAR-FFNN ; ce dernier utilise les données mensuelles pour la prévision de la charge algérienne à moyen long terme.

Le modèle proposé est un procédé en trois étapes, où la charge initiale passe à travers chaque étape et la charge est modifiée en conséquence, et la sortie d'une étape est l'entrée de la suivante. Trois méthodes sont utilisées dans le modèle X12-NAR-FFNN : la décomposition X12-ARIMA, Réseau de neurones autorégressif (NAR) et un réseau neuronal Feed Forward.

Ces méthodes sont organisées comme suit (Figure 5.1) :

- Une décomposition des données originales en trois composantes : Fluctuation saisonnière (SF), Irrégulière (IR) et Tendence-Cycle (TC) via le modèle multiplicatif de l'algorithme X12-ARIMA, dans le but de nettoyer les données et faciliter le processus de prévision. Ce dernier est complexe par la nature des séries et le peu de données disponibles.
- Une prévision mensuelle des composants : Fluctuation saisonnière (SF), Irrégulière (IR) et Tendence-Cycle (TC) est réalisée par l'intermédiaire de trois réseaux neuronaux NAR ; les 12 derniers mois de chaque composant sont utilisés comme entrée pour chaque réseau de neurones pour capturer le pattern annuel. À la fin de cette étape, on obtient les résultats des prévisions mensuelles de SF, TC et IR pour une année complète. Chaque réseau NAR est composé de 12 neurones d'entrées, 12 neurones dans la couche cachée, et un neurone de sortie.
- Enfin, le Feed Forward ANN combine les sorties des trois réseaux NAR pour les 12 mois à venir de la charge. Il est composé de trois neurones pour la couche d'entrée correspondant à chaque composante préalablement prédite par les réseaux



NAR respectifs, 6 neurones dans la couche cachée et 1 neurone pour la couche de sortie, symbolisant la prévision du mois de la production électrique.

Tous les modèles proposés dans cette étude ont été créés à l'aide des données allant de 2000 à 2010, qui nous ont fourni 132 observations mensuelles, qui sont utilisées pour la formation et la validation des phases, alors que le reste des observations (à partir des années 2011 et 2012) sera utilisé pour la phase de test. La phase d'apprentissage de tous les réseaux de neurones dans cette étude a été réalisée avec l'algorithme "Resilient Back-propagation" connu pour sa rapidité de convergence [66], et pour ce qui est de la mise à l'échelle des données, une normalisation "minmax" (5.1) est appliquée pour toutes les séries temporelles utilisées dans les réseaux sélectionnés.

$$X_{i,norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (5.1)$$

où  $X_{max}$  et  $X_{min}$  sont respectivement le maximum et le minimum de la série à normaliser.

## 5.3 Résultats et Expérimentation

Dans cette partie de la thèse, nous allons aborder les résultats et les étapes entreprises pour la prévision de la puissance en utilisant les matrices de courbes de charge, dans un horizon de moyen long terme c'est-à-dire de 12 à 24 mois, en implémentant le modèle décrit plus haut.

### 5.3.1 Pré-traitement et données utilisées

L'ensemble des données a été fourni par la société de production d'électricité algérienne (Sonelgaz) et représente la charge quotidienne, du 1er Janvier 2000 au 31 Décembre 2012, échantillonnée toutes les 15 minutes. Pour effectuer une bonne prévision à moyen long terme, et suite aux conclusions du chapitre 4, il est préférable d'utiliser des données mensuelles à la place des données annuelles. Par conséquent, il est nécessaire d'effectuer un ré-échantillonnage des matrices de courbe de charge journalière, en prenant le pic de chaque mois des données d'origine. Au final, on obtient un total de 156 observations, avec 12 valeurs pour chaque année, et un total de 156 valeurs pour les 13 années utilisées.

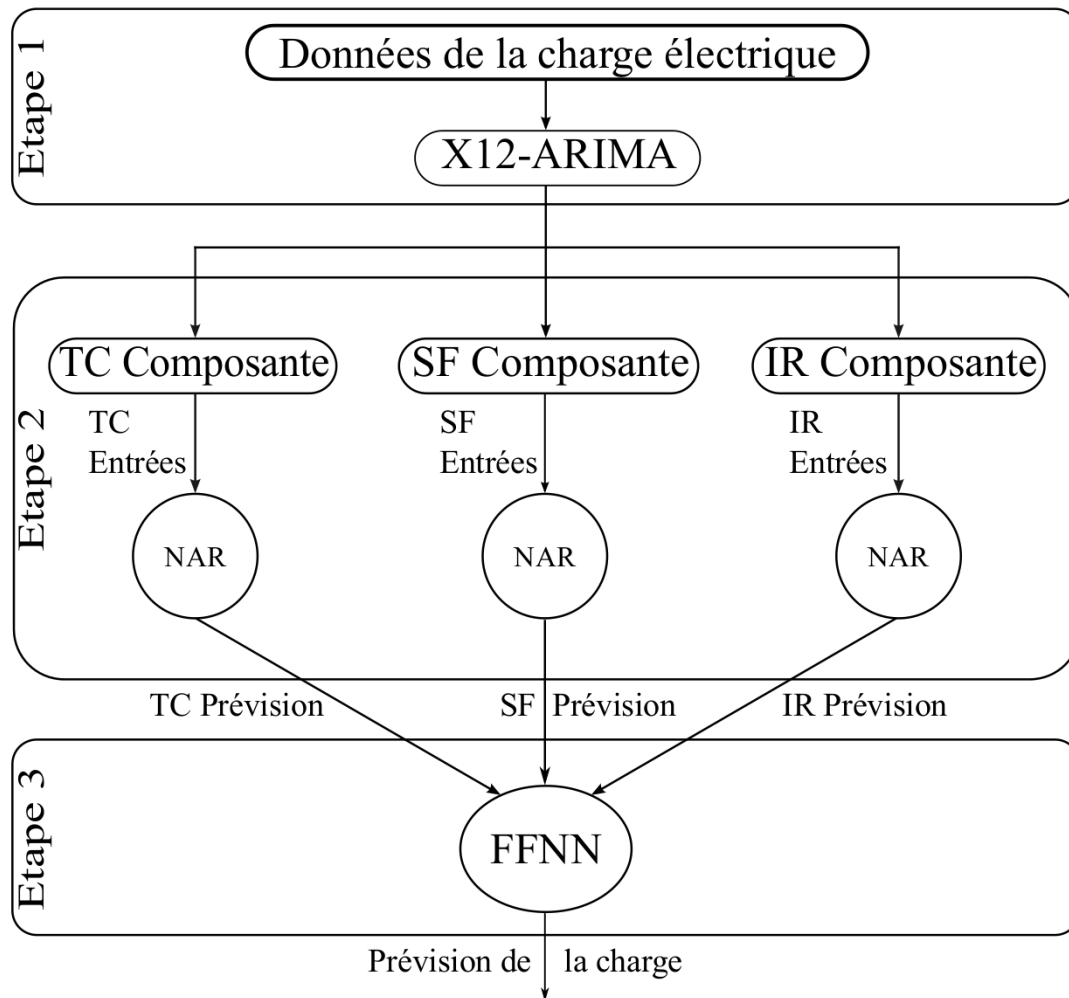


Figure 5.1 – Le modèle X12-NAR-FFNN.

La décomposition classique est souvent utilisée, mais elle souffre de plusieurs inconvénients, qui sont ; pas d'estimation de tendance pour les premières et les dernières observations, la décomposition réduit les données d'origine, ce qui devient inapproprié quand il y a quelques observations et on suppose que la composante saisonnière se répète d'année en année, ce qui n'est pas une hypothèse raisonnable pour de nombreuses séries. Plusieurs études ont utilisé la décomposition classique comme base pour surmonter les inconvénients de l'approche originale ; parmi les méthodes développées : X12-ARIMA, RMA, etc.

Le modèle X12-ARIMA a été développé par le Bureau Américain du Recensement (1998) [16] et est considéré dans ce travail comme une méthode de décomposition mensuelle des données de la charge, ce qui nous permet d'obtenir une estimation de tendance pour toutes les observations et une composante saisonnière qui peut varier lentement au fil du temps, avec une robustesse par rapport aux observations inhabituelles. Cette méthode est basée sur un modèle appelé régression-arima, et est utilisée pour améliorer la prévi-

sion et l'ajustement saisonnier. Ensuite, le modèle X11 [14] gère les sorties du modèle de régression ARIMA pour extraire TC, IR et SF de la série chronologique.

L'algorithme X12-ARIMA est implémenté sur différentes plateformes, qu'elles soient payantes comme Eviews et procX12 ou gratuites comme dans gretl, X-13-ARIMA-SEATS et x12 qui est une bibliothèque R. Le processus est simple, cependant quelques paramètres sont nécessaires pour obtenir une décomposition ; comme la méthode de décomposition (additive ou multiplicative), la transformation de la série d'origine, l'intégration des jours fériés, etc.

Dans notre étude, nous avons utilisé la méthode de décomposition multiplicative de la série mensuelle de la charge électrique algérienne comme montré dans la figure 5.2 où l'on remarque que la saisonnalité change d'une année à l'autre et plus on avance dans les années, plus la variance entre le minimum et le maximum de chaque année est plus grande. En 2000, la différence était de 1067 MW pour arriver à 2700 MW en 2012 ; en analysant les pics de production électrique, on constate qu'un changement important est constaté après l'année 2008 (où le pic de production était généralement vers la fin de l'année avant cette date), où le pic est passé au mois d'Août, cela étant expliqué par la démocratisation des climatiseurs personnels à cause des grandes chaleurs de ces dernières années.

