



République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université de Badji Mokhtar Annaba
Faculté des Sciences de l'Ingéniorat
Département d'Informatique



Thèse

Pour l'obtention du diplôme de Doctorat LMD - 3ème cycle

Apprentissage approfondi des réseaux de neurones artificiels pour l'approximation de séries temporelles

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : STIC (Sciences et Technologies de l'Information et de la Communication)

Kheir Eddine FARFAR

Directeur de thèse : Pr. Tarek Mohamed KHADIR

Devant le jury composé de :

Président	Pr. SERIDI Hassina	Université Badji Mokhtar – Annaba
Examineur	Pr. MOUSSAOUI Abdelouahab	Université de Sétif
Examineur	Pr. TOUMI Salah	Université Badji Mokhtar – Annaba
Examineur	Pr. FARAH Nadir	Université Badji Mokhtar – Annaba

Année 2017/2018

Remerciements

En premier lieu, je remercie Dieu de m'avoir permis de mener ce travail de recherche à terme.

Ce mémoire n'aurait pas vu le jour sans la contribution de plusieurs personnes à qui j'aimerais adresser mes remerciements. Je tiens tout d'abord à remercier mon directeur de thèse, Monsieur Khadir Mohamed Tarek professeur à l'université de Badji Mokhtar Annaba. Tout au long de ces cinq années, j'ai pu apprécier sa vision de recherche, son intégrité, sa disponibilité, sa générosité à partager ses connaissances. Je voudrais lui témoigner ici toute ma gratitude pour m'avoir guidé durant ces années.

Je remercie chaleureusement Monsieur Farah Nadir professeur à l'université de Badji Mokhtar Annaba et directeur du laboratoire LabGED, pour son soutien, ses nombreux conseils et l'intérêt qu'il a porté à mon travail tout au long de ces années.

Je voudrais remercier les membres de jury pour l'honneur qu'ils m'ont fait en participant à l'évaluation de ce travail :

Madame SERIDI Hassina, professeur à l'Université d'Annaba.

Monsieur MOUSSAOUI Abdelouaheb, professeur à l'Université de Sétif.

Monsieur TOUMI Salah, professeur à l'Université d'Annaba.

Monsieur FARAH Nadir, professeur à l'Université d'Annaba.

Ma gratitude s'adresse aussi à tous les enseignants de département d'informatique de l'université de Badji Mokhtar Annaba pour leurs conseils, leur encouragement et leur précieux enseignement.

Je voudrais également adresser mes remerciements à Monsieur Frédéric Alexandre pour son accueil et ses conseils pendant mon stage de perfectionnement.

Je n'oublie pas de remercier tous mes collègues et mes amis sans exception.

Merci à toute ma famille et plus particulièrement à mes très chers parents pour l'éducation qu'ils m'ont prodigué, avec tous les moyens et au prix de toutes les sacrifices qu'ils ont consentis à mon égard, pour le sens du devoir qu'ils m'ont enseigné depuis mon enfance et pour leur soutien et leur encouragement le long de ma carrière.

التعلم العميق للشبكات العصبونية الاصطناعية لتقريب دوال السلاسل الزمنية

المخلص: يمثل التنبؤ بالأحمال الكهربائية دوراً مهماً في التخطيط الجيد لأنظمة الطاقة ، حيث يتأثر الحمل بعدة عوامل يجب تحليلها وتحديدها قبل النمذجة لضمان موازنة تحميل أفضل وأكثر فورية بين العرض والطلب. تقترح هذه الرسالة منهج من مرحلتين للتنبؤ بالأحمال الكهربائية على المدى القصير. في المرحلة الأولى، يتم تحديد مجموعة من الأقسام للبيانات اليومية للحمل الكهربائي باستخدام خوارزمية تصنيفية، هذا بالموازاة مع تقدير بيانات درجة الحرارة على مدار اليوم، طريقة التقدير المقترحة مفيدة بشكل خاص في حالة عدم توفر بيانات درجة الحرارة، بينما في المرحلة الثانية، يتم استخدام نوع من الشبكات العصبونية الاصطناعية، تحديداً المرمز الآلي المصفي المكس، لبناء نماذج انحدار قادرة على التنبؤ بالحمل الكهربائي لكل صنف على حدا. خلال التجارب يتم التعليم الاصطناعي و تقييم النماذج على بيانات الطاقة الكهربائية المسجلة كل ساعة التي تقدمها الشركة الوطنية للكهرباء و الغاز الجزائرية، المقارنة مع نماذج اخرى أثبتت صحة و فعالية الطريقة المقترحة.

الكلمات المفتاحية : الشبكات العصبونية الاصطناعية ، التعلم العميق ، المرمز الآلي ، السلاسل

الزمنية ، التنبؤ بالحمل الكهربائي قصير المدى

Apprentissage profondi des réseaux de neurones artificiels pour l'approximation de séries temporelles

Résumé : La prévision de la charge électrique joue un rôle important dans la planification régulière des systèmes électriques, dans laquelle la charge est influencée par plusieurs facteurs qui doivent être analysés et identifiés avant la modélisation afin d'assurer un meilleur équilibre de charge entre l'offre et la demande. Cette thèse propose une approche en deux étapes pour la prévision de la charge électrique à court terme. Dans la première étape, un ensemble de classes des profils de charge est identifié en utilisant l'algorithme de classification K-Means ainsi que l'estimation des profils de température. La méthode d'estimation proposée est particulièrement utile en cas de manque de données historiques de la température régulière. Alors que dans la deuxième étape, l'approche Stacked Denoising Autoencoders est utilisée pour construire des modèles de régression pour prédire la consommation de chaque type de jour indépendamment. Les modèles obtenus sont établis et évalués en utilisant les données de la consommation horaires sur l'électricité fournies par la Société nationale de l'électricité et du gaz d'Algérie (Sonelgaz). Plusieurs modèles sont étudiés pour mesurer l'exactitude et l'efficacité de l'approche proposée. L'approche proposée dans cette thèse a été concrétisée par le développement d'un outil nommé PREVOS-DZ qui permet de rendre la prévision de la charge électrique accessible aux utilisateurs avec la possibilité d'effectuer des prévisions à la demande et en temps réel.

Mots clés : Réseaux de neurones artificiels, Apprentissage profondi, Auto-encodeur, Séries temporelles, Prévision de la charge électrique à court terme.

Deep learning of artificial neural networks for time series approximation

Abstract : Electrical load forecasting plays an important role in the regular planning of power systems, in which load is influenced by several factors that must be analysed and identified prior to modelling in order to ensure better and instant load balancing between supply and demand. This thesis proposes a two-stage approach for short-term electricity load forecasting. In the first stage, a set of day classes of load profiles are identified using K-means clustering algorithm alongside daily temperature estimation profiles. The proposed estimation method is particularly useful in case of lack of historical regular temperature data. While in the second stage, the stacked denoising autoencoders approach is used to build regression models able to forecast each day type independently. The obtained models are trained and evaluated using hourly electricity power data offered by Algeria's National Electricity and Gas Company. Several models are investigated to substantiate the accuracy and effectiveness of the proposed approach. The approach proposed in this thesis has been concretized by the development of a tool called PREVOS-DZ which makes it possible to make the forecast of the electric charge accessible to the users with the possibility of making forecasts on demand and in real time.

Keywords : Artificial neural networks, Deep learning, Autoencoders, Time series, Short-term load forecasting.

Liste des publications

Publications internationales

- Kheir Eddine Farfar, Mohamed Tarek Khadir "A two-stage short-term load forecasting approach using temperature daily profiles estimation" – *Neural Computing and Applications, 2018, Springer* <https://doi.org/10.1007/s00521-017-3324-x>

Conférences internationales

- Kheir Eddine Farfar, Mohamed Tarek Khadir "Day Types Identification of Algerian Electricity Load Using an Image Based Two-Stage Approach" – *25th International Conference on Artificial Neural Networks, Barcelona, Spain, September 6-9, 2016, Proceedings, Part II*
- Kheir Eddine Farfar, Mohamed Tarek Khadir, Oussama Laib "Comparison of serial and parallel approaches using artificial neural networks for Algerian short-term load forecasting" – *3rd International Conference on Advances in Computing, Electronics and Electrical Technology - CEET, Kuala Lumpur, Malaysia, April 11-12, 2015, Proceedings, p. 19-23.*

Table des matières

Introduction	1
1 État de l'art	5
1.1 Introduction	5
1.2 Les séries temporelles	6
1.2.1 Types de séries temporelles	6
1.2.2 Objectifs de l'analyse de série temporelle	7
1.2.3 La prédiction des séries temporelles	8
1.3 Prévision de la charge électrique	9
1.4 Les approches traditionnelles	11
1.4.1 Modèles de séries temporelles stochastiques	11
1.4.2 Modèles de régression	14
1.4.3 Filtre de Kalman	16
1.5 Approches basées sur l'Intelligence Artificielle	17
1.5.1 Réseau de neurones artificiels	17
1.5.2 Systèmes expert	22
1.5.3 Systèmes d'inférence flous et les systèmes à base de réseaux neuroniques et de logique floue	25
1.5.4 Machine à vecteurs de support	26
1.6 Conclusion	30
2 Réseaux de neurones artificiels profonds	33
2.1 Introduction	33
2.2 Réseaux de neurones artificiels	34
2.2.1 Neurones artificiels	34
2.2.2 Les fonctions d'activation	35
2.2.3 La normalisation	36
2.2.4 Architectures des réseaux de neurones	37
2.2.5 L'algorithme de rétro-propagation du gradient	39
2.2.6 Sélection des hyper-paramètres	43
2.3 Apprentissage non supervisé	44
2.3.1 Machine de Boltzmann restreinte	44
2.3.2 Auto-encodeur	46
2.3.3 Auto-encodeur débruiteur	47

2.4	Architectures profondes	48
2.5	Pré-apprentissage non supervisé	52
2.5.1	Deep Belief Networks	53
2.5.2	Empilement d'auto-encodeur débruiteur	54
2.6	Apprentissage approfondi pour la prédiction des séries temporelles	54
2.7	Conclusion	59
3	Analyse de données pour la prévision	61
3.1	Introduction	61
3.2	Données de la charge électrique Algérienne	62
3.3	Caractéristiques de la consommation électrique en Algérie	63
3.3.1	Tendance	64
3.3.2	Saisonnalité	65
3.3.3	Cycle Hebdomadaire	66
3.3.4	Cycle Journalier	66
3.4	Procédures de la prédiction de charge	68
3.4.1	Caractère cyclique de la consommation	68
3.4.2	Les données de prévision météorologique	69
3.4.3	L'activité économique	69
3.4.4	Les événements exceptionnels	70
3.5	Choix des paramètres du modèle	70
3.6	La relation entre la température et la consommation	71
3.7	Évaluation de la qualité de prédiction	73
3.8	Conclusion	75
4	Contribution	77
4.1	Introduction	77
4.2	Méthode proposée	80
4.2.1	Pré-traitement	80
4.2.2	Clustering	81
4.2.3	Estimation des profils de température	82
4.2.4	Normalisation des données	83
4.2.5	Stacked Denoising Auto-encoders	84
4.3	Expérimentations	85
4.4	Résultats	87
4.5	PREVOS-DZ	95
4.5.1	Prévision de la charge électrique	97

Table des matières	vii
4.5.2 Technologies utilisées	97
4.5.3 Fonctionnalités de PREVOS-DZ	98
4.6 Conclusion	102
Conclusion	105
A Calcul de la durée du jour	109
A.1 Calcul de la durée du jour	109
Bibliographie	111

Table des figures

1.1	La valeur ajoutée à l'agriculture par travailleur en Algérie. (Banque mondiale 2018)	6
2.1	Illustration qui schématise un neurone artificiel avec un index k	34
2.2	La fonction d'activation sigmoïde.	36
2.3	La fonction tangente hyperbolique	36
2.4	Réseaux de neurones à propagation avant avec 2 couche cachées et une couche de sortie avec 2 neurones.	38
2.5	Illustration d'un auto-encodeur	46
2.6	Illustration d'un auto-encodeur débruiteur en utilisant un masque à zéro	48
2.7	Illustration du schéma de pré-apprentissage non supervisé par couches, dans un RNA avec 3 couches cachées. [Goodfellow 2016]	53
3.1	Position des villes couvertes par les données de consommation régionale de charge électrique.	63
3.2	Courbes de la consommation électrique en Algérie et la tendance estimée entre 2000 et 2016	64
3.3	Produit intérieur brut et population en Algérie entre 2000 et 2016. (Banque mondiale 2018)	65
3.4	La moyenne de consommation mensuelles d'électricité en Algérie de 2000 à 2016 avec un pas de 4 ans	66
3.5	La moyenne de consommation horaires par jour de semaines en Algérie en 2008 et 2015.	67
3.6	Image de corrélation de la consommation d'électricité en Algérie entre 2013 et 2016 dans les heures quotidiennes	68
3.7	Erreur de prédiction par rapport au nombre d'entrées (heure précédent).	71
3.8	Pointe de consommation journalière d'électricité en Algérie entre 2013 et 2016 par rapport à la température maximum	72
4.1	Organigramme de l'approche.	81
4.2	La méthode du coude pour Clustering Kmeans	82
4.3	Visualisation de dix observations aléatoires de chaque clusteur.	88