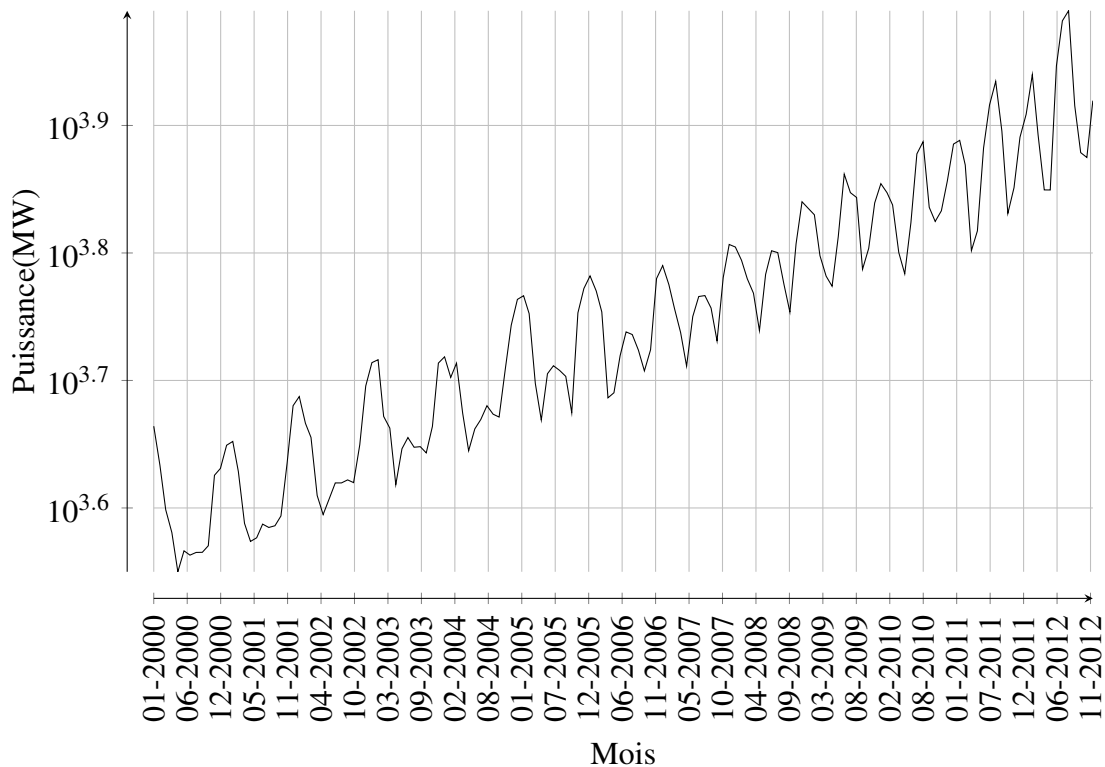


Figure 5.2 – Évolution de la production en puissance algérienne entre 01/2000 et 12/2012.

La production électrique algérienne n'a pas été suffisante pendant la période étudiée,

ce qui a engendré des périodes de délestage pour pouvoir satisfaire la demande surtout dans les périodes estivales. La nature changeante de la série temporelle de la charge électrique algérienne, qui est dûe a plusieurs variables qui sont : le réchauffement climatique, la consommation des ménages algériens, la livraison d'une grande quantité d'habitations et la bonne santé de l'économie algérienne, rend la modélisation de cette série très complexe si une intégration des différents facteurs cités est effectuée. Même dans le cas d'un modèle uni-varié ceci reste très difficile sans la décomposition saisonnière, qui nous donne 3 séries pour chaque composante (SF,IR,TC) comme le montre la figure 5.3, la composante TC nous permettra d'avoir une meilleure approximation du niveau de la série globale, la composante SF nous aidera à avoir une vision sur les fluctuations saisonnières à venir et la composante IR qui par nature est irrégulière et donc difficile à prévoir.

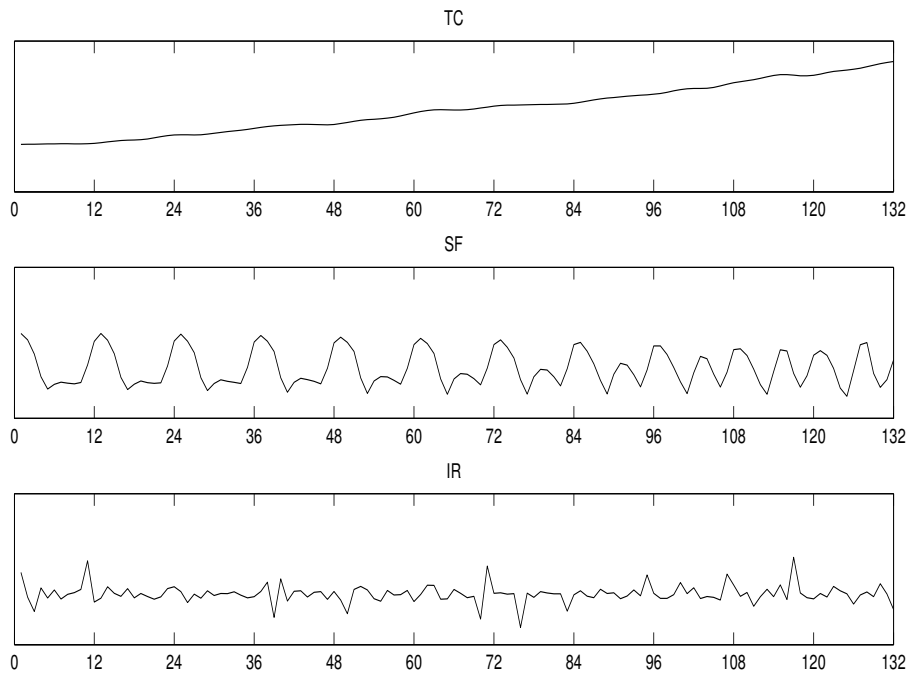


Figure 5.3 – Exemple de décomposition X12-ARIMA.

### 5.3.2 Mesures de performance

Une comparaison du modèle proposé X12-NAR-FFNN est réalisée dans cette étude avec d'autres modèles, tels que NAR et Least Squares Support Vector Regression (LS-SVR), multi-modèles comme SD(M)-LSSVR-LSSVR [31], SD(M)-LSSVR et X12-NAR. Elle a été basée sur différents types de critères parmi lesquels : l'erreur absolue moyenne en pourcentage (MAPE), et la mesure de la précision pour la prévision des valeurs des séries temporelles. MAPE est utilisée pour calculer la précision de la prédiction de chaque

modèle et est définie par la formule dans l'équation 5.2 :

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \quad (5.2)$$

où  $x_i$  est la valeur réelle et  $\hat{x}_i$  est la valeur prédite. La différence entre  $x_i$  et  $\hat{x}_i$  est divisée par la valeur réelle  $x_i$ . La somme de la valeur absolue pour chaque point est prédite et à nouveau divisée par le nombre de points  $n$  équipés. Ce nombre est multiplié par 100, étant donné un pourcentage d'erreur. Pour obtenir une étude complète et de valeur de la prédiction de tous les modèles testés, d'autres indices d'erreur ont été ajoutés, consistant en : Erreur Root-Mean-Square (RMSE) [39] (Equation 5.3) et Erreur Absolue Moyenne (MAE) [67] (Equation 5.4).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (5.3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (5.4)$$

Où  $\hat{y}_i$  et  $y_i$  sont respectivement la prévision et la valeur réelle,  $n$  est le nombre de points dans la série temporelle. En plus du calcul de la précision de prévision. Une autre comparaison des modèles est réalisée sur leurs habilités à prévoir la direction du mouvement de la courbe prédite, en utilisant le test de statistique directionnel  $D_{stat}$ , elle est formulée dans l'équation suivante :

$$D_{stat} = \frac{1}{n} \sum_{i=1}^n a_i * 100 \quad (5.5)$$

où  $a_t = 1$  quand le produit  $(x_{i+1} - x_t)(\hat{x}_{i+1} - x_t) > 1$ , sinon  $a_t = 0$ .

## 5.4 Discussion et Expérimentation

Afin de valider les résultats et la fiabilité du modèle X12-NAR-FFNN, nous allons effectuer une comparaison avec le modèle proposé par Wang et al. [31], qui prédisent la consommation hydroélectrique en Chine. Le modèle de référence est basé sur la décomposition saisonnière (SD) avec l'algorithme X12-ARIMA et la méthode des Least

Square Support Vector Regression (LSSVR). Le modèle, appelé SD(M)-LSSVR-LSSVR, décompose d'abord la série hydroélectrique en trois composantes, qui sont, à leurs tour, prédites en utilisant trois modèles LSSVR et un quatrième LSSVR combinant les prédictions des trois composantes pour obtenir la révision de la consommation finale. Les auteurs ont démontré l'efficacité de la méthode X12-ARIMA en la comparant avec d'autres méthodes de décomposition, tels que le ratio de la moyenne mobile (RMA) et la décomposition modale empirique (EMD). En outre, les auteurs ont constaté que la performance du modèle diminue lorsque l'horizon de prévision est grand.

Par conséquent, pour effectuer une comparaison du modèle proposé X12-NAR-FFNN, une sélection de modèles basée sur deux approches a été utilisée :

- Les modèles basés sur les réseaux neuronaux ; un modèle simple régressif (NAR) et un modèle X12-NAR, qui décomposent la série temporelle originale en trois composantes (à savoir TC, SF, et IR) grâce à l'algorithme X12-ARIMA. Ensuite, chaque composant est prédit à l'aide du modèle NAR. Enfin, pour obtenir la charge finale, une multiplication des composants prédits est effectuée : les NAR ont 12 neurones d'entrée pour capturer le pattern annuel de la composante saisonnière et 12 neurones dans la couche cachée, et un neurone de sortie ; la fonction d'activation sigmoïde et l'algorithme de propagation "resilient" ont été utilisés pour tous les modèles neuronaux.
- Les modèles basés sur la méthode LSSVR ; un modèle de LSSVR simple et un modèle X12-LSSVR, qui suit la même approche par rapport à celle utilisée pour le modèle X12-NAR. Cependant, au lieu de réseaux de neurones, le modèle LSSVR a été utilisé en plus du modèle de référence appelé SD(M)-LSSVR-LSSVR. Un noyau gaussien RBF est utilisé pour tous les modèles de LSSVR présentés dans cette étude.

Une comparaison entre les modèles neuronaux et les modèles de LSSVR est effectuée. Ces derniers ont prouvé leurs efficacités et leurs capacités d'apprendre et de prévoir sur un horizon de moyen long terme, avec un petit échantillon de données [31]. Après ré-échantillonnage de la charge quotidienne algérienne en mois, les modèles NAR et LSSVR ont été formés en utilisant les données de Janvier 2000 à Décembre 2010. Les méthodes d'évaluation, avec les mesures de performance décrites ci-dessus, ont été effectuées sur plusieurs horizons ; 3, 6, 12 et 24 mois pour comparer la capacité des modèles à prévoir sur différents horizons.

L'expérimentation a été réalisée en trois parties ; dans chaque partie, le réseau de neurone et le paradigme LSSVR sont comparés. La première partie évalue le modèle NAR et LSSVR en utilisant MAPE, RMSE et l'erreur MAE, ces dernières donnent les erreurs

de 4,44%, 557,11, 377,05 respectivement pour le LSSVR, ce qui est plus efficace que le modèle NAR, qui atteint des erreurs respectives de 5,40%, 626,13 et 456,13 pour un horizon de 24 mois (voir le tableau 5.1). Le modèle de LSSVR est plus efficace que la NAR dans le moyen long terme, car il conserve le principe de la SRM et possède de bonnes capacités de généralisation, bien que, les deux ont la même  $D_{stat}$ , avec un pourcentage de 78,28% (Figure 5.4).

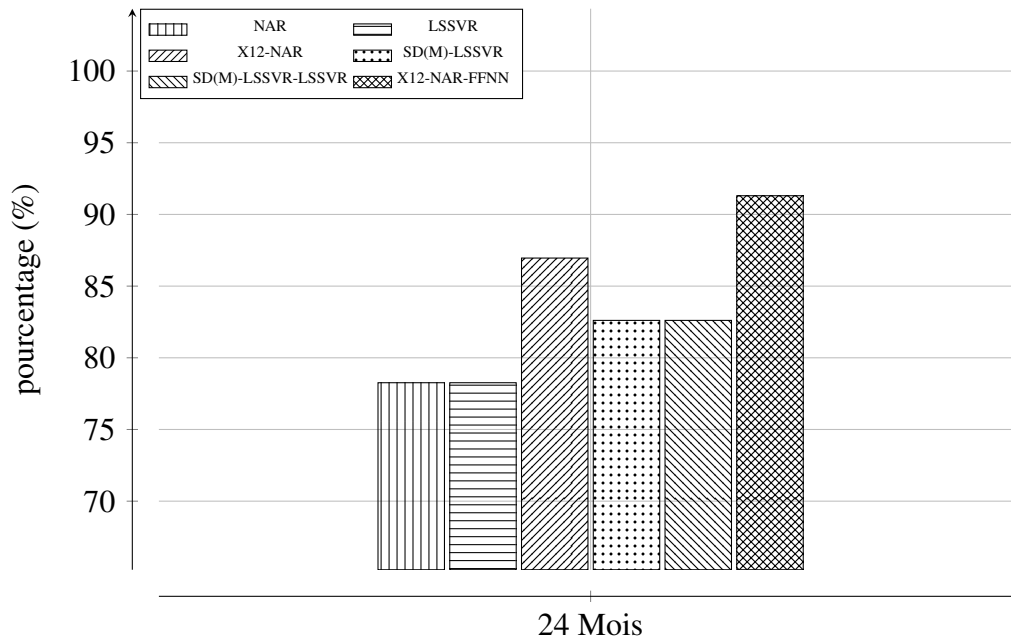


Figure 5.4 –  $D_{stat}$  des modèles pour l’horizon 24 mois.

La deuxième partie de l’expérimentation est basée sur la décomposition de la charge en Tendence-Cycle, la Fluctuation saisonnière et la composante irrégulière, en utilisant le modèle multiplicatif de l’algorithme X12-ARIMA et sa valeur ajoutée par rapport à un modèle simple. Par la suite, trois modèles NAR et LSSVR ont été utilisés pour prédire TC, SF et IR appelés respectivement X12-NAR et SD(M)-LSSVR. Les sorties de la suite sont le résultat du produit des trois composants obtenus à partir des modèles NAR et LSSVR, avec 2,49%, 250,22, 199,86 pour les erreurs MAPE, RMSE et MAE, respectivement, et un  $D_{stat}$  de 86,9% pour l’horizon de 24 mois. Le X12-NAR est plus précis que SD(M)-LSSVR, qui atteint une erreur de 3,08%, 353,40 et 254,46 pour les mêmes critères, et un  $D_{stat}$  de 82,6%.

La décomposition de la charge permet aux modèles basés sur les réseaux de neurones, qui sont connus pour être gourmands en données, d’être efficaces avec un petit ensemble de données d’apprentissage avec des résultats de près de 3% de plus en erreur de MAPE que les modèles simples NAR (Figure 5.6). Le X12-NAR surpasse le SD(M)-LSSVR sur tous les horizons, comme indiqué dans le tableau 5.1, et est plus efficace pour prédire la direction du mouvement des données.

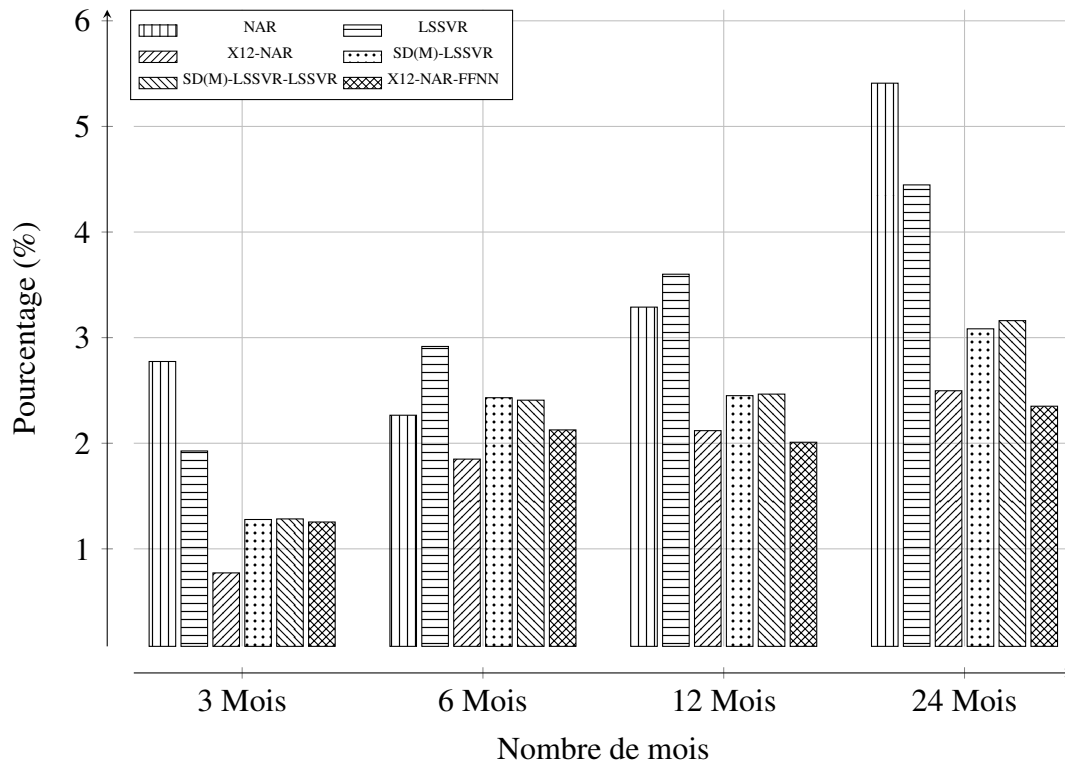


Figure 5.5 – Les erreurs MAPE de tous les modèles sur les différents horizons de prévision.

Les résultats montrent que la décomposition X12-ARIMA a augmenté la précision de la prévision en utilisant un petit ensemble de données d'apprentissage. L'ajout d'un modèle de LSSVR pour combiner les trois composants n'améliore pas les performances de prédiction et  $D_{stat}$  est à peu près égale à celle de SD(M)-LSSVR-LSSVR. Contrairement à ce dernier, le modèle X12-NAR-FFNN améliore l'erreur de prédiction par rapport à X12-NAR pour obtenir une valeur MAPE, RMSE et MAE respectivement de 2,35%, 270,95, 189,65 et  $D_{stat}$  de 91,30%, ce qui en fait le modèle le plus performant sur tous les horizons. Les figures 5.3 et 5.5 résument respectivement l'erreur des critères MAPE cités ci-dessus et  $D_{stat}$  de tous les modèles pour un horizon de 24 mois, et montrent clairement que le modèle X12-NAR-FFNN surpasse en termes d'erreur et de mouvement directionnel les autres modèles utilisés dans la présente étude.

La prévision de la charge de Janvier 2011 à Décembre 2012, donnée par le modèle X12-NAR-FFNN, est représentée sur la figure. 5.6. Le modèle est validé en comparant la prévision par rapport à la charge réelle, y compris l'augmentation observée à la fin 2012 en raison de la croissance à la fois du secteur du logement et des industries algériennes. Les réseaux de neurones ont prouvé leur efficacité dans un certain nombre de scénarios dans les domaines de la prévision de la charge électrique, mais comme tout modèle, ils ont des avantages et des inconvénients. Pour cette raison, la tendance évolue vers les multi-modèles afin de combiner les avantages de différents modèles. Le modèle X12-



		<b>X12-NAR-FFNN</b>	<b>X12-NAR</b>	<b>SD(M)-LSSVR</b>	<b>SD(M)-LSSVR-LSSVR</b>	<b>LSSVR</b>	<b>NAR</b>
<b>3 M</b>	<b>MAPE</b>	1.25	0.77	1.27	1.28	1.92	2.77
	<b>RMSE</b>	115.07	63.24	116.52	117.17	146.55	235.18
	<b>MAE</b>	95.3	59.1	98.03	98.43	146.49	209.49
<b>6 M</b>	<b>MAPE</b>	2.12	1.84	2.43	2.4	2.91	2.26
	<b>RMSE</b>	184.03	148.53	205.73	205.6	258.73	188.3
	<b>MAE</b>	151.85	127.89	170.65	169.16	210.56	164.45
<b>12 M</b>	<b>MAPE</b>	2	2.11	2.45	2.46	3.6	3.28
	<b>RMSE</b>	220.39	198.46	220.92	222.33	328.22	322.32
	<b>MAE</b>	150.06	157.76	181.47	182.66	274.93	254.92
<b>24 M</b>	<b>MAPE</b>	2.35	2.49	3.08	3.16	4.44	5.4
	<b>RMSE</b>	270.95	250.22	353.4	358.83	557.11	626.43
	<b>MAE</b>	189.65	199.86	254.46	260.71	377.05	456.13

Tableau 5.1 – Mesures de performance des modèles sur 3,6,12 et 24 mois.