4.4	Les clusters de K-means de la charge électrique algérienne (données réduit par ACP)	88
4.5	Nombre d'observations par jour de la semaine et par mois dans chaque classe.	89
4.6	La variations de la température horaires estimées sur une période de 24 heures.	90
4.7	Résultats de prévision de 24 heures en utilisant les modèles sélectionnés RNA et SDAE pour chaque Clusteur.	93
4.8	Volet accueil de PREVOS-DZ	98
4.9	Volet données de PREVOS-DZ	98
4.10	Volet profils de PREVOS-DZ	99
4.11	Volet prévisions RLM de PREVOS-DZ	100
4.12	Volet prévisions hebdomadaires de PREVOS-DZ	102
A.1	Schéma de l'orbite de la Terre. [Wikipedia 2018a]	109

Liste des tableaux

1.1	Applications de la prévision de la charge électrique [Wood 2012] . . .	9
1.2	Tableau récapitulatif classant certaines études sur la prévision de charge électrique à court terme par : Type d’approche utilisée, Horizon et Pays.	30
2.1	Des études sur l’utilisation de l’apprentissage profondi pour la prédiction des séries temporelles dans secteur énergétique et météorologique	59
3.1	Intervalle de temps du jours des 5 modèles RNA parallèles	67
4.1	Résumé des entrées et sorties	86
4.2	Paramètres utilisés dans différentes configurations expérimentales de SDAE et RNA séparément. Le symbole U signifie uniformément échantillonné.	87
4.3	Taille des données pour l’apprentissage, la validation et le test. . . .	89
4.4	Effet d’estimation du profil de température sur MAE pour chaque cluster	91
4.5	MAE et MAPE de chaque modèle de cluster en utilisant la régression linéaire multiple	91
4.6	Moyenne et l’écart type pour MAE et MAPE de chaque modèle de Cluster sur les ensembles de tests calculés sur les 10 meilleures erreurs de validation.	92
4.7	MAPE du modèle global et moyenne pondérée des modèles RNAs sur les ensembles de test calculés sur les 10 meilleures erreurs de validation.	92
4.8	Paramètres utilisés pour obtenir les résultats des modèles SDAE sélectionnés.	92
4.9	Paramètres utilisés pour obtenir les résultats des modèles RNA sélectionnés.	92
4.10	Evaluation du Modèle Saisonnier (été) pour les deux mois Juillet-Aout 2016.	94
4.11	Performances de certaines études sur la prévision de charge électrique à court terme	96
4.12	Fonctionnalités du volet RLM et RNA	101

Liste des acronymes

Acronyme	Description
STLF	Prédiction de la charge électrique à court terme
AR	Auto-régressif
ARMA	Auto-Régressifs et Moyenne Mobile
RLM	Régression linéaire multiple
RBF	Réseaux à Fonctions de Base Radiales
RNR	Réseau de neurones récurrents
SE	Systèmes expert
SIF	Systèmes à Inférences Floues
NFIS	Systèmes à base de réseaux neuroniques et de logique floue
ANFIS	Système d'inférence adaptative neuro-flou
SOM	Cartes auto-adaptatives
RBM	Machine de Boltzmann restreinte
AE	Auto-Encodeur
DAE	Auto-Encodeur débruiteur
DBN	Deep Belief Networks

SDAE Empilement d'auto-encodeur débruiteur

Introduction

Contexte général

Le système d'énergie électrique est une infrastructure critique, qui influence le développement économique, à travers le monde, car les industries dépendent presque entièrement de son utilisation. Ainsi, c'est un système complexe qui comprend trois étapes, à partir de la génération dans les centrales électriques, la transmission vers les centres de charge à travers le réseau de transmission et ensuite la distribution aux consommateurs industriels et résidentiels à travers un réseau de circuits de distribution. L'exploitation et la planification efficaces et optimales des systèmes de production d'énergie électrique ont toujours occupé une place importante dans le secteur de l'énergie électrique, une augmentation ou une diminution de la charge du système entraîne une augmentation ou une diminution de la production d'énergie.

La prédiction de la charge à court terme joue un rôle important dans les opérations quotidiennes des systèmes de gestion de l'énergie électrique. Elle fournit des données qui pourrait aider à éviter les contingences, dus à des charges inattendues et fournir des informations essentielles pour la prise de décision et les études de flux de charge afin de contrôler tout risque technologique ou économique. Ainsi, les nouveaux paradigmes et les derniers avancés des systèmes électriques reposent sur l'utilisation de l'intelligence artificielle à plusieurs étapes de sa couche physique, donnant naissance à des concepts tels que le Smart Grids et le Smart Buildings. Afin de minimiser les pertes de transmission et optimiser la production et la consommation.

Une série temporelle est une séquence d'observations indicées par le temps. Ce type de données peut être trouver dans divers domaines tels que l'économie, la météorologie, le commerce, l'ingénierie, etc. Il y a plusieurs raisons d'enregistrer et d'analyser les données d'une série temporelle, notamment pour développer un modèle statistique capable de prédire le comportement futur de la série sur la base des connaissances du passé, ce qui aide les décideurs à faire des choix entre des alternatifs d'actions.

Généralement, les séries temporelles de la charge électrique sont complexes, elles présentent un comportement non stationnaire et dépendant, en plus de ses données historiques, de nombreux facteurs aléatoires exogènes rendent la tâche de prévision difficile notamment les changements saisonniers et météorologiques. La consommation électrique suit des variations journalières et hebdomadaires similaires en cours de l'année, et par conséquent comprendre les modèles de consommation d'énergie est

utile pour concevoir et valider un modèle en termes de mesure de précision, où une attention particulière doit être accordée aux comportements de charge, les activités industrielles et les conditions météorologiques.

Différents modèles de prédiction peuvent être conçus pour chaque type de jour et il existe de nombreuses techniques et des travaux de recherche qui ont été appliqués sur des données provenant de plusieurs pays. Ces techniques sont généralement basées soit sur les approches traditionnelles qui incluent une certaine variation des modèles classiques des séries temporelles, soit sur l'intelligence artificielle (IA) qui a connu de nombreuses expériences et des tests pratiques réussies, particulièrement en utilisant les réseaux de neurones artificiels. Différentes approches hybrides de ces techniques (techniques classiques et des techniques basé sur l'IAs) ont également été appliquées pour augmenter l'efficacité du modèle.

Les Réseaux de Neurones Artificiels (RNA) ont connu une tendance prometteuse dans le domaine de l'intelligence artificielle et plus particulièrement les perceptrons multi-couches en utilisant l'algorithme d'apprentissage par rétro-propagation du gradient. Le progrès de recherche dans le domaine d'apprentissage automatique a ouvert le défi à des problèmes de grande complexité ce qui nécessite une architecture de RNA profonde avec plusieurs niveaux de non-linéarités leur permettant de représenter des fonctions fortement non-linéaires.

Cependant, l'optimisation à base de gradient à partir de l'initialisation aléatoire d'un large réseau à plusieurs couches cachées apparaît généralement moins performante que les RNA avec une ou deux couches cachées. Ce phénomène peut être expliqué par deux hypothèses :

- La descente de gradient peut facilement se coincer dans un mauvais minimum local.
- Si la descente de gradient est capable de trouver un bon minimum en termes d'erreur d'apprentissage, il n'y a aucune garantie que la configuration des paramètres associés fournira une bonne performance de généralisation.

Dans cette thèse, nous nous intéressons à des stratégies d'apprentissage plus efficaces permettant l'apprentissage des architectures profondes du RNA. Ces stratégies ont montré des promesses pour la modélisation de données statiques dans des domaines telles que la reconnaissance d'objets et leur application aux données de séries chronologiques soulève de plus en plus d'attention.

L'objectif de cette thèse est de construire un modèle de prédiction de charge électrique à court terme, afin d'aider les opérateurs de dispatching, en vue d'optimiser la production et la consommation d'énergie électrique. Tout en étudiant les

algorithmes proposés dans le cadre de l'apprentissage profond des RNAs dans le but de proposer des améliorations conceptuelles aux algorithmes existants afin d'améliorer en temps d'apprentissage et en performance l'approximation des séries temporelles.

Motivé par les nouveaux progrès d'apprentissage des RNAs profond d'une part, et la capacité potentielle de ces derniers, d'autre part, nous essayons d'améliorer en performance ces larges RNAs dans la prédiction de charge électrique à court terme.

Dans l'approche proposée dans cette thèse, les courbes de la charge électrique quotidienne en Algérie sont regroupées à l'aide de l'algorithme de clustering K-Means pour créer des classes de profils de charge. Bénéficiant des avancées récentes dans les architectures des réseaux de neurones profonds, une modélisation en utilisant l'empilement d'auto-encodeur débruiteur est établit et testées sur les données de charge de chaque type de jours séparément.

Plan de thèse

La thèse proposée s'articule autour de 4 chapitres et une conclusion afin de résumer notre contribution et souligner les futures perspectives :

Chapitre : État de l'art : Ce chapitre présente quelques notions sur les séries temporelles. Par la suite, l'accent est mis sur l'importance du problème de prédiction des séries temporelles et plus particulièrement la prédiction de la charge électrique à l'horizon court terme. Un état de l'art est présenté sur les différents travaux qui ont été réalisés dans la littérature sur cette problématique au cours des dernières années.

Chapitre : Réseaux de neurones artificiels profonds : Ce chapitre présente la théorie des réseaux de neurones artificiels et les problématiques rencontrées dans l'apprentissage des architectures profondes avec une focalisation sur les réseaux de neurones de type à propagation avant qui représente le paradigme utilisé dans cette thèse. Ainsi que les blocs de construction et les motivations des recherches récentes sur les réseaux de neurones profonds. Par la suite, des recherches sur l'approximation des séries temporelles sont présentées, particulièrement en utilisant les réseaux de neurones profonds.

Chapitre : Analyse de données pour la prévision : Ce chapitre décrit les données utilisées dans les expérimentations de cette thèse qui sont fournies par la compagnie chargée de la production, du transport et de la distribution de l'électricité et du gaz en Algérie (SONELGAZ). Ensuite, une analyse introductive de la charge

électrique en Algérie est montrée, cette analyse est importante pour le développement d'un modèle de prédiction et elle permet d'extraire ses caractéristiques pour mieux comprendre son évolution.

Chapitre : Contribution : Ce chapitre présente la contribution de cette thèse pour la prévision de charge électrique à court terme par une approche en deux étapes. Premièrement, l'algorithme de classification K-Means est utilisé pour définir un ensemble de classes de jour de profils et une méthode d'estimation de profile de température est présenté. Dans la deuxième étape, l'empilement d'auto-encodeur débruiteur est utilisé pour réaliser indépendamment des modèles de prédiction de chaque type de jours. Une discussions sur les résultats obtenus avec cette approche et une présentation du logiciel PREVOS-DZ qui concrétise cette études sont établi à la fin de ce chapitre.

Conclusion : Cette partie présente une conclusion sur les travaux réalisés dans cette thèse, ainsi qu'une discussion sur les perspectives des recherches futures.

État de l'art

Sommaire

1.1	Introduction	5
1.2	Les séries temporelles	6
1.2.1	Types de séries temporelles	6
1.2.2	Objectifs de l'analyse de série temporelle	7
1.2.3	La prédiction des séries temporelles	8
1.3	Prévision de la charge électrique	9
1.4	Les approches traditionnelles	11
1.4.1	Modèles de séries temporelles stochastiques	11
1.4.2	Modèles de régression	14
1.4.3	Filtre de Kalman	16
1.5	Approches basées sur l'Intelligence Artificielle	17
1.5.1	Réseau de neurones artificiels	17
1.5.2	Systèmes expert	22
1.5.3	Systèmes d'inférence flous et les systèmes à base de réseaux neuroniques et de logique floue	25
1.5.4	Machine à vecteurs de support	26
1.6	Conclusion	30

1.1 Introduction

La prédiction d'événements et de conditions futurs est très importante dans de nombreux types d'organisations car les prédictions d'événements futurs doivent être intégrées dans le processus de prise de décision. Toute organisation doit être capable de faire des prévisions afin de prendre des décisions intelligentes.

Dans la prévision des événements qui se produiront à l'avenir, il faut compter sur l'information concernant les événements qui se sont produits dans le passé. C'est-à-dire, pour préparer une prévision, nous devons analyser les données passées et baser la prévision sur les résultats de cette analyse. Ces données doivent être

collectées d'une série temporelle. Dans notre cas le phénomène observé est celui de la consommation horaire d'électricité.

Les séries temporelles peuvent être rencontrées dans multitude de secteurs comme dans le domaines de l'ingénierie, de la sociologie et de l'économie. Dans ce chapitre nous allons voir quelques notions sur les séries temporelles. Par la suite, nous allons présenter un état de l'art sur les différents travaux qui ont été fait dans la problématique de la prédiction de la charge électrique à l'horizon court terme.

1.2 Les séries temporelles

Une série temporelle est une collection de valeurs d'un certain phénomène observé dans le passé, indicées par le temps, noté $\{X_t : t \in T\}$. Pour chaque t , X_t est une variable aléatoire dont on a une observation, x_t .