NAR-FFNN proposé est capable d'apprendre à partir d'un petit échantillon de données et peut prévoir jusqu'à 24 mois. Ce modèle est exempt de variables exogènes. En outre, le processus de décomposition intégré dans le modèle, rend le modèle plus robuste aux valeurs incorrectes dans la série chronologique.

Un paramètre important pour l'industrie de l'énergie est de prédire la pointe de la charge annuelle pour répondre à la demande future. Les prévisions à moyen long terme permettent aux fournisseurs d'énergie de décider si les installations actuelles peuvent répondre à la demande future ; sinon, il devient nécessaire de planifier la construction de nouvelles centrales électriques.

Dans ce cas, le X12-NAR-FFNN atteint le pic d'Août 2011 et 2012, et donne une bonne représentation de la saisonnalité et la tendance, qui sont bien reproduits par le modèle proposé (Figure. 5.6). X12-NAR-FFNN a aussi le meilleur MAPE, RMSE, MAE ce qui en fait le meilleur modèle en terme de précision de prédiction entre les modèles qui ont déjà prouvé leurs efficacités, comme le SD (M)-LSSVR-LSSVR modèle.

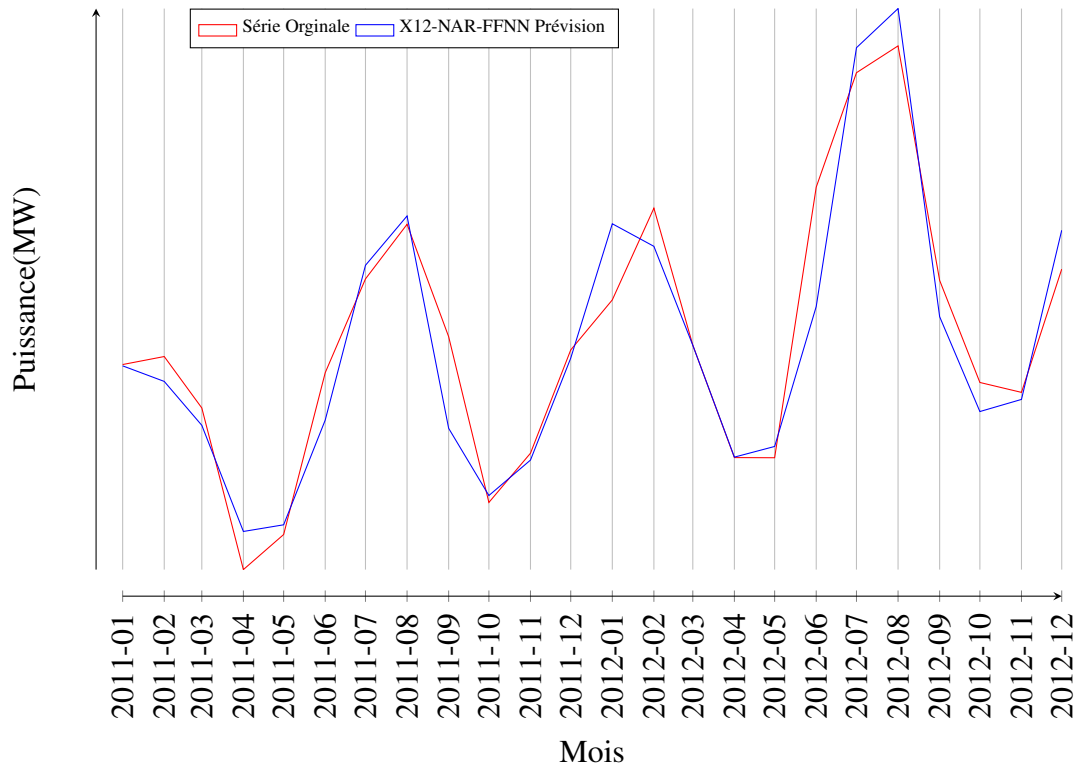


Figure 5.6 – La prévision de la charge électrique algérienne de Janvier 2010 à Décembre 2012 du modèle X12-NAR-FFNN.

## 5.5 Conclusion

La charge électrique dépend fortement des variables exogènes et de la modélisation du modèle. Souvent, ça implique l'incorporation des variables dites explicatives (exogènes) dans le modèle, ce qui est fastidieux et peut s'avérer complexe en la prévision de chaque variable exogène. C'est pour cela que les modèles uni-variés suscitent un intérêt grandissant.

Chaque méthode computationnelle a ses avantages et inconvénients. Notamment, les réseaux de neurones artificiels, qui ont été utilisés dans la littérature pour la prévision de la charge électrique. Cependant, ils sont connus pour être gourmands en données d'apprentissage, ce qui est problématique dans certains cas où l'échantillon de données fournies est petit. Des modèles dit hybrides ou multi-modèles ont vu le jour ces derniers temps, leur principe est de combiner des modèles issus de l'intelligence artificielle pour avoir de meilleurs résultats en comparaison avec les modèles simples.

Dans ce chapitre, un multi-modèle uni-varié basé sur le paradigme des réseaux de neurones artificiels est proposé pour la prévision de la charge électrique algérienne pour le moyen et long terme. Le modèle a été conçu dans l'objectif d'avoir de bonnes per-

formances avec un petit échantillon de données d'apprentissage. Le X12-NAR-FFNN est une combinaison de deux types de réseaux de neurones artificiels : Réseau de neurones autorégressif (NAR) et réseau Feed forward, qui prennent comme entrée la charge électrique préalablement décomposée en trois composantes par la méthode X12-ARIMA. Le X12-NAR-FFNN montre de meilleures performances que les modèles proposés dans la littérature jusqu'à 24 mois d'horizon (de janvier 2011 à Décembre 2012), cela peut être déduit des résultats obtenus. La décomposition saisonnière augmente considérablement les performances du modèle, et la combinaison des NARs et FFNN donne des résultats concluants dans le moyen et long terme en comparaison aux modèles qui ont prouvé leur efficacité dans la littérature.

## **Chapitre 6**

## **Conclusion**

La prévision à moyen long terme de la charge électrique a un rôle vital dans la production, la transmission et la planification du réseau. Pour cette raison, de nombreuses recherches ont été menées dans le but de trouver des solutions. Les premières méthodes utilisées sont des méthodes statistiques telles que les modèles de régression, qui ne pouvaient pas fournir une représentation fidèle du comportement de la charge électrique dans un marché de l'électricité volatil.

Au cours des dernières années, les méthodes issues de l'intelligence artificielle ont été beaucoup appliquées pour la prévision de la charge ; les réseaux neurones artificiels, les systèmes neuro-flous, SVM, etc. qui contrairement aux méthodes statistiques, fournissent des solutions prometteuses malgré les divers facteurs qui doivent être pris en compte lors de la prévision de charge, ce qui rend cette dernière complexe. La meilleure méthode de prévision serait donc celle qui pourrait trouver des relations non linéaires entre la charge et divers facteurs économiques et autres, et est adaptable aux changements.

Nous avons présenté dans cette thèse une étude de la charge électrique algérienne, où nous avons proposé différentes approches prévisionnelles à moyen et à long terme. Une grande partie de notre étude est basée sur la compréhension et l'analyse des données fournies par notre partenaire SONELGAZ. Cela étant fait, nous nous sommes intéressés à la première partie qui consiste en la proposition d'une prévision à long terme allant jusqu'en 2030 pour les différents niveaux d'énergies et les secteurs d'activités. La prévision s'est faite sur 3 scénarios (Faible, moyen et fort) pour chaque modèle.

La difficulté fût de proposer un modèle assez robuste pour gérer la variance des différents scénarios, où généralement le scénario fort a plus une forme exponentielle que linéaire. Suite à notre étude, plus de 90 modèles ont été proposés et quarante-quatre ont été sélectionnés pour la prévision les différents niveaux de tension et secteurs d'activités. En plus, un mode mixte qui combine entre, une régression exponentielle pour le moyen terme suivie d'une régression linéaire multiple qui prends le relais pour le long terme. Les résultats proposés ont été validés avec notre partenaire SONELGAZ. Dans un second temps, nous avons présenté le système de profils de charge électrique en calculant le facteur de charge qui est multiplié par la prévision de l'énergie annuelle. Une ventilation de la charge annuelle en 8760 ce qui correspond au nombre d'heures dans l'année, l'intérêt des profils c'est qu'ils sont modifiables et interchangeables, ce qui rend la prévision à court terme (heure, jour, semaine) dynamique.

Un outil nommé PREVELEC-DZ a été conçu pour SONELGAZ dans lequel nous avons modélisé les modèles régressifs pour les différentes prédictions des niveaux d'énergie, ainsi que les sociétés de distribution, secteurs d'activités et scénarios de décompositions. Le logiciel conduit non seulement à des résultats appréciables validés, mais offre

également la convivialité, l'aspect ergonomique et facilité d'utilisation, une visualisation des données historiques et des prévisions ainsi que la gestion des profils. Des extensions sont déjà prévues comme le cadre du logiciel, en raison de son aspect décomposable, est facilement extensible. Enfin, PREVELEC-DZ est la meilleure preuve qu'une bonne combinaison entre universitaires et industriels conduit inévitablement à la réussite. L'outil est prévu pour prendre en considération tous les aspects relatifs à la prévision et prise de décision pour des experts de SONELGAZ.

Nous avons choisi le modèle de régression en raison du manque de données. Cependant, il existe d'autres méthodes qui peuvent être appliquées dans ce cas particulier, notamment la théorie Grey, qui est adaptée à la prévision de charge [68] [26], parce que son principe est simple et le calcul est pratique dans un petit échantillon. Notre recherche future concerne l'application de cette théorie à notre jeu de données pour prédire l'électricité algérienne des prochaines années, et de pouvoir la comparer avec ce que nous avons réalisé.

La charge est fortement tributaire des variables exogènes et du modèle. Pour le moyen et long terme, cela implique l'intégration de ces variables, ce qui est souvent une tâche compliquée ; d'où l'intérêt porté aux modèles univariés, qui nécessitent que les valeurs de la charge. Les réseaux de neurones artificiels ont été largement utilisés pour les prévisions de charge ; cependant, ils ont l'inconvénient d'être avides de données d'apprentissage. Un autre aspect de notre thèse est la proposition d'un multi-modèle basé sur les réseaux de neurones pour la prévision de la charge électrique algérienne.

Un multi-modèle univarié basé sur les réseaux neuronaux a été proposé pour la prévision de la charge électrique algérienne pour le moyen et le long terme, avec un petit échantillon de données de formation (Janvier 2000 à Décembre 2010). Le modèle X12-NAR-FFNN proposé surpasse ses concurrents pour une prévision de 24 mois (de Janvier 2011 à Décembre 2012), on peut déduire à partir des résultats obtenus, que la décomposition saisonnière augmente considérablement les performances de prévision, et la combinaison des réseaux neuronaux NAR et FFNN fournit des résultats concluants pour le moyen long terme, comparé à d'autres modèles avec une efficacité prouvée dans la littérature.

Une deuxième partie de l'étude a été réalisée pour répondre au besoin de visualisation de la consommation à court terme de la charge (horaire) par l'expert. Par conséquent, l'approche de profil a été modifiée pour une décomposition par mois au lieu d'une décomposition par saison comme réalisé dans la partie régressive. La combinaison des deux approches permet une prévision mensuelle à moyen terme et une visualisation à court terme avec des résultats fiables, comme indiqué dans l'expérimentation.

## 6.1 Perspectives et travaux futurs

L'apport majeur de notre thèse, est de proposer des modèles prévisionnels qu'ils soient régressifs ou réseaux neuronaux pour des utilisateurs incapables de construire ou de paramétrer un modèle mathématique, ou de réaliser un apprentissage sur des données quelconques, dans l'intérêt de faire une prévision notamment pour le secteur énergétique.

Pour arriver à l'objectif voulu, c'est-à-dire un système de prévision autonome qui ne nécessite pas de paramétrage, ni de connaissances préalables sur l'apprentissage automatique ou statistique, ni de traitement de données préalable, un ensemble de propositions a été formulé dans cette recherche. Toutefois, les travaux initiés dans cette thèse reste en cours pour aller vers les perspectives suivantes :

- Proposer des modèles ARIMA pour les tensions qui sont auto-corrélées notamment la BT.
- Proposer la décomposition X12-ARIMA dans PREVELEC-DZ.
- Implémenter le modèle X12-NAR-FFNN dans PREVELEC-DZ.
- Réaliser une version PREVELEC-DZ basée serveur.
- Perfectionner la décomposition saisonnière en introduisant les jours fériés et les fêtes religieuses spécifiques au pays, dans l'algorithme X12-ARIMA.
- Prévoir des profils de charge, en se basant sur la prévision de la composante saisonnière du modèle X12-NAR-FFNN.
- Proposer un multi-modèle neuro-flou pour la prévision de la charge électrique algérienne à moyen long terme.
- Étudier le comportement du X12-NAR-FFNN sur différents types de séries temporelles et essayer de démontrer que c'est un modèle universel capable de prévoir n'importe quelle série dans le temps.

## **Annexe A**

# **Calculs de la régression linéaire et exponentielle**



## A.1 Les modèles régressifs

Une situation courante en sciences est d'avoir à sa disposition deux ensembles de données de taille  $n$ ,  $y_1, y_2, \dots, y_n$  et  $x_1, x_2, \dots, x_n$ , obtenus expérimentalement ou mesurés sur une population. Le problème de la régression consiste à rechercher une relation pouvant éventuellement exister entre les  $x$  et les  $y$ , par exemple de la forme  $y = f(x)$ .

Lorsque la relation recherchée est affine, c'est-à-dire de la forme  $y = ax + b$ , on parle de régression linéaire. Mais même si une telle relation est effectivement présente, les données mesurées ne vérifient pas, exactement, cette relation en général.

Pour tenir compte, dans le modèle mathématique, des erreurs observées, on considère les données  $y_1, y_2, \dots, y_n$  comme autant de réalisations d'une variable aléatoire  $Y$  et parfois aussi les données  $x_1, x_2, \dots, x_n$  comme autant de réalisations d'une variable aléatoire  $X$ . On dit que la variable  $Y$  est la variable dépendante ou variable expliquée et que la variable  $X$  est la variable explicative.

Dans l'analyse de régression, nous nous sommes intéressés à évaluer le rôle d'une variable prédictive ( $x$ ) qui est supposée être fixée, pour expliquer la variabilité d'une variable de résultat ( $y$ ). Nous évaluons cette relation par échantillonnage d'une certaine population pour estimer le modèle :

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (\text{A.1})$$

Où :  $\beta_0$  est le résultat lorsque l'indicateur  $y$  est 0.  $\beta_1$  représente la pente, ou la quantité de changements dans le résultat  $y$  par unité de variation du prédicteur  $x$ .  $\varepsilon$  est le terme d'erreur.

### A.1.1 La régression linéaire simple

En statistique, la régression linéaire simple est l'estimateur des moindres carrés d'un modèle de régression linéaire avec une seule variable explicative. Dans la régression linéaire, nous avons habituellement beaucoup de différentes valeurs de la variable explicative, et on suppose généralement que les valeurs entre les valeurs observées des variables explicatives sont aussi des valeurs possibles des variables explicatives. Nous postulons une relation linéaire entre la moyenne de la population du résultat et la valeur de la variable explicative.

Réduire la moyenne ou le total des erreurs carrées entre les observations et les sorties

du modèle est la principale préoccupation derrière les paramètres de réglage des modèles de la régression linéaire simple.

Dans le modèle de la régression linéaire simple[69], il y a juste une variable, appelée : variable dépendante et l'autre connue comme variable indépendante, où la variable dépendante doit être estimée en fonction de l'autre. L'équation A.2 montre la relation entre les variables dépendantes et indépendantes où  $\beta_0$  et  $\beta_1$  sont appelés coefficients du modèle de régression simple entre les variables dépendantes  $x$  et la variable indépendante  $y$ .

$$y = \beta_0 + \beta_1 x \quad (\text{A.2})$$

Ainsi, le modèle structurel dit que pour chaque valeur de  $x$ , la moyenne de population  $y$  peut être calculée en utilisant l'expression linéaire simple  $\beta_0 + \beta_1 x$ . Bien sûr nous ne pouvons pas faire le calcul exactement en pratique, parce que les deux paramètres sont inconnus. En pratique, nous procédons aux estimations des paramètres et nous remplaçons les estimations de l'équation.

Dans la vraie vie, nous savons que, bien que l'équation fasse une prédiction de la moyenne vraie du résultat pour chaque valeur fixe de la variable explicative, il ne serait pas judicieux d'utiliser une extrapolation pour faire la prédiction hors de la plage des valeurs de  $x$  dont nous disposons pour l'étude. D'autre part, il est raisonnable d'interpoler, c'est à dire, de faire des prédictions pour les valeurs de  $x$  non observées entre les valeurs  $x$  observées. Le modèle de structure est essentiellement issu de l'hypothèse de linéarité, au moins dans la plage des données explicatives observées. [70]

Il est important de réaliser que le « linéaire » dans la régression linéaire n'implique pas que seuls des relations linéaires peuvent être étudiées. Techniquement, il dit seulement que le bêta ne doit pas être sous une forme transformée. Il est possible de transformer  $x$  ou  $y$ , ce qui permet à de nombreuses relations non linéaires d'être représentées à une nouvelle échelle qui rend la relation linéaire.

Le modèle d'erreur utilisé est que pour chaque  $x$ , nous avons ou nous pouvons collecter plusieurs sujets avec cette valeur, leur distribution autour de la moyenne de la population est gaussienne avec une diffusion, dite  $\varepsilon$ , qu'il y est la même valeur pour chaque valeur de  $x$  (et moyenne de la population correspondante de  $y$ ). Bien sûr, la valeur de  $y$  est un paramètre inconnu, et nous pouvons faire une estimation de celle-ci à partir des données.

### A.1.1.1 Hypothèses

Ces hypothèses pèsent sur les propriétés des estimateurs (biais, convergence) et l'inférence statistique (distribution des coefficients estimés).

- **Hypothèse 1 :** Hypothèses sur  $Y$  et  $X$ .  $X$  et  $Y$  sont des grandeurs numériques mesurées sans erreur.  $X$  est une donnée exogène dans le modèle. Elle est supposée non aléatoire.  $Y$  est aléatoire par l'intermédiaire de  $\varepsilon$  c'est-à-dire la seule erreur que l'on a sur  $Y$  provient des insuffisances de  $X$  à expliquer ses valeurs dans le modèle.
- **Hypothèse 2 :** Hypothèses sur le terme aléatoire  $\varepsilon$ . Les  $\varepsilon_i$  sont indépendants et identiquement distribués.
  - $E(\varepsilon_i)$  en moyenne les erreurs s'annulent c'est-à-dire le modèle est bien spécifié.
  - $V(\varepsilon_i) = \sigma^2$  la variance de l'erreur est constante et ne dépend pas de l'observation. C'est l'hypothèse d'homoscédasticité.
  - En particulier, l'erreur est indépendante de la variable exogène c'est-à-dire  $cov(x_i, \varepsilon_i) = 0$
  - Indépendance des erreurs. Les erreurs relatives à 2 observations sont indépendantes c'est-à-dire  $cov(x_i, \varepsilon_i) = 0$ . On parle de « non autocorrélation des erreurs ».