Si une série temporelle ne peut être exprimée qu'en termes de distributions de probabilités, elle est dite stochastique ou statistique. Sinon elle est dite déterministe, si elle peut être exprimée comme une fonction du temps.[Kriebol 1999]

La figure 1.1 présente un exemple d'une série temporelle de la productivité du secteur agricole en Algérie entre 1980-2016, Où le graphique suggère la possibilité d'ajuster une tendance quadratique ou exponentielle sur les données.

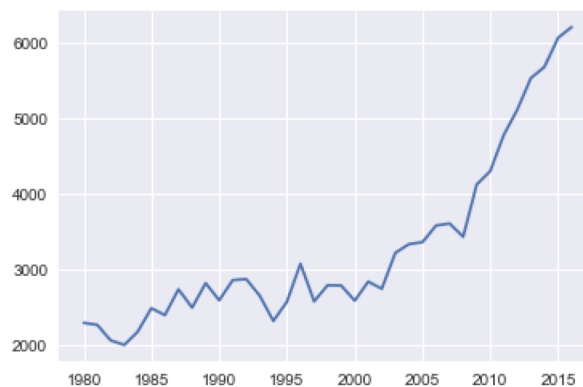


FIGURE 1.1 – La valeur ajoutée à l'agriculture par travailleur en Algérie. (Banque mondiale 2018)

1.2.1 Types de séries temporelles

Les observations d'une série temporelle peuvent être continues dans le temps ou échantillonnées comme un ensemble discret de points dans le temps. Il y a par

conséquent deux types de séries temporelles :

1.2.1.1 Série temporelle continue

On dit une série temporelle continue lorsque les observations sont faites continuellement dans le temps, telle qu'une mesure de l'activité électrique du cœur enregistrée à partir d'un Electrocardiogramme. Ce type de série temporelle est échantillonnée à des intervalles de temps assez petit et égaux de tel sorte que peu ou pas d'information est perdu. [Chatfield 2000]

1.2.1.2 Série temporelle discrète

On dit une série temporelle discrète lorsque les observations sont prises seulement à des moments précis, généralement également espacés. [Kriebol 1999]

Ce type de série temporelle est généralement survenu de trois manière [Chatfield 2000], à savoir :

- Échantillonnée à partir d'une série continue, par exemple : température mesurée à des intervalles horaires.
- Agrégée sur une période de temps, par exemple : total des ventes par mois.
- Une série intrinsèquement discrète, par exemple : les dividendes versé par une société au cours des années successives.

1.2.2 Objectifs de l'analyse de série temporelle

Pour des raisons souvent économiques, il est intéressant d'analyser l'historique d'une variable. L'analyse d'une série temporelle vise à étudier sa structure interne (saisonnalité, tendance, etc.) afin de comprendre le processus qui l'a produit, et parmi les principaux objectifs de l'analyse des séries temporelles on peut citer [Palit 2006] :

- Décrire et classifier les données en utilisant des statistiques récapitulatives et / ou des méthodes graphiques.
- Trouver un modèle statistique approprié pour décrire le processus par lequel les données des séries temporelles sont générées.
- Prédire le comportement futur de la série sur la base des connaissances du passé, ce qui aide les décideurs à faire des choix entre des alternatifs d'actions.

Principalement, ces objectives sont non exclusifs et interconnectées. Par exemple, la prédiction d'une série temporelle demande la description et la modélisation de la série temporelle.

Afin de pouvoir atteindre les objectifs de l'analyse des séries temporelles, des concepts issus de différentes disciplines comme les probabilités et les statistiques sont utilisés. Dans cette thèse, diverses approches de la prédiction des séries temporelles seront considérées et leur application pour la prévision de la charge électrique sera démontrée.

1.2.3 La prédiction des séries temporelles

La prédiction est l'un des objectifs principaux de l'analyse des séries temporelles. Elle consiste à prédire certaines valeurs futures de la variable observée en utilisant uniquement des valeurs passées provenant de la même variable (prédiction univariée) ou en utilisant des données historiques provenant de plusieurs variables corrélées qui peuvent aider à expliquer la variation de la série temporelle (prédiction multivariée) [Kling 1985]. Ce qui implique l'identification d'autres variables liées à la variable à prédire. Une fois ces variables connexes identifiées, un modèle statistique décrivant la relation entre ces variables et la variable à prédire est développé.

Généralement, la prédiction d'un point de vue mathématique est une approximation d'une fonction, qui peut être inconnue, dont l'objectif est de reconstruire une fonction qui permet d'approcher la fonction cible. C'est à dire, au lieu d'une formule explicite, seulement un ensemble de points (la série temporelle) est fourni. [Wikipedia 2018b]

Pour construire des modèles de série temporelle qui permettent de décrire le mécanisme générant la série, la méthodologie Box et Jenkins (Box et Jenkins, 1976) propose une chaîne d'actions itératives qui inclut 4 phases [Palit 2006] :

- Phase d'identification du modèle, dans laquelle on détermine le nombre de paramètres du modèle nécessaires afin d'assurer que le modèle mathématique à construire approxime avec une précision désirée aux données de la série temporelle étudiée.
- Phase d'estimation du modèle, dans laquelle les valeurs des paramètres du modèle sont estimées en minimisant la somme des carrés des résidus.
- Phase de validation du modèle, dans laquelle la qualité de l'ajustement des données est vérifiée et l'amélioration possible du modèle est établie.
- Phase de prévision du modèle, dans laquelle le modèle est utilisé pour établir les limites de confiance de la prévision.

Cette méthodologie, peut être précédée par un nettoyage des données collectées, par la détection et la correction des valeurs qui peuvent mener à des fausses modélisations. Une fois qu'un modèle satisfaisant est développé, il peut être utilisé de

TABLE 1.1 – Applications de la prédiction de la charge électrique [Wood 2012]

Fonction	Horizon de prévision	Intervalles de prévision
Dispatching économique	Heure prochaine	30 secondes
Écoulement de puissance optimal	2 jours suivants	5 minutes
Stabilité de la tension	Heure prochaine	120 échantillons par seconde
Planification de l'opération des unités de production (UC)	14 prochains jours	Horaire
Planification énergétique	Prochain 1-6 mois	Hebdomadaire

différentes manières en fonction du domaine d'application.

1.3 Prédiction de la charge électrique

Le système d'énergie électrique est une infrastructure critique et complexe, à partir de la génération et la transmission de l'énergie aux circuits de distribution. La principale préoccupation de ce système est de fournir en permanence aux clients une énergie électrique fiable et de bonne qualité.

La plupart des fonctions de décision ou de contrôle dans ces systèmes nécessitent une connaissance des conditions futures, en particulier de la demande. Ce qui implique que l'un des principaux objectifs d'une entreprise d'électricité est de prévoir avec précision les besoins des clients.

Le tableau 1.1 présente un résumé des utilisations actuelles et potentielles des prévisions de la charge électrique par différentes unités de système d'énergie électrique.

Par exemple, le but de l'unité de la répartition économique est de trouver la politique d'exploitation optimale c'est à dire, la détermination des sorties de puissance individuelles pour les unités génératrices. De plus, on trouve la planification de l'opération des unités de production (Unit Commitment (UC)) qui sert à résoudre le problème de décision : quand et quelle unité de production devrait être utilisée pour fournir le coût d'exploitation minimal qui satisfieraient parmi la demande prévue un ensemble complet d'unités génératrices [Wood 2012].

Par conséquent, les scénarios de prédiction sont réalisés sur différents horizons temporels à savoir [Soliman 2010] :

1. **Court Terme** : La prévision à court terme comprend une période d'une demi-heure à une semaine d'avance, et qui fait l'objet de cette thèse. Cette prévision met en jeu le pilotage du système en temps réel par un dispatching de la charge électrique entre les clients. Afin que les coûts de fonctionnement quotidien et la répartition soient minimisés.
2. **Moyen Terme** : La prévision à moyen terme inclut la prédiction des pics mensuels et annuels jusqu'à une période d'une année. Elle porte sur : la planification des besoins en énergie, la préparation des plans de charge des ressources disponibles, les plans d'approvisionnements par exemple en combustibles ou encore les programmes d'entretien des équipements.
3. **Long Terme** : La prévision à long terme consiste à faire la prédiction jusqu'à une période de 50 ans. Elle est utile pour les décisions d'investissement en équipements et en infrastructure de telle sorte que la planification de l'expansion peut être facilitée.

Ces prévisions sont obtenues en utilisant des techniques qualitatives ou quantitatives. Les techniques qualitatives sont souvent utilisées pour les prévisions à long terme où les résultats sont basés sur l'avis d'un ou plusieurs experts du domaine. Ainsi, cette méthode est utile généralement lorsque les données historiques ne sont pas disponibles ou rares [Wood 2012].

Parmi ces trois catégories, la prédiction de la charge électrique à court terme est la pierre angulaire de la mission des compagnies d'électricités. Bien qu'elle est également utile dans d'autres études de réseau hors ligne où les ingénieurs préparent une liste d'actions correctives pour différents types de fautes attendues. Ces actions peuvent être le démarrage des unités génératrices ou le délestage qui consiste à arrêter l'alimentation d'une zone particulière ou d'un groupe de clients afin d'éviter la surcharge du système qui engendre les pannes de courant [Kyriakides 2007].

Cependant, diverses techniques quantitative ont été appliquées pour la construction des modèles de prévision de charge électrique à court terme. Ces approches se regroupent sous deux catégories : les approches traditionnelles et les approches basées sur l'intelligence artificielle. Bien que plusieurs chercheurs ont essayé de combiner les deux approches. Dans la sous-section suivante nous présentons un aperçu des différents travaux qui utilisent ces deux approches.

1.4 Les approches traditionnelles

Les approches traditionnelles incluent une certaine variation des modèles de séries temporelles classiques et les méthodes de régression ainsi que les techniques basées sur le filtrage de Kalman.

1.4.1 Modèles de séries temporelles stochastiques

Dans la méthode des séries temporelles stochastiques, la charge est modélisée comme la sortie d'un filtre linéaire et un bruit blanc. Selon les caractéristiques du filtre linéaire, différents modèles de charge peuvent être formulés [Soliman 2010].

1.4.1.1 Le modèle auto-régressif (AR)

Dans le modèle auto-régressif, une valeur de la série temporelle est expliquée linéairement par ses valeurs passées plutôt que par d'autres variables et un bruit blanc. Un modèle auto-régressif d'ordre p , noté $AR(p)$ est donné par :

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + \varepsilon_t \quad (1.1)$$

Où :

- $\theta_1 \dots \theta_p$ sont les paramètres du modèle.
- ε_t est un bruit blanc.
- p est la valeur précédente la plus ancienne à laquelle $Y(t)$ est régressé.

Plusieurs méthodes peuvent être utilisées pour calculer les paramètres du modèle. [Mbamalu 1993] ont proposé d'utiliser la méthode des moindres carrés dans une procédure d'estimation des paramètres d'un modèle auto-régressif (AR) multiplicatif saisonnier rencontré lors de la prévision de la charge électrique pour l'entreprise du service public canadienne Nova Scotia Power. [Huang 1997] a employé des techniques de construction et de détermination d'un nombre de seuils optimaux pour proposer une méthode de prévision qui utilise des modèles auto-régressifs à seuil. Avec cet algorithme, moins de paramètres sont nécessaires pour capturer la composante aléatoire dans la dynamique de la charge.

1.4.1.2 Le modèle Auto-Régressifs et Moyenne Mobile (ARMA)

Box et Jenkins (1976) [Box 2015] ont développé une méthodologie pour les modèles de moyenne mobile auto-régressive (ARMA). Ces modèles sont souvent appelés modèles Box-Jenkins. Dans ce modèle, la valeur courante de la série temporelle Y_t est

exprimée linéairement en termes de valeurs des périodes précédentes (Y_{t-1}, Y_{t-2}, \dots) et en termes des valeurs actuelles et précédentes d'un bruit blanc ($\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$).

L'ordre du processus ARMA est sélectionné à la fois par la plus ancienne valeur précédente de la série (p) et par la plus ancienne valeur du bruit blanc (q) à laquelle Y_t est régressé. Un modèle ARMA(p, q) est écrit comme suit :

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \phi_q \varepsilon_{t-q} \quad (1.2)$$

[Fan 1994] utilise l'algorithme des moindres carrés pondérés récursifs (WRLS) pour estimer les paramètres du modèle ARMA, pour faire une prévision de charge électrique à court terme adaptative aux conditions météorologiques en temps réel.

[Barakat 1992] propose une méthode pour analyser et prévoir les pics de demande d'un système électrique en pleine croissance. Dans cette méthode, les données des demandes historiques sont lissées ensuite elles sont décomposées en composantes de charge déterministe et stochastique. Le composant de charge déterministe est constitué d'une composante de tendance à long terme et d'une composante de température saisonnière. Tandis que, le composant stochastique est déterminé par un modèle ARMA. Les prévisions produites par le modèle développé se comparent favorablement aux demandes réelles, sauf dans certaines périodes où des écarts inhabituels qui ont rapport aux conditions météorologiques sont observés ou lorsque des événements spéciaux sont signalés.