**Remarque :** Cette hypothèse est toujours respectée pour les coupes transversales. En effet, l'échantillon est censé être construit de manière aléatoire et les observations indépendantes et identiquement distribuées. Nous pouvons donc intervertir aléatoirement les lignes sans porter atteinte à l'intégrité des données. En revanche, la question se pose pour les données temporelles. Il y a une contrainte qui s'impose à nous (contrainte temporelle - les données sont ordonnées) dans le recueil des données.

- $\varepsilon_i = N(0, \sigma_\varepsilon)$  L'hypothèse de normalité des erreurs est un élément clé pour l'inférence statistique. [71]

### A.1.1.2 Estimation des moindres carrés ordinaires (MCO)

Notre objectif est de déterminer les valeurs de  $a$  et  $b$  en utilisant les informations apportées par l'échantillon. Nous voulons que l'estimation soit la meilleure possible, c'est-à-dire ; la droite de régression doit approcher au mieux le nuage de points. [71]

Si graphiquement la solution semble intuitive, il nous faudrait un critère numérique

qui réponde à cette spécification pour réaliser les calculs sur un échantillon de données. Le critère des moindres carrés consiste à minimiser la somme des carrés des écarts (des erreurs) entre les vraies valeurs de Y et les valeurs prédites avec le modèle de prédiction (Figure A.1). L'estimateur des moindres carrés ordinaires (MCO) des paramètres a et b doit donc répondre à la minimisation de  $\varepsilon_i$ .

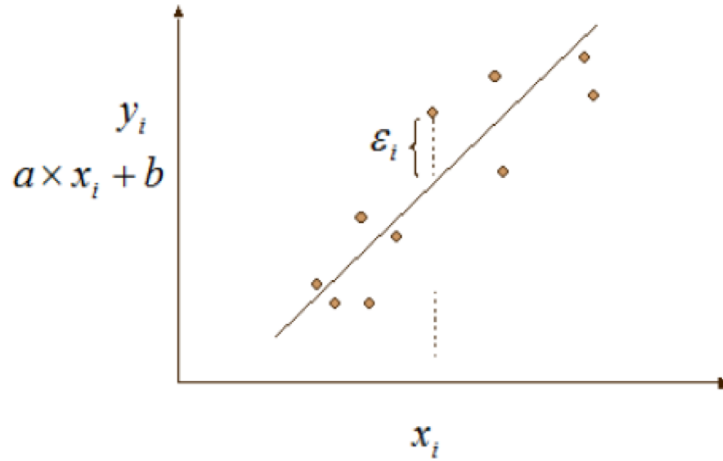


Figure A.1 – Comptabilisation de l'erreur : écart entre Y observé et Y prédit par le modèle linéaire.

$$S = \sum_{i=1}^n \varepsilon_i^2 \quad (\text{A.3})$$

$$= \sum_{i=1}^n [y_i - (ax_i + b)]^2 \quad (\text{A.4})$$

$$= \sum_{i=1}^n [y_i - ax_i - b]^2 \quad (\text{A.5})$$

Pour déterminer les valeurs de a et b, les conditions suivantes sont nécessaires :

$$\begin{cases} \frac{\partial S}{\partial a} = 0 \\ \frac{\partial S}{\partial b} = 0 \end{cases} \quad (\text{A.6})$$

En appliquant ces dérivées partielles, nous obtenons les équations normales :

$$\begin{cases} \sum_i x_i y_i - a \sum_i x_i^2 - b \sum_i x_i = 0 \\ \bar{y} - a\bar{x} - b = 0 \end{cases} \quad (\text{A.7})$$

que l'on trouve également sous la forme suivante dans la littérature :

$$\begin{cases} \sum_i x_i \varepsilon_i = 0 \\ \sum_i \varepsilon_i = 0 \end{cases} \quad (\text{A.8})$$

En appelant  $\hat{a}$  et  $\hat{b}$  les solutions de ces équations normales, nous obtenons les estimateurs des moindres carrés :

$$\hat{a} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (\text{A.9})$$

$$\hat{b} = \bar{y} - \hat{a}\bar{x} \quad (\text{A.10})$$

### A.1.1.3 Le point moyen

Nous pouvons définir la droite de régression linéaire ainsi :

- Le point moyen est le point qui a pour coordonnées la moyenne de X et la moyenne de Y, on l'appelle également le centre de gravité.
- La droite de régression qui passe par le point moyen. C'est aussi la droite qui minimise la somme des carrés des écarts des observations.

Une fois connue, l'équation de cette droite permet de résumer la série et de faire des prévisions. [72]

### A.1.1.4 Calcul des coefficients

L'équation de la droite de régression se calcule ainsi : Soit la droite d'équation :

$$y = ax + b \quad (\text{A.11})$$

Si nous voulons que cette droite soit ajustée à un nuage de points dans le plan  $X, Y$ , il est nécessaire de calculer les coefficients  $a$  et  $b$  en appliquant les formules suivantes :

$$a = \frac{\text{cov}(x, y)}{\sigma_x^2} \quad (\text{A.12})$$

$$b = \bar{\bar{y}} - a\bar{\bar{x}} \quad (\text{A.13})$$

Où  $cov(x, y)$  représente la covariance de  $(x, y)$  et se calcule ainsi :

$$cov(x, y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{\bar{x}} \bar{\bar{y}} \quad (\text{A.14})$$

Par conséquent la formule détaillée de  $a$  est de :

$$a = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{\bar{x}} \bar{\bar{y}}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{\bar{x}}^2} \quad (\text{A.15})$$

### A.1.2 La régression linéaire multiple

La régression linéaire multiple représente une généralisation à plus d'une seule variable explicative du modèle de régression linéaire simple. Le but de ce type de régression est de modéliser la relation entre une variable expliquée aléatoire et un certain nombre de variables explicatives. Strictement parlant, les valeurs de la variable explicative sont supposées être connues, ou sous le contrôle de l'enquêteur, en d'autres termes, elles ne sont pas considérées comme des variables aléatoires. [73]

Le modèle de régression linéaire multiple concerne une variable expliquée à un ensemble de variables explicatives. La relation supposée est linéaire (en termes de paramètres plutôt qu'en termes de variables explicatives), et les paramètres du modèle (généralement connus sous le nom de coefficients de régression) sont généralement estimés par les moindres carrés. Un cadre d'inférence est ajouté en faisant certaines hypothèses de répartition sur les termes d'erreur dans le modèle.

Supposons que l'objet est affecté par des facteurs multiples tels que :  $x_1, x_2, \dots, x_k$  qui sont linéaires avec  $y$  et puis il doit établir un modèle de régression linéaire multiple. C'est-à-dire :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (\text{A.16})$$

$y$  est la variable expliquée,  $x_j (j = 1, 2, \dots, k)$  sont les variables explicatives et  $\beta_j (j = 1, 2, \dots, k)$  sont des paramètres inconnus.

Pour un groupe d'observation  $y_i$ ,  $x_{1i}, x_{2i}, \dots, x_{ki}$  ( $i = 1, 2, \dots, n$ ) les systèmes d'équations sont :

$$y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \quad (i = 1, 2, \dots, n) \quad (\text{A.17})$$

Après avoir cherché des variables et des observations, le calcul de régression linéaire peut induire le modèle de prévision de la régression linéaire multiple.

### A.1.2.1 Les hypothèses

Comme pour la régression simple, les hypothèses permettront de déterminer les propriétés des estimateurs (biais, convergence) et les lois de distribution (loi de Student pour chaque coefficient pris individuellement, loi de Fisher dès que l'on traite un groupe de coefficients). Nous distinguons :

- Les hypothèses stochastiques
  - Les  $X_j$  sont non aléatoires c'est-à-dire les  $x_{ij}$  sont observés sans erreur.
  - $E[\varepsilon_i] = 0$ , l'espérance de l'erreur est nulle. En moyenne, le modèle est bien spécifié.
  - $E[\varepsilon_i^2] = \sigma^2$ , la variance de l'erreur est constante, c'est l'hypothèse de homoscedasticité.
  - $cov(\varepsilon_i, \varepsilon_i)$  pour  $i \neq 1$ , les erreurs sont indépendantes, c'est l'hypothèse de non autocorrélation des résidus.
  - $cov(x_{i,j}, \varepsilon_i) = 0$ , l'erreur est indépendante des variables exogènes.
  - $\varepsilon_i = N(0, \sigma_\varepsilon)$ , les erreurs sont distribuées selon une loi normale.
- Les hypothèses structurelles
  - La matrice  $(\dot{X}X)$  est régulière c'est-à-dire  $(\dot{X}X) \neq 0$  et  $(\dot{X}X)^{-1}$  existe. Elle indique l'absence de colinéarité entre les variables exogènes. Nous pouvons aussi voir cette hypothèse sous l'angle  $\text{rang}(X) = p + 1$  et  $\text{rang}(\dot{X}X) = p + 1$
  - $-\frac{\text{num}(\dot{X}X)}{n}$  tend vers une matrice non singulière lorsque  $n \rightarrow +\infty$ .
  - $n > p + 1$ , le nombre d'observations est supérieur au nombre de paramètres à estimer. Dans le cas où  $n = p + 1$ , nous avons une interpolation, la droite passe exactement par tous les points. Lorsque  $n < p + 1$ , la matrice  $(\dot{X}X)$  n'est plus inversible. [71]

### A.1.2.2 Minimisation de la somme des carrés des erreurs

Comme pour la régression simple, on cherche les coefficients qui permettent de minimiser la quantité suivante :

$$S = \sum_{i=1}^n \varepsilon_i^2 \quad (\text{A.18})$$

$$\text{Où : } \varepsilon_i^2 = [y_i - (a_0 + a_1x_{i,1} + \dots + a_px_{i,p})]^2$$

On passe de nouveau par les dérivées partielles que l'on annule pour obtenir les  $(p+1)$  équations normales.

$$\begin{cases} a_0 + a_1\bar{x}_1 + \dots + a_p\bar{x}_p = \bar{y} \\ a_0 \sum_i x_{i,p} + a_1 \sum_i x_{i,1}x_{i,p} + \dots = a_p \sum_i x_{i,p}x_{i,p} = x_{i,1}y_p \end{cases} \quad (\text{A.19})$$