[Chen 1995] propose un modèle ARMA adaptatif pour la prévision de charge. Cette capacité d'adaptation est réalisée grâce à un simple calcul des coefficients des erreurs d'apprentissage et de l'erreur de prévision à un pas en avant. Les résultats des prévisions ont montré que le schéma adaptatif est plus précis que l'approche conventionnelle de Box-Jenkins, et en particulier pour le cas des 24 heures.

1.4.1.3 Le modèle Moyenne Mobile Auto-régressif Intégré (ARIMA)

Les séries temporelles modélisées en utilisant les modèles AR ou ARMA sont appelées processus stationnaire. Sachant que la série temporelle est dit stationnaire si la structure reste la même, cela signifie que son moyennes et son covariances est stationnaires par rapport au temps. Si le processus modélisé est non stationnaire, il est d'abord transformé en une série stationnaire avant d'être modélisé par le processus AR ou ARMA. Ceci peut être réalisé via un processus de différenciation [Soliman 2010]. En introduisant la définition de l'opérateur retard L , une série temporelle différenciée d'ordre 1 peut être exprimée comme suit :

$$\nabla Y_t = Y_t - Y_{t-1} = (1 - L)Y_t \quad (1.3)$$

Par conséquent, une série temporelle différenciée d'ordre d s'écrit :

$$\nabla^d Y_t = (1 - L)^d Y_t \quad (1.4)$$

La série stationnaire différenciée peut être modélisée comme AR ou ARMA pour donner un ARI ou ARIMA. Une série qui doit être différenciée d fois et a des ordres p et q pour les composantes AR et ARMA (*p. ex.* ARIMA (p, d, q)), le modèle est écrit comme suit :

$$\theta(L)\nabla^d Y_t = \phi(L)\varepsilon_t \quad (1.5)$$

La méthodologie de prévision présentée dans [Elrazaz 1989] utilise la technique ARIMA et a été appliquée avec succès pour prévoir la charge de pointe hebdomadaire. Cette méthodologie est basée sur la décomposition des charges hebdomadaires en trois composantes : la composante tendancielle de la charge, la charge sensible aux conditions météorologiques et la charge cyclique non-météorologique.

[Juberias 1999] a expliqué l'utilisation du modèle ARIMA dans un système de contrôle en temps réel et inclut les mesures de la météorologie comme variable explicative. Cette variable est justifiée comme variable explicative parce qu'elle peut être considérée comme le moyen le plus simple et le plus efficace pour estimer l'influence de la météorologie à chaque heure de la journée.

Les résultats de l'analyse de [Barakat 1990] ont montré que la structure unique de l'énergie et de la demande dans les zones à croissance rapide avec des événements spéciaux cycliques en mouvement est difficile à analyser et à prédire en appliquant directement des méthodes standard. Pour cette raison, les auteurs proposent d'utiliser le modèle ARIMA saisonnier pour isoler les effets de ces événements et prévoir séparément les composants statiques et dynamiques de la demande du système pour avoir des meilleurs résultats.

La méthode de série temporelle ARIMA a été la méthode la plus populaire pour la modélisation de la charge [Hagan 1987]. Cependant, pour de plus longues périodes de prédictions, une composante saisonnière ou périodique doit être incluse dans ces processus. Cette inclusion aboutit à ce que l'on appelle un processus saisonnier, nommé SARMA et SARIMA [Soliman 1997, Nezzar 2016].

Dans [Amjady 2001], une nouvelle approche pour STLF a été développée, qui incorpore la modélisation de série temporelle de l'ARIMA avec la connaissance d'opérateurs humains expérimentés. D'un point de vue mathématique, cette méthode

utilise l'estimation des opérateurs comme prévision initiale. Ensuite, elle combine cette prévision initiale avec les données de la température et de la charge pour obtenir une meilleure prévision.

En général, les méthodes des séries temporelles donnent des résultats satisfaisants s'il n'y a pas de changement dans les variables qui affectent la demande de charge (telles que les variables environnementales ou sociales). S'il y a un changement brusque dans l'une de ces variables, les méthodes des séries temporelles ne sont pas aussi précises [Kyriakides 2007].

1.4.2 Modèles de régression

La régression est l'une des techniques traditionnelles les plus utilisées malgré le grand nombre d'alternatifs. L'objectif général de la régression multiple est d'en savoir plus sur la relation entre plusieurs variables indépendantes ou prédictives, et une variable dépendante ou critère. La régression multiple est basée sur les moindres carrés : le modèle est ajusté de sorte que la somme des carrés des différences des valeurs observées et prédites est minimisée [Weron 2007].

Pour les prévisions de la charge électrique, les méthodes de régression sont généralement utilisées pour modéliser la relation entre la charge et les variables liées aux facteurs météorologiques, au type de jour et à la classe de la clientèle. Ces variables influentes sont identifiées sur la base d'une analyse de corrélation avec la charge. Mathématiquement, le modèle de charge utilisant cette approche peut être écrit comme suit :

$$L_t = a_0 + a_1 L_t^1 + \dots + a_k L_t^k + \varepsilon_t \quad (1.6)$$

Où :

- L_t est la charge.
- L_t^1, \dots, L_t^k sont des variables explicatives corrélées à la charge.
- $a_0 \dots a_k$ sont les paramètres du modèle.
- ε_t est un bruit blanc.

Les variables explicatives peuvent être simples, comme l'humidité, ou des fonctions complexes de variables simples, telles que la différence au carré entre les températures quotidiennes maximales et minimales.

Dû à la relation non linéaire entre les températures et les charges, [Haida 1994] présente une méthode de prévision de charge de pointe quotidienne basée sur la régression multiple avec une technique de transformation, afin de traiter le changement saisonnier, la croissance annuelle de la charge et le dernier changement de

charge quotidien.

La régression non paramétrique présentée dans [Charytoniuk 1998] permet de calculer une prévision de charge directement à partir des données historiques comme une moyenne locale des charges passées observées avec la mesure du voisinage local (en termes de temps, température et autres facteurs pertinents) et les poids spécifiques sur les charges définies par un noyaux associés multivariés.

Ramanathan et al. [Ramanathan 1997] proposent une approche pour la prévision à court terme des charges horaires basée sur un modèle de régression multiple, un pour chaque heure de la journée (avec les week-ends modélisés séparément). Avec une structure d'erreur dynamique ainsi que des ajustements adaptatifs pour corriger les erreurs de prévision des heures précédentes.

El-Hawary et Mbamalu [El-Hawary 1990] présentent une procédure qui exploite la convenance de la fonction d'autocorrélation (ACF) et la fonction d'autocorrélation partielle (PACF) des données des charges précédentes dans l'identification d'un modèle sous-optimal. Les paramètres du modèle identifiés sont estimés de façon itérative à l'aide de la procédure des moindres carrés pondérés. Les résultats montrent des performances supérieures comparés aux résultats basés sur la méthode des moindres carrés ordinaires pour prédire la charge horaire de 24 heures à l'avance. Un autre modèle basé sur des techniques de régression linéaire par les moindres carrés pondérés est présenté par Papalexopoulos et al. [Papalexopoulos 1990], qui tient en compte la charge pendant les vacances en utilisant des variables binaires et la température en utilisant des fonctions de degré de chauffage et de refroidissement. Ainsi, ils utilisent des techniques pour atténuer les effets des erreurs potentielles dans les variables explicatives afin d'éviter que les prévisions de pointe quotidienne ne soient biaisées négativement.

Un algorithme adaptatif sensible à la météo a été présenté dans [Ruzic 2003] pour la prévision de la charge à court terme. Cette méthodologie comprend deux étapes principales : L'énergie quotidienne totale est prédite indépendamment dans la première étape tandis que les charges horaires sont prédites dans la deuxième étape. Tous les paramètres du modèle sont automatiquement calculés et mis à jour en utilisant les données réalisées dans la période d'identification, c'est-à-dire avec des conditions météorologiques similaires. Cette période d'identification est obtenue en utilisant la distance euclidienne entre le jour à prédire et les jours de la base de données.

Dans [Hyde 1997], un modèle de prévision a été élaboré pour identifier deux composantes. Une première composante de charge normale (insensible aux intempéries) identifiée par l'analyse de régression linéaire des données de charge et de météoro-

logie. Une deuxième composante, qui est sensible à la température, est estimée à l'aide des paramètres de l'analyse de régression.

Bien que les méthodes de régression soient largement utilisées par les compagnies d'électricité, elles souffrent d'un certain nombre de limites théoriques. En raison de la difficulté de capturer une relation non linéaire entre la demande de la charge et les facteurs d'influence. Ainsi que les modèles qui ne sont pas très flexibles pour des changements de charge rapides du système [Papalexopoulos 1994].

1.4.3 Filtre de Kalman

Dans cette méthode, la charge est modélisée comme une variable d'état en utilisant une formulation d'espace d'état qui est désignée par deux variables ; l'estimation d'état du système et une mesure de la précision de l'état estimé [Moghran 1989].

[Trudnowski 2001] propose une stratégie de prédiction de charge à très court terme pour améliorer les performances de contrôle de génération automatique (AGC) dans un réseau électrique, en utilisant deux estimateurs de Kalman. Un estimateur prédit les composantes de la charge qui changent lentement tandis que l'autre prédit les composantes les plus rapides. Les paramètres du modèle de Kalman sont déterminés en faisant correspondre la réponse en fréquence de l'estimateur aux résidus de charge. Où la charge totale a une composante stochastique et un composant piloté par plusieurs entrées globales telles que l'heure de la journée, la température, etc.

Park et al. [Park 1991] propose un modèle pour une prédiction de charge d'une journée composé de trois éléments : la charge nominale, la charge type et la charge résiduelle. La charge nominale est modélisée de telle sorte que le filtre de Kalman peut être utilisé et les paramètres du modèle sont adaptés par la méthode des moindres carrés récursifs exponentiellement pondérés. Le composant de chargement de type est extrait pour la prédiction de la charge des week-ends en utilisant une méthode de lissage exponentielle. La charge résiduelle est prédite par le modèle AR, dont les paramètres sont estimés en utilisant la méthode des moindres carrés récursifs.

[Sargunraj 1997] présente un algorithme de filtrage de Kalman avec l'incorporation d'une «fading memory» pour la prévision de charge au niveau des sous-stations. Deux étapes sont effectuées dans cette algorithme qui sont la moyenne qui est d'abord prédite ensuite une correction est incorporée en temps réel en utilisant un retour d'erreur des heures précédentes.

En générale, l'une des principales difficultés dans l'utilisation des techniques basées sur l'intelligence artificiel pour la prédiction de la charge est d'identifier les paramètres du modèles. En particulier, pour le filtrage de Kalman ce sont les ma-

trices de covariance de bruit qui ne sont pas facilement estimées [Moghram 1989].

1.5 Approches basées sur l'Intelligence Artificielle

Les techniques fondées sur l'intelligence artificielle comprennent les systèmes experts, les réseaux neuronaux artificiels, l'inférence floue ainsi que les modèles de neurones flous.

1.5.1 Réseau de neurones artificiels

Dans l'article de synthèse de [Hippert 2001] une collection de documents (publiés entre 1991 et 1999) qui rapportent l'application des RNAs à la prévision de charge à court terme a été examinée. Les Réseaux de Neurones Artificiels (RNAs) sont représentées par un grand nombre d'articles rapportant des expériences réussies et des tests validés.

[Chen 1992] propose une méthode de prévision sensible aux conditions météorologique, le modèle consiste en un RNA principal et trois RNAs de soutien. Le RNA principal est utilisé pour fournir la référence de prévision de base du modèle, les trois autres RNAs supplémentaires sont ajoutés pour augmenter la capacité d'apprentissage du modèle proposé, ce qui permet au modèle de mieux extraire les relations entre les différentes catégories d'entrées (données de charge, température, heure de la journée et jour de la semaine) et d'obtenir une meilleure précision.

Dans [Peng 1992] une approche est proposée pour la sélection des cas d'apprentissage pour le RNA basé sur une mesure de distance minimale entre les entrées et l'utilisation des fonctions linéaires et non-linéaires dans le réseau. Cette stratégie montre l'avantage de contourner le problème des changements drastiques dans les événements inhabituels comme les vacances ou les dérèglements climatiques.

La mise en œuvre d'un RNA pour la prédiction de la charge électrique à court terme est présentée dans [Papalexopoulos 1994]. Où une modification d'un logiciel de prédiction basée sur la régression a permis de modéliser avec précision les événements spéciaux, tels que les vacances, les vagues de chaleur, les vagues de froid et d'autres conditions qui perturbent la configuration normale de la charge. Un RNA avec un algorithme d'apprentissage adaptatif basé sur le filtre de Kalman est présenté dans [Dash 1995b] qui permet aussi de prévoir les charges sensibles aux conditions météorologiques, ainsi le modèle permet de faire la différence entre les changements en semaine et les week-ends.

Pour le développement d'un modèle de prévision pour le Centre de Contrôle de

l'Energie de la société grecque d'énergie publique (PPC), un RNA avec 24 sorties est proposé et une attention particulière a été accordée à la modélisation précise des vacances [Bakirtzis 1996]. Une technique de prévision de charge électrique basée sur la compensation des conditions météorologiques neuronales est présentée dans [Chow 1996]. Cette technique est une généralisation non linéaire de l'approche de Box et Jenkins pour la prédiction des séries temporelles non stationnaires.