Nous avons  $(p+1)$  équations à  $(p+1)$  inconnues. Nous pouvons en extraire les estimations  $(\hat{a}_0, \hat{a}_1, \dots, \hat{a}_p)$ , mais cette écriture est difficile à manipuler.[74]

Avec l'écriture matricielle, nous pouvons produire une écriture condensée. Soit  $\varepsilon$  le vecteur des erreurs, avec  $\varepsilon' = (\varepsilon_1, \dots, \varepsilon_n)$ . La somme des carrés des erreurs devient :

$$S = \sum_i \varepsilon_i^2 = \varepsilon \varepsilon' \quad (\text{A.20})$$

Développons l'expression :

$$\begin{aligned} \varepsilon \varepsilon' &= (Y - Xa)'(Y - Xa) \\ &= Y'Y - Y'Xa + a'X'Y + a'X'Xa \\ &= Y'Y - 2a'X'Y + a'X'Xa \\ S &= Y'Y - 2a'X'Y + a'X'Xa \end{aligned} \quad (\text{A.21})$$

Quelques éléments sur les calculs matriciels pour comprendre les développements ci-dessus :

$$(Xa)' = a'X'(Y'Xa)' = a'X'Y \quad (\text{A.22})$$



La transposée d'un scalaire est égale à lui-même. Or, en se référant aux dimensions des vecteurs et matrice, on constate que  $(a'X'Y)$  est de dimension  $(1, 1)$ . Pour déterminer le minimum de  $S$ , nous réalisons la dérivation matricielle que nous annulons :

$$\frac{\partial S}{\partial a_0} = -2(X'Y) + 2(X'X)a = 0(X'X)a = X'Y \quad (\text{A.23})$$

L'estimateur des moindres carrés ordinaires (MCO) pour obtenir les coefficients du modèle s'écrit :

$$\hat{a} = (X'X)^{-1}X'Y \quad (\text{A.24})$$

## A.2 La régression exponentielle

Parfois nous savons que la liaison n'est pas linéaire, soit parce que nous avons des connaissances expertes sur le problème que nous traitons, soit parce que nous le constatons visuellement en construisant le nuage de points. Nous sommes alors confrontés à un double problème : déterminer la forme de la liaison, la fonction reliant  $Y$  à  $X$  ; en estimant les paramètres éventuels à partir des données disponibles.

Il existe cependant une classe de fonctions que nous pouvons linéariser en appliquant les transformations adéquates. Dans ce cas, l'estimation des paramètres devient possible. L'interprétation des résultats est modifiée cependant, notamment en ce qui concerne la pente. Dans le modèle exponentiel, la relation s'écrit :

$$y = e^{(ax+b)} \quad (\text{A.25})$$

Le coefficient de la pente se lit :

$$a = \frac{\frac{\partial y}{\partial x}}{y} \quad (\text{A.26})$$

Le taux de variation de  $Y$  est proportionnel à la variation de  $X$ . Ce type de modèles est surtout utilisé quand  $X$  correspond au temps, ainsi  $\partial x = 1$ . Dans ce cas, la croissance (ou décroissance) de  $Y$  est constante dans le temps. Ce type d'évolution (exponentielle) ne dure pas longtemps. On linéarise la relation de la manière suivante :

$$\ln y = \ln b + axd \quad (\text{A.27})$$

### A.3 Exemple

Calculons  $a$  et  $b$  dans le cas de la série  $S$  :

$$S = \{\{1; 3.5\}; \{3; 3.6\}; \{4; 4\}; \{6; 5\}; \{7; 6.6\}; \{8; 6.8\}\}$$

Afin de faciliter les calculs, adoptons la disposition du tableau A.1 : Calculons  $a$  :

<b>x</b>	<b>y</b>	<b>xy</b>	<b>x<sup>2</sup></b>	<b>y<sup>2</sup></b>
1	3.5	3.5	1	12.25
3	3.6	10.8	9	12.96
4	4	16	16	16
6	5	30	36	25
7	6.6	46.2	49	43.56
8	6.8	54.4	64	46.24
<b>29</b>	<b>29.5</b>	<b>160.9</b>	<b>175</b>	<b>156</b>

Tableau A.1 – Les valeurs nécessaires au calcul de la droite.

$$a = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{y} \bar{x}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}$$

$$a = \frac{\frac{160.9}{6} - \frac{29}{6} * \frac{29.5}{6}}{\frac{175}{6} - \left(\frac{29}{6}\right)^2}$$

Une fois  $a$  connu, nous en déduisons  $b$  :

$$b = \bar{y} - a\bar{x}$$

$$b = \left(\frac{29}{6}\right) - 0.5258 * \left(\frac{29}{6}\right) = 2.37512$$

L'équation de la droite de régression est donc :

$$y = 0.5258x + 2.37512$$

# Bibliographie

- [1] PK Nag. *Power plant engineering*. Tata McGraw-Hill Education, 3 edition, 2008.
- [2] Philip Graham Hill. *Power generation : resources, hazards, technology, and costs*. MIT Press, Cambridge, MA, 1977.
- [3] Jean-Christophe Victor. Chères terres rares - le dessous des cartes, Mars 2015.
- [4] Siegbert Hopf. Planning of electric power distribution. Technical report, Siemens AG, Mozartstr. 31c D-91052 Erlangen, Germany, 2016.
- [5] H Lee Willis and JED Northcote-Green. Comparison tests of fourteen distribution load forecasting methods. *IEEE Transactions on Power Apparatus and Systems*, PAS-103(6) :1190–1197, June 1984.
- [6] Hossein Daneshi, Mohammad Shahidehpour, and Azim Lotfjou Choobbari. Long-term load forecasting in electricity market. *2008 IEEE International Conference on Electro/Information Technology, IEEE EIT 2008 Conference*, pages 395–400, June 2008.
- [7] H Lee Willis, Linda A Finley, and MJ Buri. Forecasting electric demand of distribution system planning in rural and sparsely populated regions. *IEEE transactions on power systems*, 10(4) :2008–2013, November 1995.
- [8] Patrice H Henault, Robert B Eastvedt, John Peschon, and Laszlop P Hajdu. Power system long-term planning in the presence of uncertainty. *IEEE Transactions on power apparatus and systems*, PAS-89(1) :156–164, January 1970.
- [9] Soliman Abdel-hady Soliman and Ahmad Mohammad Al-Kandari. *Electrical load forecasting : modeling and model construction*. Butterworth-Heinemann, April 2010.
- [10] Daniel Eilstein, Alain Le Tertre, Abdelkrim Zeghnoun, Sylvie Cassadou, Laurent Filleul, Laurence Pascal, Hélène Prouvost, Christophe Declercq, Philippe Saviuc, Agnès Lefranc, Catherine Nunes, Benot Chardon, Jean-Francois Jusot, Myriam D’Helf, Pascal Fabre, Sylvia Médina, and Philippe Quénel. Séries temporelles et modèles de régression. Technical report, Institut de veille sanitaire Département santé environnement, 2003.

- [11] Wayne A Fuller. *Introduction to statistical time series*, volume 428. John Wiley & Sons, 2009.
- [12] Benabbas Farouk. *méthodes heuristiques pour la prédiction des séries temporelles*. PhD thesis, Universté Badji Mokhtar Annaba UBMA, 2012.
- [13] Reda Mohamed Nezzar, Nadir Farah, Mohamed Tarek Khadir, and Lakhdar Chouireb. Mid-long term load forecasting using multi-model artificial neural networks. *International Journal on Electrical Engineering and Informatics*, 8(2) :389, 2016.
- [14] Estela Bee Dagum. Modelling, forecasting and seasonally adjusting economic time series with the x-11 arima method. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 27(3/4) :203–216, 1978.
- [15] Dominique Ladiray and Benoît Quenneville. Comprendre la méthode x11. *Institut National de la Statistique et des Études Économiques*, 1999.
- [16] David F Findley, Brian C Monsell, William R Bell, Mark C Otto, and Bor-Chung Chen. New capabilities and methods of the x-12-arima seasonal-adjustment program. *Journal of Business & Economic Statistics*, 16(2) :127–152, 1998.
- [17] Bernard Widrow and Marcian E Hoff. Adaptive switching circuits. In *Neurocomputing : foundations of research*, pages 123–134. MIT Press, 1988.
- [18] Hans-Dieter Block. The perceptron : A model for brain functioning. i. *Reviews of Modern Physics*, 34(1) :123, 1962.
- [19] GE Hinton, DE Rumelhart, and RJ Williams. Learning internal representations by back-propagating errors. *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, 1, 1985.
- [20] Badar Ul Islam. Comparison of conventional and modern load forecasting techniques based on artificial intelligence and expert systems. *IJCSI International Journal of Computer Science Issues*, 8(5) :504–513, 2011.
- [21] Shi Zhang and Dingwei Wang. Medium and long-term load forecasting based on pca and bp neural network method. In *Energy and Environment Technology, 2009. ICEET'09. International Conference on*, volume 3, pages 389–391. IEEE, 2009.
- [22] Qian Zhang and Tongna Liu. Research on mid-long term load forecasting base on wavelet neural network. In *Computer Engineering and Applications (ICCEA), 2010 Second International Conference on*, volume 2, pages 217–220. IEEE, 2010.
- [23] Qian Zhang and Tongna Liu. A fuzzy rules and wavelet neural network method for mid-long-term electric load forecasting. In *Computer and Network Technology (ICCNT), 2010 Second International Conference on*, pages 442–446. IEEE, 2010.
- [24] Lu Yue, Yao Zhang, Huifan Xie, and Qing Zhong. The fuzzy logic clustering neural network approach for middle and long term load forecasting. In *Grey Systems*