[Ranaweera 1996] présente une méthode pour inclure les incertitudes ou les variables d'entrées liées aux conditions météorologiques dans les modèles de prévision de charge électrique basée sur un réseau neuronal. Cette méthode consiste en réseaux neuronaux formés de manière traditionnelle et en un ensemble d'équations qui calcule la valeur moyenne et les intervalles de confiance de la charge prévue. Sachant que la disponibilité des intervalles de confiance aide l'opérateur à prendre de meilleures décisions lorsque les variables d'entrées sont incertaines.

[AlFuhaid 1997] utilise l'algorithme d'apprentissage en cascade des RNAs pour la prévision de la charge électrique à court terme, en incluent la prévision d'énergie maximale, minimale et quotidienne en tant que données d'entrée supplémentaires pour l'étape de prévision finale. Les résultats de la simulation indiquent que cette approche peut être efficace dans une grande variété de conditions d'exploitation du système électrique, y compris les conditions météorologiques changeant rapidement.

[Kiartzis 1997] introduit un prétraitement des données d'entrée afin d'améliorer la précision des prévisions ainsi que la sensibilité du modèle aux changements de température rapides. Parmi les étapes de prétraitement on trouve la phase de création de la courbe de charge "typique". Ces courbes de charge produites sont utilisées comme des modèles de comparaison pour le lissage de la charge. En fait, les courbes typiques contiennent l'historique du comportement de la charge au cours des années précédentes et sont différents pour chaque saison et chaque jour de la semaine (28 courbes sont créées).

[Yoo 1999] a développé un réseau de neurones adaptatifs auto-supervisé pour effectuer des prévisions de charge à court terme (STLF). Cette méthode utilise le réseau auto-supervisé pour extraire les caractéristiques de corrélation des données de température et de charge. Pour les prévisions journalières, les résultats surpassaient les autres systèmes de prévision de charge d'énergie qui utilisaient généralement des structures et des algorithmes très complexes. Cependant, pour les prévisions horaires, les résultats étaient similaires à ceux rapportés dans la littérature.

La méthode de prédiction proposée dans [Senjyu 2002] utilise la sélection des jours similaires au jour de prédiction. Où la puissance de charge prévue est obtenue en ajoutant une correction aux données de jour similaires sélectionnées en se basant

sur le déviation de température. Par conséquent, il n'est pas nécessaire que le réseau de neurones apprenne toutes les données des jours similaires. Ainsi, la structure du réseau neuronal et le temps d'apprentissage peuvent être réduites.

Une prédiction d'ensemble météorologique consiste en plusieurs scénarios pour une variable météorologique. Dans [Taylor 2002], J. Taylor and R. Buizza ont montré comment les prévisions d'ensemble météorologique peuvent être utilisées dans les prévisions de charge par un RNA pour les délais d'un à dix jours. Les scénarios de température, la vitesse du vent et la couverture nuageuse sont utilisés pour produire plusieurs scénarios de charge. Les résultats montrent que la moyenne des scénarios de charge est une prévision de charge plus précise que celle produite en utilisant les prévisions météorologiques traditionnelles. Ainsi, ils utilisent les scénarios de charge pour estimer l'incertitude dans la prévision de charge. Cela se compare favorablement aux estimations basées uniquement sur les erreurs de prévision de charge historique.

[Abdel-Aal 2005] présente une approche formelle pour appliquer la technique de comité qui est une stratégie de division dans laquelle les réponses de plusieurs RNA (experts) sont combinées en une seule réponse en utilisant des données historiques de charge et de température couvrant plusieurs années, avec des membres individuels du comité formés sur différentes années.

La plupart de ces applications ont utilisé des grandes architectures des réseaux neuronaux, ce qui soulève la question méthodologique du sur-apprentissage. Dans [Hippert 2005] H.S. Hippert et al. examine ce problème en comparant plusieurs méthodes de prévision sur un échantillon de demandes d'électricité horaires, y compris les grands réseaux de neurones et les méthodes classiques de régression. Où les résultats montrent des bonnes performances pour les grands réseaux de neurones et ils proposent une analyse montrant que la prévision du vecteur de 24 éléments de la demande quotidienne en électricité peut être particulièrement favorable à cette approche.

Une description et une application originale d'un type de RNA appelé réseaux à Fonctions de Base Radiales (RBF) est présentée dans [Ranaweera 1995]. La capacité prédictive des modèles RBF et leur capacité à produire des mesures précises pouvant être utilisées pour estimer les intervalles de confiance pour les prévisions à court terme. Ces mesures de fiabilité sont très utiles aux opérateurs et fournissent une solution raisonnable aux besoins de non satisfaction dans les industries.

Un autre type de RNA qui est les cartes auto-adaptatives est utilisé dans [Benabbas 2008, Khadir 2011, Benabbas 2011] afin de distinguer les différents types de profils de charge. Cette approche les réseaux de neurones non supervisés avec K-means sur la

demande d'énergie électrique. Les vecteurs de données de charge de l'ensemble d'apprentissage ont d'abord été subdivisés en clusters à l'aide d'un réseau neuronal non supervisé (cartes auto-adaptatives). Cette opération est utile, notamment lorsque l'ensemble d'apprentissage est constitué de données non homogènes.

Une approche de réseau de neurones récurrent pour la prévision de courbe de charge du jour suivant en utilisant le concept de jours similaires est présentée dans [Senjyu 2004]. Une norme euclidienne avec des facteurs pondérés est utilisée pour la sélection de jours similaires. La charge prévue est obtenue en ajoutant une correction aux données des jours similaires sélectionnés et la correction est obtenue à partir du réseau de neurones. Il a été constaté que l'adoption des données de charge corrigées des jours précédents en tant que données de jours similaires donne un meilleur résultat de prévision que l'utilisation des données de charge d'origine des jours passés en tant que données de jours similaires.

Un modèle de prévision de charge à court terme utilisant un réseau neuronal récurrent avec un algorithme d'apprentissage adaptatif afin d'améliorer la vitesse de convergence est proposé dans [Tran 2006]. Le modèle proposé est formulé par une combinaison de plusieurs modèles individuels. La transformée en ondelettes adoptée dans le modèle proposé est utilisée pour décomposer le profil de charge en une représentation multi résolutions. Chaque modèle individuel est utilisé pour prévoir les sous-profil à chaque niveau de résolution obtenu par la transformée en ondelettes.

L'étude de [Ringwood 2001] se concentre sur l'utilisation des RNAs sur trois différentes échelles de temps qui intéresse les centrales électriques : annuelle (jusqu'à quinze ans à l'avance), hebdomadaire (jusqu'à trois ans à l'avance) et horaire (jusqu'à 24 heures à l'avance). La demande d'électricité présente des caractéristiques considérablement différentes à ces différentes échelles de temps, à la fois en termes de processus auto-régressifs et d'entrées appropriées à chaque échelle temporelle.

[Fidalgo 2007] montrent l'importance de la température sur le niveau de consommation de charge et décrit une étude où plusieurs techniques de régression qui ont été expérimentées : k-plus proches voisins, régression linéaire multiple, arbres de régression et RNA. Le RNA ayant obtenu les meilleurs résultats, il a été choisie comme un outil principal pour le processus de prévision de la charge globale au Portugal.

Un autre algorithme proposé dans [Tian 2009], combine efficacement les caractéristiques de l'algorithme d'Optimisation par Essaims Particulaires Quantiques (QPSO) et du réseaux à Fonctions de Base Radiales (RBF). L'application pour la prévision de charge à court terme montre que la méthode peut accélérer la vitesse de convergence du réseau et augmenter la précision de la prévision par rapport au

RBF traditionnel.

[Hamid 2010] propose un algorithme d'apprentissage du système immunitaire artificiel (AIS) pour le RNA. La comparaison du RNA avec l'algorithme d'apprentissage par rétro-propagation montre que la performance de l'algorithme d'apprentissage AIS proposé est comparable à celle de l'algorithme d'apprentissage par rétro-propagation et peuvent conduire à l'amélioration de la prédiction.

Chen et al. [Chen 2010] présente une méthode qui combine la sélection d'un jour similaire, la décomposition en ondelettes et les réseaux de neurones pour prévoir la charge de demain. L'idée est de sélectionner une charge de jour similaire en tant que charge d'entrée basée sur l'analyse de corrélation, et d'utiliser la décomposition en ondelettes et des réseaux neuronaux pour capturer les principales caractéristiques de charge.

L'adaptation de la production d'énergie à la demande a traditionnellement été très importante pour les services publics afin d'optimiser la consommation de ressources. Cela est particulièrement vrai dans les micro-centrales où de nombreux éléments intelligents doivent adapter leur comportement en fonction des conditions de génération et des consommations futures. Les micro-centrales étant des espaces physiques caractérisés par une faible consommation d'électricité par rapport aux grandes régions. [Hernández 2014] présente un modèle basé sur trois niveaux : qui commence par la reconnaissance des formes par une cartes auto-adaptatives, une classification de la partition précédente par l'algorithme k-means, et enfin la prévision de la demande avec un perceptron multi-couches pour chaque cluster obtenus dans le processus de clustering. Le modèle produit de faibles erreurs sur les données d'un environnement de la taille d'un micro-centrale par rapport à d'autres modèles simples qui ne sont pas spécialisés en terme de la classification supervisée et de la classification non supervisée.

Les nouveaux paradigmes et les derniers développements des réseaux électriques reposent sur l'introduction de l'intelligence distribuée à plusieurs étapes dans sa couche physique, donnant naissance à des concepts tels que les réseaux électriques intelligents, centrale électrique virtuelle, Smart Environments et Smart Buildings. Ce qui implique la réalisation d'une gestion adéquate de la demande d'énergie, combinant à la fois des sources conventionnelles et renouvelables. Pour cela plusieurs techniques ont été proposées pour développer et évaluer des modèles de prédiction à court terme de la demande d'électricité basé sur les RNAs [Mena 2014, Platon 2015, Azadeh 2013].

Ainsi, les prévisions précises dans les sous-stations électriques sont obligatoires pour l'efficacité des fonctions d'automatisation de la distribution avancée dans les

systèmes de distribution. [Ding 2016] décrit la conception d'un modèle de RNA, pour les prévisions de charge dans des sous-stations à moyenne tension / basse tension. Les résultats montrent que les modèles basés sur le RNA surpassent les modèles des séries temporelles et que la méthodologie de conception proposée garantit une meilleure capacité de généralisation en terme de capacité de RNA pour cartographier les relations complexes entre les entrées et les sorties.

Extreme Learning Machine (ELM) fait référence à un type de RNA. Sa spécificité est de n'avoir qu'une seule couche cachée, où les poids entre les entrées les nœuds cachés sont jamais mis à jour. Les poids entre des nœuds cachés et les sorties sont appris en une seule étape, ce qui revient essentiellement à l'apprentissage d'un modèle linéaire, donnant un processus d'apprentissage beaucoup plus rapide. Dans [Li 2016], un nouveau modèle de prévision d'ensemble pour la prévision de charge à court terme est proposé basé sur l'ELM. Où plusieurs outils d'apprentissage automatique sont exploitées tel que l'apprentissage ensembliste, les ondelettes, la méthode d'apprentissage Levenberg-Marquardt, la régression des moindres carrés partiels. Ainsi, une méthode de sélection de caractéristiques basée sur l'information mutuelle conditionnelle est développée pour sélectionner un ensemble de variables d'entrée pour le modèle de prévision. La méthode proposée est capable d'atténuer de nombreuses difficultés, telles que les poids et les biais aléatoires du réseau, la sélection des paramètres d'ondelettes et le sur-apprentissage.

Un algorithme génétique est utilisé dans [Bianchi 2015] pour ajuster les paramètres du RNA de type Echo State Network (ESN), où les performances obtenues sur les données de la société qui gère la distribution d'électricité à Rome ont montré que un ESN peut surpasser un système ARIMA dans la prédiction.

[Cecati 2015] étudie l'efficacité de certains algorithmes d'apprentissage automatique tel que les machines à vecteurs de supports et ELM pour la prévision de charge électrique de 24 heures. Afin de les évalués et comparés aux réseaux RBFs proposés. Où des meilleurs résultats en termes d'erreurs d'apprentissage et de validation, de taille de réseau et de temps de calcul ont été obtenus avec l'algorithme proposé.

1.5.2 Systèmes expert

Les systèmes expert sont une branche de l'intelligence artificielle qui fait un usage intensif des connaissances spécialisées pour résoudre un problème dans un domaine. Ces programmes informatiques imitent le processus de raisonnement d'un expert humain ou son comportement de manière experte dans un domaine pour lequel ils ne sont pas disponibles. Ils raisonnent généralement avec des informations incertaines

et imprécises [Kandel 1991].

Un système expert se compose généralement d'au moins trois parties : un moteur d'inférence, une base de faits et une base de règles. La base de faits est comme une base de données qui augmente avec le temps et elle est utilisée pour stocker les informations obtenues de l'utilisateur du système. Tandis que la base de règles contient les connaissances d'un expert du domaine de prévision de charge. Ces connaissances sont représentées comme des faits et des règles Si-Alors, et comprennent l'ensemble des relations entre les changements dans la charge et les changements dans les facteurs qui affectent l'utilisation de l'électricité, comme les conditions météorologiques qui ont été largement examinées afin d'élaborer des règles différentes, autant que la saison et le jour de la semaine. Le moteur d'inférence utilise la base de règles avec la base de faits acquises sur un problème pour fournir une solution experte [Merritt 2012].

L'approche par système expert du problème de prévision de charge développée et testée dans [Rahman 1988] où il a fourni un algorithme fiable qui a la performance souhaitée. Avec l'avantage que les données et les exigences de calcul de l'algorithme sont également minimales. L'algorithme a été développé sur la base des relations logiques et syntaxiques entre le temps et la charge, et les formes de charge quotidiennes dominantes.

L'algorithme proposé dans [Rahman 1993] combine des caractéristiques issues des techniques statistiques et les systèmes à base de connaissances qui utilise la comparaison par paires. Cet algorithme a produit une prévision de charge assez bonne sur un délai de 24 heures pour de nombreux sites divers, toute en incluant les variables météorologiques et leurs effets relatif sur la charge. Parmi les avantages signalés dans cette technique c'est qu'elle est indépendante du site, permet l'intervention de l'opérateur si nécessaire, et elle ne nécessite pas plus de trois ans de données passées. Cependant, elle est facilement personnalisable en ajoutant les caractéristiques dépendantes du site. Ces caractéristiques sont formulées sous la forme de règles de sélection et d'ajustement. Une fois ajoutées, ces règles devraient améliorer les performances de l'algorithme pour un site spécifique.

Une minorité des techniques décrites dans la littérature ont examiné la charge d'énergie électrique d'un point de vue indépendant du site. Une nouvelle technique indépendante du site pour STLF est analysée dans [Rahman 1996]. La charge est modélisée à l'aide d'une base de règles paramétrée et d'une base de données de paramètres. Le modèle de charge, les règles et les paramètres ont été conçus sans connaissances spécifiques sur un site particulier. Cette technique basée sur un système expert est testée en utilisant des données provenant de différents sites à travers

les États-Unis, y compris la Floride, le Massachusetts, le Texas, la Virginie et Washington.

La mise en œuvre en temps réel d'un système expert de prévision appelée ALFA est décrit dans [Jabbour 1988]. Ce système qui tient compte des jours fériés et des événements spéciaux et ajuste la croissance de la charge, utilisant une longue base de données historiques sur 10 ans d'observations horaires de 12 variables météorologiques et la demande de charge. Les observations et les prévisions météorologiques requises sont obtenues en temps réel auprès du Service météorologique national. Les rapports des changements horaires mettent à jour la base de données et permettent la correction adaptative des erreurs.

Un autre système expert à base de connaissances a été développé dans [Ho 1990] qui a été écrit dans PROLOG en utilisant une base de données de 5 ans enregistré à Taïwan. Les connaissances des opérateurs et les observations horaires de la charge du système au cours des ces années sont utilisées pour établir les types de onze jours pour le système. Ce système inclus l'effet des variables météorologiques telles que les températures et les humidités sur la prévision de la charge. L'évaluation de cette méthode a montré une amélioration en terme de performances par rapport à la méthode statistique conventionnelle Box-Jenkins. Parmi les types des charges spéciales considérés par le système expert : la semaine du nouvel an chinois, les caractéristiques de charge spéciales des jours suivant une tempête tropicale, la fermeture partielle de certaines usines le samedi.

Brown R. et al. [Brown 1999] ont proposé une méthode qui associe les connaissances existantes sur le système et la charge, les patterns de charge de croissance et l'information sur l'horizon pour permettre l'élaboration de scénarios multiples pour la progression de croissance des charges à partir des modèles existants. Cette méthode réduit considérablement les besoins en données et en main-d'œuvre par rapport aux techniques de prévision de charge traditionnelles. De plus, la méthode nécessite uniquement des données à un niveau de détail cohérent avec la quantité de détails de prévision souhaitée pour une partie particulière du système électrique.

Généralement, les systèmes experts sont une composante d'un modèle hybride global pour améliorer les performances de prédiction. Par exemple, en Égypte, Mohamad et al. [Mohamad 1996] ont proposé une technique basée sur un modèle généralisé combinant les caractéristiques du RNA et un système expert. La méthodologie le rend particulièrement adapté aux modèles de charge qui sont influencés principalement par les activités sociales, et la météo. Par exemple, les nombreuses occasions sociales qui dépendent des préférences religieuses qui ne peuvent pas être décidées bien à l'avance.

A Taiwan aussi, dans [Chiu 1997] un RNA et combiné avec la sortie fournie par un système expert basé sur des règles est conçu pour démontrer que l'inclusion de la prédiction d'un système expert améliorerait la capacité prédictive du RNA. De plus, l'approche combinée proposée converge beaucoup plus rapidement que le réseau de neurones conventionnel et la méthode du système expert basée sur des règles.

1.5.3 Systèmes d'inférence flous et les systèmes à base de réseaux neuroniques et de logique floue

Le système d'inférence floue est une méthode de calcul basée sur les concepts de la théorie des ensembles flous, des règles Si-Alors floues et du raisonnement flou. Des applications réussies peut être trouver dans plusieurs domaines, tels que la prévision des séries temporelles, la robotique et la reconnaissance des formes. En raison de sa nature multidisciplinaire, le système d'inférence floue est connu par de nombreux autres noms, tels que système basé sur les règles floues, système expert flou [Jang 1997].

La combinaison de la logique floue et les RNAs crée un système hybride avec la capacité à apprendre de l'expérience et la capacité à synthétiser de nouvelles relations entre la demande de charge et les variables d'entrée. L'idée fondamentale derrière cette combinaison est simple : le système de logique floue modélise la connaissance du système et de ses paramètres d'entrée, quantitatifs et qualitatifs, et le RNA capture la relation inexplicable entre les entrées et les sorties floues. Les chercheurs ont proposé plusieurs façons de combinaison afin d'améliorer la performance globale de la prévision. Ces travaux peuvent être classées en quatre catégories[Srinivasan 1995] :

1. Système de logique floue à la sortie de RNA pour manipuler la sortie [Kim 1995].
2. Système de logique floue à l'entrée d'un RNA pour pré-traiter les entrées [Srinivasan 1994].
3. RNA floue intégré pour créer une base de règles floues à partir des données historiques d'apprentissage [Dash 1995a].
4. Système de logique floue et RNA distincte pour prévoir les différents composants de la charge [Gottschalk 1994].

Pour faire face au problème des conditions de charge anormales pendant les vacances,[Kim 2000] propose une technique hybride en utilisant les RNAs et les systèmes d'inférence flous pour améliorer la performance de prédiction dans les jours spéciaux, qui comprennent les jours fériés, les jours fériés consécutifs et les jours

précédant et suivant les jours fériés. Cette méthode comporte un modèle RNA pour prévoir les courbes de charge dans chaque type de jour et deux modèles d'inférence floue sont utilisés pour prévoir les charges maximales et minimales des jours spéciaux. Enfin, les résultats des RNAs et des modèles d'inférence floue sont combinés pour prévoir les charges des 24 heures des jours spéciaux. Les résultats des tests ont montré que la méthode de prévision proposée pourrait améliorer considérablement la précision des prévisions pour les jours spéciaux. Elle pourrait assister efficacement le système conventionnel et l'expert humain.

L'étude de L. Ying et M. Pan [Ying 2008] a adopté le système d'inférence adaptative neuro-flou (ANFIS) pour prévoir les charges d'électricité régionales à Taïwan et a comparé les résultats d'ANFIS avec ceux d'autres modèles tel que : RNA, les machines à vecteur de support avec le modèle des algorithmes génétiques (RSVMG). Les résultats empiriques obtenus par ANFIS dans cette étude, donnent des prévisions de charge régionales plus précises que les autres modèles et démontrent que ANFIS est une alternative valide et prometteuse pour la prévision des charges d'électricité régionales.

Çevik H. et Çunkaş M. [Çevik 2015] ont comparé les résultats de prévision horaire de charge électrique pour la Turquie en utilisant la logique floue et l'ANFIS. Où, les méthodes ANFIS et la logique floue trouvent des valeurs très proches des données réelles et produisent de bons résultats de prévision. Les résultats montrent que le modèle ANFIS pourrait avoir une meilleure précision de prévision que la méthode de logique floue en terme d'erreur de prédiction. Ainsi, les paramètres d'entrée supplémentaires tels que la couverture nuageuse et les variables de vitesse du vent peuvent être utilisés pour produire des résultats encore meilleurs. En outre, pour les pays ayant de grandes superficies, les données de température prises à partir d'un seul point conduiraient à une augmentation d'erreurs de prévision.

1.5.4 Machine à vecteurs de support

Les Machines à Vecteurs de Support (SVMs) sont une méthodologie d'apprentissage qui a connu une attention importante dans beaucoup d'application d'intelligence artificielle et de classification de données. Drucker H. et al. [Drucker 1997] ont proposé une méthode utilisant les SVMs pour résoudre les problèmes de régression, tout en gardant les caractéristiques principales qui caractérisent l'algorithme tel que la minimisation de l'erreur et l'identification de l'hyperplan qui maximise la marge. Vue que la sortie est un nombre réel, une marge de tolérance est établie.

En 2001, le réseau européen d'excellence sur les technologies intelligentes pour

les systèmes intelligents adaptatifs (EUNITE) a organisé un concours visant à prévoir la charge maximale quotidienne des 31 prochains jours. Pendant la compétition, [Chen 2004] ont proposé un SVM qui était l'entrée gagnante pour les données d'hiver seulement. Parmi les différentes solutions proposées, le choix des segments de données appropriés semble améliorer les performances du modèle. Certaines conclusions importantes des résultats sont que la température (ou d'autres types d'informations climatiques) construites avec une estimation continue peuvent être inexacts à la fin et que l'introduction du concept des séries temporelles pourrait améliorer la prévision.

La méthodologie des machines à vecteur de support des moindres carrés (LS-SVM) à taille fixe a été appliquée à la prévision de charge à court terme dans [Espinoza 2005]. Où il est montré qu'il est possible de construire un modèle de régression non linéaire à grande échelle à partir d'un ensemble de données avec 26 000 échantillons en utilisant différents sous-échantillons comme vecteurs de support. Les résultats donnent une performance de prévision satisfaisante pour les deux modes de prévision (une heure en avance et 24 heures en avance) lorsque le mappage non linéaire est calculé en utilisant 800 vecteurs de support sélectionnés, qui représentent 3% de l'échantillon disponible.

Dans [Perez 2008], une discussion de certaines méthodologies pour la prévision de charge est établie pour faire une classification entre différentes techniques basées sur l'intelligence artificielle comme la logique floue et les SVMs. Les résultats montrent que la prédiction faite à l'aide des SVMs rend la prédiction de la demande si réussie en réduisant l'erreur beaucoup plus que la prédiction du système ANFIS. Une analyse comparative est faite pour chaque technique et les résultats présentent les avantages de chaque technique.

Référence	Type d'approche utilisée	Pays	Horizon
[Mbamalu 1993]	AR	Canada	168 heures à l'avance
[Huang 1997]	AR	Taiwan	1/24 heures à l'avance
[Fan 1994]	ARMA	Géorgie	24/168 heures à l'avance
[Barakat 1992]	ARMA	Arabie Saoudite	Pointe de charge mensuelle
[Chen 1995]	ARMA	Taiwan	24/168 heures à l'avance
[Elrazaz 1989]	ARIMA	Arabie Saoudite	168 heures à l'avance
[Juberias 1999]	ARIMA	Espagne	1/6 heures à l'avance

[Barakat 1990]	ARIMA	Arabie Saoudite	Pointe de charge mensuelle
[Hagan 1987]	ARIMA	États-Unis	1/5 minutes à l'avance
[Amjady 2001]	ARIMA	Iran	Pointe de charge quotidienne
[Nezzar 2016]	ARIMA	Algérie	Pointe de charge mensuelle
[Soliman 1997]	ARIMA	Koweït	24 heures à l'avance
[Haida 1994]	RLM	Japon	Pointe de charge quotidienne
[Charytoniuk 1998]	Régression non paramétrique	États-Unis	168 heures à l'avance
[Ramanathan 1997]	RLM	États-Unis	Une heure à l'avance
[El-Hawary 1990]	RLM	Canada	1/24 heures à l'avance
[Papalexopoulos 1990]	ARIMA et RLM	États-Unis	Pointe de charge quotidienne
[Ruzic 2003]	RLM	Serbie	24 heures à l'avance
[Hyde 1997]	RLM	Irlande	1/10 jours à l'avance
[Trudnowski 2001]	Filtre de Kalman	États-Unis	30/120 minutes à l'avance
[Park 1991]	Filtre de Kalman	Corée du Sud	1/24 heures à l'avance
[Sargunaraaj 1997]	Filtre de Kalman	Inde	1/168 heures à l'avance
[Chen 1992]	RNA	États-Unis	1/168 heures à l'avance
[Peng 1992]	RNA	États-Unis	1/168 heures à l'avance
[Dash 1995b]	RNA et Filtre de Kalman	États-Unis	1/24 heures à l'avance
[Bakirtzis 1996]	RNA	Grèce	24 heures à l'avance
[Chow 1996]	RNA et ARMA	Île de Hong Kong	1/24 heures à l'avance
[Ranaweera 1996]	RNA	États-Unis	Pointe de charge quotidienne
[AlFuhaid 1997]	RNA en Cascade	Koweït	30 minutes / 24 heures à l'avance
[Kiartzis 1997]	RNA	Grèce	24 heures à l'avance