- and Intelligent Services, 2007. GSIS 2007. IEEE International Conference on*, pages 963–967. IEEE, 2007.
- [25] Haiyong Yu and Qian Zhang. Application of variable structure artificial neural network for mid-long term load forecasting. In *Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on*, pages 450–453. IEEE, 2010.
- [26] Wenqing Zhao and Dongxiao Niu. A mid-long term load forecasting model based on improved grey theory. In *Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on*, pages 633–635. IEEE, 2010.
- [27] Fuwei Zhang and Wei Chen. Application of grey ls-svm in mid and long term power load forecasting. In *Management and Service Science, 2009. MASS’09. International Conference on*, pages 1–4. IEEE, 2009.
- [28] Wei Sun and Wei Zhao. Mid-long term power load forecasting based on mg-caco and svm method. In *Future Computer and Communication (ICFCC), 2010 2nd International Conference on*, volume 1, pages V1–118. IEEE, 2010.
- [29] Dongxiao Niu, Haitao Lv, and Yunyun Zhang. Combination forecasting model for mid-long term load based on least squares support vector machines and a mended particle swarm optimization algorithm. In *Bioinformatics, Systems Biology and Intelligent Computing, 2009. IJCBS’09. International Joint Conference on*, pages 525–528. IEEE, 2009.
- [30] Jianzhou Wang, Xiaolong Ma, Jie Wu, and Yao Dong. Optimization models based on gm (1, 1) and seasonal fluctuation for electricity demand forecasting. *International Journal of Electrical Power & Energy Systems*, 43(1) :109–117, 2012.
- [31] Shuai Wang, Lean Yu, Ling Tang, and Shouyang Wang. A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in china. *Energy*, 36(11) :6542–6554, 2011.
- [32] Qia Ding. Long-term load forecast using decision tree method. In *Power Systems Conference and Exposition, 2006. PSCE’06. 2006 IEEE PES*, pages 1541–1543. IEEE, 2006.
- [33] Mohammad Moradi Dalvand, Seyed Bahram Zahir Azami, and Hadi Tarimoradi. Long-term load forecasting of iranian power grid using fuzzy and artificial neural networks. In *Universities Power Engineering Conference, 2008. UPEC 2008. 43rd International*, pages 1–4. IEEE, 2008.
- [34] Zhi-gang Du, Lin Niu, and Jian-guo Zhao. Long-term electricity demand forecasting using relevance vector learning mechanism. In *International Symposium on Neural Networks*, pages 465–472. Springer, 2007.

- [35] Yingling Shi, Hongsong Yang, Yawei Ding, and Nansheng Pang. Research on long term load forecasting based on improved genetic neural network. In *Computational Intelligence and Industrial Application, 2008. PACIA'08. Pacific-Asia Workshop on*, volume 2, pages 80–84. IEEE, 2008.
- [36] Otavio AS Carpinteiro, Rafael C Leme, Antonio C Zambroni de Souza, Carlos AM Pinheiro, and Edmilson M Moreira. Long-term load forecasting via a hierarchical neural model with time integrators. *Electric Power Systems Research*, 77(3) :371–378, 2007.
- [37] Arash Ghanbari, S Farid Ghaderi, and M Ali Azadeh. Adaptive neuro-fuzzy inference system vs. regression based approaches for annual electricity load forecasting. In *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*, volume 5, pages 26–30. IEEE, 2010.
- [38] Y Aslan, S Yavasca, and C Yasar. Long term electric peak load forecasting of kutahya using different approaches. *International Journal on Technical and Physical Problems of Engineering*, 3(7) :87–91, 2011.
- [39] Toly Chen. A collaborative fuzzy-neural approach for long-term load forecasting in taiwan. *Computers & Industrial Engineering*, 63(3) :663–670, 2012.
- [40] Rong-Jong Wai, Yu-Chih Huang, and Yi-Chang Chen. Design of intelligent long-term load forecasting with fuzzy neural network and particle swarm optimization. In *Machine Learning and Cybernetics (ICMLC), 2012 International Conference on*, volume 4, pages 1644–1650. IEEE, 2012.
- [41] Donald F Specht. A general regression neural network. *IEEE transactions on neural networks*, 2(6) :568–576, 1991.
- [42] Hong-Ze Li, Sen Guo, Chun-Jie Li, and Jing-Qi Sun. A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. *Knowledge-Based Systems*, 37 :378–387, 2013.
- [43] M Nosrati Maraloo, AR Koushki, C Lucas, and A Kalhor. Long term electrical load forecasting via a neurofuzzy model. In *Computer Conference, 2009. CSICC 2009. 14th International CSI*, pages 35–40. IEEE, 2009.
- [44] Hongyan Liu, Liya Cai, and Xiaojuan Wu. Grey-rbf neural network prediction model for city electricity demand forecasting. In *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM'08. 4th International Conference on*, pages 1–5. IEEE, 2008.
- [45] A.K. Chakraborty, S.C. Bhattacharya, and B.G. Varma. *The Story of Electricity*. Children's Book Trust, 1985.
- [46] J. Bonallack. *Switched On : The Story of Electricity*. Orbit investigations. Learning Media, 2007.

- [47] H. Schobert. *Energy : The Basics*. The Basics. Taylor & Francis, 2013.
- [48] Dave Barnett and Kirk Bjornsgaard. *Electric power generation : a nontechnical guide*. PennWell Books, 2000.
- [49] Issam Bechinia. Identification des éléments d'une centrale électrique thermique approche multi-modèle. Master's thesis, Université Badji Mokhtar, Annaba, 2010.
- [50] Malcolm Atkinson and Ronald Morrison. Orthogonally persistent object systems. *The VLDB Journal—The International Journal on Very Large Data Bases*, 4(3) :319–402, 1995.
- [51] Jerome T Connor, R Douglas Martin, and Les E Atlas. Recurrent neural networks and robust time series prediction. *IEEE transactions on neural networks*, 5(2) :240–254, 1994.
- [52] Nazih Abu-Shikhah, Fawwaz Elkarmi, and Osama M Aloquili. Medium-term electric load forecasting using multivariable linear and non-linear regression. *Smart Grid and Renewable Energy*, 2(02) :126, 2011.
- [53] Dongxiao Niu, Yan Lu, and Yanan Wei. Research on mid-long term power load forecasting models of anhui province. *International Journal of Advancements in Computing Technology*, 5(5), 2013.
- [54] HM Al-Hamadi and SA Soliman. Long-term/mid-term electric load forecasting based on short-term correlation and annual growth. *Electric power systems research*, 74(3) :353–361, 2005.
- [55] Rafal Weron. *Modeling and forecasting electricity loads and prices : A statistical approach*, volume 403. John Wiley & Sons, 2007.
- [56] MDKZ Ghiassi, David K Zimbra, and H Saidane. Medium term system load forecasting with a dynamic artificial neural network model. *Electric Power Systems Research*, 76(5) :302–316, 2006.
- [57] Changhao Xia, Jian Wang, and Karen McMenemy. Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. *International Journal of Electrical Power & Energy Systems*, 32(7) :743–750, 2010.
- [58] Qian Zhang Qian Zhang and Tongna Liu Tongna Liu. Research on the mid-long term electric load forecasting based on fuzzy rules. *Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on*, pages 461–463, 2010.
- [59] Biao Wang, Gang Feng, Jing Huang, and Xian-Yong Xiao. Electric power load forecasting based on fuzzy optimal theory. In *Intelligent System Design and Engineering Application (ISDEA), 2012 Second International Conference on*, pages 821–824. IEEE, 2012.

- [60] Guo Pan and Aijia Ouyang. Medium and long term power load forecasting using cpso-gm model. *Journal of Networks*, 9(8) :2121–2128, 2014.
- [61] Hong Zhang, Zhi Guo Lei, Yue Cheng, and Yu Ming Wang. Medium and long-term load forecasting using grey theory based on rough sets. In *Applied Mechanics and Materials*, volume 672, pages 1405–1408. Trans Tech Publ, 2014.
- [62] Xiang Shuo He, Li Yang, and Xiao Na Yu. Research of combined optimum grey model to mid and long term electric load forecasting. In *Advanced Materials Research*, volume 986, pages 1379–1382. Trans Tech Publ, 2014.
- [63] Zhigang Ji, Peijun Zhang, and Zhiwei Zhao. Application of wavelet neural network and rough set theory to forecast mid-long-term electric power load. In *Education Technology and Computer Science, 2009. ETCS'09. First International Workshop on*, volume 1, pages 1104–1108. IEEE, 2009.
- [64] Qian Zhang and Tongna Liu. A fuzzy rules and wavelet neural network method for mid-long-term electric load forecasting. In *Computer and Network Technology (ICCNT), 2010 Second International Conference on*, pages 442–446. IEEE, 2010.
- [65] Wei-Chiang Hong. Electric load forecasting by support vector model. *Applied Mathematical Modelling*, 33(5) :2444–2454, 2009.
- [66] H Braun and M Riedmiller. Rprop : a fast adaptive learning algorithm. In *Proceedings of the International Symposium on Computer and Information Science VII*, 1992.
- [67] Jianming Hu, Jianzhou Wang, and Guowei Zeng. A hybrid forecasting approach applied to wind speed time series. *Renewable Energy*, 60 :185–194, 2013.
- [68] He-Rui Cui, Wei Li, and Bo Zhang. A long-term forecasting model based on grey feed-back modification. In *2006 International Conference on Machine Learning and Cybernetics*, pages 413–416. IEEE, 2006.
- [69] Abbas Mahmoudabadi. Comparison of weighted and simple linear regression and artificial neural network models in freeway accidents prediction. In *Computer and Network Technology (ICCNT), 2010 Second International Conference on*, pages 392–396. IEEE, 2010.
- [70] Howard J Seltman. Experimental design and analysis. *Pittsburgh : Carnegie Mellon University*, 428, 2012.
- [71] R Rakotomalala. *Econométrie, la régression linéaire simple et multiple*. Université Lumière Lyon 2, 2016.
- [72] Fabrice Mazerolle. *Mémentos LMD - Statistique descriptive*. Gualino Editeur, 2005.
- [73] Sabine Landau and Brian Everitt. *A handbook of statistical analyses using SPSS*, volume 1. Chapman & Hall/CRC Boca Raton, FL, 2004.



- [74] M. Diener F. Diener. Méthode des moindres carrés. Université de Nice - Sophia Antipolis, 2007.