[Yoo 1999]	RNA	États-Unis	1/24 heures à l'avance
[Senjyu 2002]	RNA	Japon	1 heures à l'avance
[Benabbas 2008]	SOM	Algérie	1/24 heures à l'avance
[Khadir 2011]	SOM	Algérie	1/24 heures à l'avance
[Benabbas 2011]	SOM	Algérie	1/24 heures à l'avance
[Taylor 2002]	RNA et ARMA	Angleterre et pays de Galles	1/10 jours à l'avance
[Abdel-Aal 2005]	RNA	Arabie Saoudite	Pointe de charge quotidienne
[Hippert 2005]	RNA	Brésil	24 heures à l'avance
[Ranaweera 1995]	RBF	États-Unis	Pointe de charge quotidienne
[Senjyu 2004]	RNR	Japon	1/168 heures à l'avance
[Tran 2006]	RNR	États-Unis	1/24 heures à l'avance
[Ringwood 2001]	RNA	Irlande	1/24 heures à l'avance
[Fidalgo 2007]	RNA	Portugal	1/24 heures à l'avance
[Tian 2009]	RBF	Chine	1/24 heures à l'avance
[Hamid 2010]	RNA	Malaisie/États-Unis	1 heures à l'avance
[Chen 2010]	Réseaux d'ondelettes et RNA	Nouvelle-Angleterre	1/24 heures à l'avance
[Hernández 2014]	RNA	Espagne	1/24 heures à l'avance
[Ding 2016]	RNA	France	Moyenne quotidienne
[Li 2016]	ELM	Nouvelle-Angleterre	1/24 heures à l'avance
[Bianchi 2015]	ESN	Italie	10 minutes / 24 heures à l'avance
[Cecati 2015]	RBF	Nouvelle-Angleterre	24 heures à l'avance
[Rahman 1988]	SE	États-Unis	1/24 heures à l'avance
[Rahman 1993]	SE	États-Unis	1/24 heures à l'avance
[Rahman 1996]	SE	États-Unis	1/24 heures à l'avance
[Jabbour 1988]	SE	États-Unis	1/48 heures à l'avance
[Ho 1990]	SE	Taiwan	1/24 heures à l'avance

[Mohamad 1996]	SE et RNA	Égypte	1/24 heures à l'avance
[Chiu 1997]	SE et RNA	Taiwan	1 heures à l'avance
[Kim 1995]	SIF et RNA	Corée du Sud	1 heures à l'avance
[Srinivasan 1994]	NFIS	Singapour	1/24 heures à l'avance
[Dash 1995a]	NFIS	États-Unis	Pointe de charge quotidienne
[Kim 2000]	ANFIS	Corée du Sud	24 heures à l'avance
[Ying 2008]	ANFIS	Taiwan	Pointe de charge annuelle
[Çevik 2015]	ANFIS	Turquie	1/24 heures à l'avance
[Chen 2004]	SVM	Europe	Pointe de charge quotidienne
[Espinoza 2005]	SVM	Belgique	1/24 heures à l'avance
[Perez 2008]	ANFIS et SVM	Colombie	1/24 heures à l'avance

TABLE 1.2 – Tableau récapitulatif classant certaines études sur la prévision de charge électrique à court terme par : Type d'approche utilisée, Horizon et Pays.

La spécificité de chaque pays conduit à un problème de prévision plus difficile et joue un rôle important dans la modélisation et la sélection des paramètres. La table 1.2 classe certaines études sur la prévision de charge électrique à court terme présentée dans cette section. Avec les données historiques de la consommation, plusieurs types de données sont utilisées afin d'avoir une meilleure performance comme les variables calendaires telles que le type de jours, les jours fériés, la représentation saisonnière et mensuelle. Ces études traitent chaque type de jours séparément ou dans un modèle global. Nous pouvons aussi trouver d'autres paramètres comme : le temps du lever et le coucher de soleil. Ainsi, les variables météorologiques, principalement la température, la vitesse du vent, la nébulosité et l'humidité relative.

De plus, le phénomène observé dans notre étude qui les profils quotidiens, mensuels et annuels de l'électricité Algérienne est très spécifique, à cause du changement climatique de la méditerranée avec quatre saisons distinctes et un pays couvrant plus de 3 types de climat : méditerranéen, hautes terres et déserts, en raison de sa taille et sa position géographique.

1.6 Conclusion

Ce chapitre présentait l'importance du problème de prédiction des séries temporelles et plus particulièrement la prédiction de la charge électrique à court terme, Où

nous avons présenté les méthodes les plus utilisées dans la littérature au cours des dernières années. Ce problème a connue une attention particulière des chercheurs vue son importance vitale pour l'industrie électrique dans l'économie déréglementée.

Réseaux de neurones artificiels profonds

Sommaire

2.1	Introduction	33
2.2	Réseaux de neurones artificiels	34
2.2.1	Neurones artificiels	34
2.2.2	Les fonctions d'activation	35
2.2.3	La normalisation	36
2.2.4	Architectures des réseaux de neurones	37
2.2.5	L'algorithme de rétro-propagation du gradient	39
2.2.6	Sélection des hyper-paramètres	43
2.3	Apprentissage non supervisé	44
2.3.1	Machine de Boltzmann restreinte	44
2.3.2	Auto-encodeur	46
2.3.3	Auto-encodeur débruiteur	47
2.4	Architectures profondes	48
2.5	Pré-apprentissage non supervisé	52
2.5.1	Deep Belief Networks	53
2.5.2	Empilement d'auto-encodeur débruiteur	54
2.6	Apprentissage approfondi pour la prédiction des séries temporelles	54
2.7	Conclusion	59

2.1 Introduction

Nous abordons dans ce chapitre la théorie des neurones artificiels, ainsi que les blocs de construction des réseaux de neurones. Par la suite nous focalisons sur la théorie des réseaux de neurones à propagation avant et nous clôturons avec les

réseaux neuronaux profonds ainsi que les recherches récentes sur l'approximation de la fonction des séries temporelles en utilisant ce type de réseau, qui représente le paradigme utilisé dans cette thèse.

2.2 Réseaux de neurones artificiels

Le cerveau humain peut prendre des 100–200 ms pour reconnaître un visage familier intégré dans une scène inconnue. Il a la capacité de s'adapter et de construire ses propres règles de comportement à travers ce que nous appelons habituellement l'expérience. Les RNAs sont inspirés par le cerveau humain et modélés sur la structure du cerveau biologique. Ils se composent d'un modèle abstrait de neurones interconnectés, dont la structure spécial et la liaison peuvent être utilisées pour résoudre des problèmes d'applications basés sur le calcul dans divers domaines tels que les statistiques, la technologie et l'économie [Haykin 2009].

2.2.1 Neurones artificiels

Les unités de calcul qui font partie d'un réseau de neurones sont appelées des neurones artificiels. La figure 2.1 montre la composition d'un neurone artificiel. Nous identifions trois éléments de base du modèle neuronal [Haykin 2009] :

- Un ensemble de synapses, chacun caractérisé par un poids. Plus précisément, un signal x_i à l'entrée de la synapse i relié au neurone k est multiplié par le poids synaptique w_{ki} .
- Un fonction de combinaison pour sommer les signaux d'entrée.
- Une fonction d'activation pour limiter l'amplitude de la sortie d'un neurone y_k .

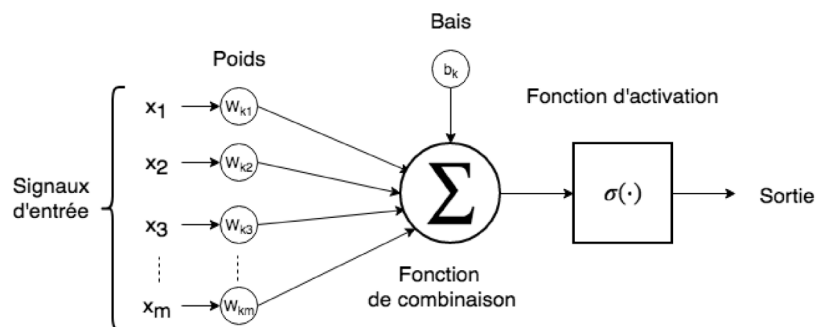


FIGURE 2.1 – Illustration qui schématise un neurone artificiel avec un index k .

En termes mathématiques, le neurone k représenté sur la figure 2.1 peut être décrit par la paire d'équations :

$$a_k = \sum_{i=1}^m w_{ki} x_i \quad (2.1)$$

Et

$$y_k = \sigma(a_k + b_k) \quad (2.2)$$

Le biais b_k est un paramètre externe du neurone k , il est possible de considérer sa présence dans l'équation 2.1 en ajoutant une synapse avec une entrée $x_0 = 1$ et son poids $w_{k0} = b_k$.

2.2.2 Les fonctions d'activation

La fonction d'activation $\sigma(\cdot)$ définit la sortie d'un neurone en fonction de a_k . Dans cette section, nous présentons deux fonctions d'activation : la fonction sigmoïde et la fonction tangente hyperbolique.

La fonction sigmoïde

La fonction sigmoïde est de loin la forme la plus commune de la fonction d'activation utilisée dans la modélisation des réseaux de neurones [Haykin 2009]. La plupart des articles cités dans cette thèse utilisent la fonction sigmoïde qui est différentiable, une propriété très utile lors de l'apprentissage des réseaux de neurones. Un exemple de la fonction sigmoïde est la fonction logistique définie par :

$$\sigma(x) = \frac{1}{1 + \exp^{-\alpha x}} \quad (2.3)$$

Où α est le paramètre de pente de la fonction sigmoïde, on faisant varier le paramètre α , nous obtenons des fonctions sigmoïdes de pente différentes comme illustré dans la Figure 2.2.

La fonction sigmoïde produit des valeurs continues entre 0 à 1.

La fonction tangente hyperbolique

Dans certaines situations, il est préférable d'avoir une fonction d'activation dont la sortie va de -1 à +1, dans ce cas là nous utilisons la fonction tangente hyperbolique définie par :

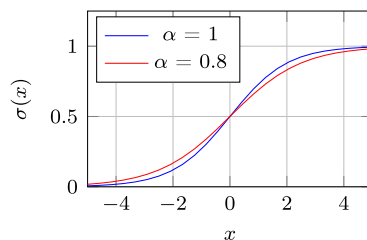


FIGURE 2.2 – La fonction d’activation sigmoïde.

$$\sigma(x) = \tanh(x) = \frac{\sinh(x)}{\cosh(x)} \quad (2.4)$$

La fonction tangente hyperbolique produit des valeurs continues entre -1 à 1, comme illustré dans la Figure 2.3.

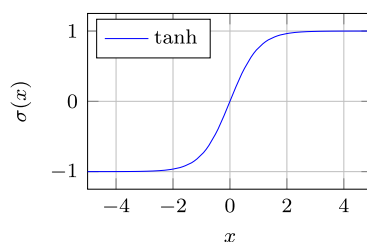


FIGURE 2.3 – La fonction tangente hyperbolique

2.2.3 La normalisation

Généralement dans l’apprentissage automatique, les données sont collectées à partir de plusieurs sources. La collection de ces données doit être effectuée soigneusement pour éviter la redondance et l’incohérence afin d’améliorer la précision et accélèrent le processus d’apprentissage. Ensuite, ces données doivent être transformées en formes adaptées à l’algorithme d’apprentissage. La transformation des données implique la normalisation parmi d’autres processus comme la construction des attributs.

Dans le cas des réseaux de neurones comme d’autre type d’algorithme d’apprentissage automatique, la normalisation est particulièrement utile et permettra d’accélérer la phase d’apprentissage. Un attribut est normalisé en mettant à l’échelle ses valeurs afin qu’elles se situent dans un intervalle spécifié tel que $[0, 1]$. Il existe de

nombreuses méthodes pour la normalisation des données, citons : la normalisation min-max et la normalisation Centrer-Réduire.

Normalisation min-max

La normalisation min-max transforme une valeur x de l'attribut S à x' dans l'intervalle $[nouveaumin_s, nouveamax_s]$, notant que min_s et max_s sont les valeurs minimum et maximum réelles de S . Pour cela, la normalisation min-max effectue une transformation linéaire sur les données d'origine en calculant :

$$x' = (nouveamax_s - nouveaumin_s) * \left(\frac{x - min_s}{max_s - min_s} \right) + nouveaumin_s \quad (2.5)$$

Normalisation Centrer-Réduire

La normalisation Centrer-Réduire est basé sur la moyenne et l'écart-type de l'attribut S . Cette normalisation se calcule comme suit :

$$x' = \frac{x - \mu}{\sigma} \quad (2.6)$$

Où μ est la moyenne de l'attribut S cependant σ est son écart-type. cette normalisation est utile lorsque le minimum et le maximum réels de l'attribut S sont inconnus.

2.2.4 Architectures des réseaux de neurones

Les RNAs peuvent être structurés de différentes manières. Selon la conception du RNA, un algorithme d'apprentissage correspondant est appliqué pour former le réseau. Fondamentalement les architectures des RNAs peuvent être identifier en trois différentes classes [Haykin 2009] :

Réseaux à propagation avant à une seule couche

Dans le réseau de neurones à propagation avant à une seule couche, les neurones sont organisés sous la forme de couches. Dans la forme la plus simple d'un réseau en couches, nous trouvons une couche d'entrée qui projette directement sur une couche de neurones de sortie. En d'autres termes, ce réseau est strictement de type à propagation avant. Un tel réseau est appelé un réseau monocouche, avec la désignation "monocouche" se référant à la couche de sortie [Haykin 2009].

Réseaux de neurones multicouches

Réseaux de neurones multicouches ou perceptrons multicouches sont différents des réseaux à propagation avant à une seule couche par la présence d'une ou de plusieurs couches cachées. Ce type de réseaux fait l'objet de cette thèse. Le terme "caché" fait référence au fait que cette partie du réseau de neurones n'est pas vue directement à partir de l'entrée ou de la sortie du réseau. La tâche de ces unités cachées est de faire partie de l'analyse des données circulant entre les couches d'entrée et de sortie. En ajoutant une ou plusieurs couches cachées, le réseau peut être capable d'extraire plus d'informations de ses entrées [Haykin 2009].

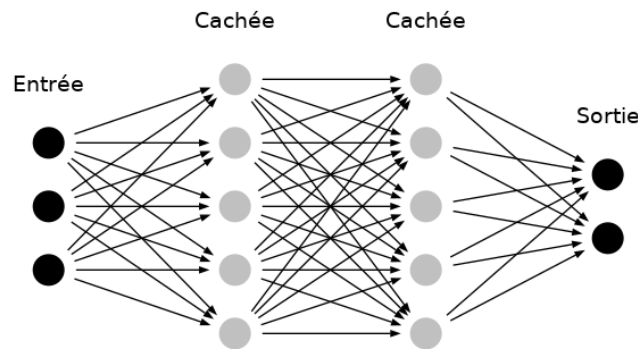


FIGURE 2.4 – Réseaux de neurones à propagation avant avec 2 couche cachées et une couche de sortie avec 2 neurones.

La figure 2.4 illustre l'architecture d'un réseau de neurones à propagation avant multicouches pour le cas de deux couches cachées. Où chaque neurone dans chaque couche du réseau est connecté à tous les autres neurones dans la couche avant adjacente.

Réseaux de neurones récurrents

Un réseau de neurones récurrent se distingue d'un réseau neuronal à propagation avant par le fait qu'il comporte au moins une boucle de rétro-action. Par exemple, un réseau récurrent peut être constitué d'une seule couche de neurones, chaque neurone transmet son signal de sortie aux entrées de tous les autres neurones [Haykin 2009].

2.2.5 L'algorithme de rétro-propagation du gradient

La méthode la plus courante pour mettre à jour les poids dans la théorie des réseaux de neurones est l'algorithme du rétro-propagation du gradient [Rumelhart 1986]. Cet algorithme met à jour les poids en calculant le gradient de l'erreur pour chaque neurone du réseau, de la dernière couche vers la première.

L'algorithme de rétro-propagation est appliqué en deux phases. Dans la première phase qui est la phase de propagation, au cours de laquelle chaque échantillon de l'ensemble d'apprentissage passe à travers les poids synaptiques d'une couche à la suivante, jusqu'à ce que les données sortent finalement dans les neurones de sortie tout en ajoutant les biais comme suit :

$$a^{(1)} = x \quad (2.7)$$

$$z^{(2)} = w^{(1)} a^{(1)} \quad (2.8)$$

$$a^{(2)} = g(z^{(2)}) \quad (2.9)$$

Jusqu'à :

$$z^{(L)} = w^{(L-1)} a^{(L-1)} \quad (2.10)$$

$$a^{(L)} = y = g(z^{(L)}) \quad (2.11)$$

Où :

- x est le vecteur d'entrée.
- L représente le nombre de couches dans le réseau.
- g est la fonction d'activation d'un neurone.
- $a_i^{(j)}$ est l'activation du neurone i dans la couche j .
- $w^{(j)}$ est la matrice des poids contrôlant le mappage de la couche j vers la couche $j + 1$.
- y représente la sortie.

Dans la deuxième phase qui est la phase de rétro-propagation, au cours de laquelle nous commençons par les nœuds de sortie et traversons toutes les couches dans le réseau en calculant récursivement le gradient, δ , pour chaque neurone dans chaque couche. Pour un neurone j , la règle de delta est donné par :

$$\Delta w_j = \alpha (target_j - y_j) g'(z_j) x = \alpha \delta_j x \quad (2.12)$$

Où :

- α est le pas d'apprentissage
- $target_j$ est la sortie cible.
- z_j est la somme pondérée des entrées du neurone.

De cette façon, nous mettons à jour les poids synaptiques en fonction de la règle de Delta (l'équation 2.12). Dans la couche de sortie, δ est simplement l'erreur (voir l'équation 2.13) multipliée par la dérivée première de sa fonction d'activation.

$$E = \sum \frac{1}{2} (target - y_j)^2 \quad (2.13)$$

$$\delta_j = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial z_j} = (target_j - o_j) o_j (1 - o_j) \quad (2.14)$$

Nous utilisons l'équation 2.14 si la fonction d'activation utilisée est la fonction sigmoïde pour calculer les changements de poids de toutes les connexions menant à la couche de sortie ($o_j = y$). Lorsque nous obtenons δ de la couche de sortie, nous pouvons alors continuer en calculant les δ s pour la couche juste avant la couche de sortie à l'aide de l'équation 2.15.

$$\delta_j = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial z_j} = \left(\sum_{\ell \in V} w_{j\ell} \delta_\ell \right) o_j (1 - o_j) \quad (2.15)$$

Où :

- V sont les neurones recevant la sortie du neurone j .

Ce calcul récursif se poursuit couche par couche en propageant les changements à tous les poids synaptiques du réseau. En résumé, l'algorithme de rétro propagation fonctionne en passant les données dans une propagation avant et une propagation arrière de l'erreur. L'algorithme de rétro propagation est présenté dans le pseudo code 1.

Il y a deux façons de transmettre les données à travers l'algorithme durant l'apprentissage. Ces méthodes sont appelées : mode d'apprentissage séquentiel et mode d'apprentissage par lots (batch). La différence majeure est que l'algorithme par lots maintient les poids du réseau constants tout en calculant l'erreur associée à chaque échantillon d'apprentissage. Cependant, la version séquentiel met constamment à jour ses poids, son calcul d'erreur et l'estimation de gradient utilise des poids différents pour chaque échantillon d'apprentissage. Il existe encore un troisième mode, appelé apprentissage par mini lots (mini-batch) qui est la méthode choisie pour l'apprentissage dans les expérimentations de cette thèse. Ce mode est une hybridation

Algorithm 1 L'algorithme de rétro-propagation du gradient

L'ensemble d'apprentissage : $(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})$ où m est le nombre d'échantillons d'apprentissage.

Initialiser les poids de réseau (petites valeurs aléatoires)

Répéter

Pour i de 1 jusqu'à m **Faire**

PASSAGE D'AVANTS

Mettre $a^1 = x^i$ (Le vecteur d'entrée)

Faire la propagation en avant pour calculé $a^{(l)}$ pour $l = 2, 3, \dots, L$

PASSAGE ARRIÈRE

Utiliser $y^{(i)}$ pour calculer l'erreur sur les neurones de sortie.

Calculer Δw pour tous les poids de la dernière couche cachée vers la couche de sortie en utilisant l'équation 2.14.

Calculer Δw pour tous les poids de la couche d'entrée vers les couches cachées en utilisant l'équation 2.15.

Mette à jour les poids.

Fin Pour

Jusqu'à Tous l'ensemble d'apprentissage soient classés correctement ou qu'un autre critère d'arrêt soit satisfait.

et le calcul de l'erreur et l'estimation de gradient se fait sur un sous-ensemble de données.

Plusieurs choix de conception qui ont été testés dans le but d'améliorer la performance de l'algorithme de rétro-propagation. Parmi les méthodes éprouvées et qui ont amélioré significativement les performances, on peut citer cette liste [Haykin 2009] :

1. Maximiser le contenu de l'information pour rechercher plus dans l'espace de poids. Soit, en utilisant des exemples d'apprentissage présenté à l'algorithme de rétro-propagation qui rendent l'erreur d'apprentissage plus grande en se basant sur leurs contenu d'informations. Ou en utilisant des exemples qui sont différents de tous ceux utilisés précédemment. Ceci peut se faire en randomisons l'ordre dans lequel les échantillons d'apprentissage sont présentés au RNA d'une époque à l'autre.
2. Pour accélérer le processus d'apprentissage, le choix le plus judicieux est d'utiliser une fonction d'activation sigmoïde ou la fonction tangente hyperbolique $g(v) = a \tanh(bv)$ avec $a = 1,7159$ et $b = \frac{2}{3}$.
3. Afin d'éviter que l'algorithme de rétro-propagation conduit les paramètres libres du réseau à l'infini, ce qui ralentit le processus d'apprentissage et engendre la saturation des couches cachées. Les valeurs cibles doivent se situer dans la plage de la fonction d'activation sigmoïde. Plus précisément, la valeur maximale de la réponse souhaitée $target_j$ pour le neurone j dans la couche de sortie du RNA doit être décalée d'une certaine valeur ε par rapport à la valeur limite de la fonction d'activation sigmoïde.
4. Chaque variable d'entrée doit être pré-traité de sorte que sa valeur moyenne, calculée sur les échantillons d'apprentissage, soit proche de zéro [LeCun 1993]. La normalisation des les entrées devraient également inclure aussi les deux mesures ci-dessous afin d'accélérer le processus d'apprentissage par l'algorithme de rétro-propagation [LeCun 1993] :
 - Il est préférable que les variables d'entrée de l'ensemble de d'apprentissage ne soient pas corrélées.
 - La covariance devrait être approximativement la même assurant que les différents poids synaptiques du réseau apprennent à peu près à la même vitesse.
5. L'initialisation des poids synaptiques doit éviter les valeurs trop élevées et les valeurs trop petites. Il est recommandées d'initialiser les poids en tirant des valeurs aléatoires à partir d'une distribution uniforme de moyenne nulle et une variance égale à $\frac{1}{\text{le nombre de connexions synaptiques d'un neurone}}$.

6. L'objectif de l'apprentissage à partir d'exemples est de former un modèle pouvant représenter le mappage entrée-sortie de la fonction inconnue $f(\cdot)$. S'il y a une connaissance préalable de $f(\cdot)$, cette connaissance devrait être utilisée dans l'apprentissage du réseau. Les indices de la fonction $f(\cdot)$ peuvent être utilisés pour accélérer la recherche de sa réalisation approximative, ainsi pour améliorer la qualité de l'estimation finale.

2.2.6 Sélection des hyper-paramètres

En pratique, nombreux algorithmes d'apprentissage automatique tel que les RNAs impliquent des hyper-paramètres avec les données d'entrées et sorties, ces hyper-paramètres doivent être définies avant d'optimiser les paramètres du modèle. Les RNAs peuvent avoir plusieurs hyper-paramètres, y compris ceux qui spécifient la structure du réseau lui-même, comme le nombre de couches cachées et le nombre de neurones par couche, et ceux qui déterminent la façon dont le réseau est formé tel que le pas d'apprentissage et le critère d'arrêt. Généralement, on enveloppe l'algorithme d'apprentissage disons A par un autre algorithme B , pour sélectionner les hyper-paramètres, par exemple on se base sur l'erreur d'ensemble de validation. Parmi les méthodes de sélection de ces hyper-paramètres on peut citer :

Recherche manuelle

En utilisant les connaissances acquises sur le problème, On suit un processus itératif jusqu'à trouver une configuration des hyper-paramètres qui fonctionnent bien. A chaque étape de ce processus, on essaie des paramètres et on observe le résultat, ensuite on modifie les hyper-paramètres sur la base de ces résultats.

Recherche en grille

Le RNA est entraîné en utilisant toutes les combinaisons de paramètres afin de sélectionner la combinaison qui fonctionne le mieux. Ces hyper-paramètres sont identifiés sur des intervalles basés sur les connaissances acquises sur le problème et généralement répartis uniformément. Après la sélection des hyper-paramètres il est possible de refaire le processus de recherche sur un domaine plus restreint centré sur les paramètres les plus performants [Laroche 2007]. Cette méthode elle a l'avantage de faciliter le lancement de la recherche pour plusieurs paramètres en parallèle. Par contre, le fait de changer un seul hyper-paramètre, ça rend le calcul redondant et demande une forte capacité computationnelle.

Recherche aléatoire

Cette méthode est semblable et a les même avantages que la recherche en grille, c'est à dire en utilisant les connaissance d'un expert sur le problème pour identifier des intervalle pour les hyper-paramètres. Cependant, au lieu de choisir les valeurs d'une manière méthodique, on sélectionne aléatoirement, ce qui rend le calcul moins redondant vu que les hyper-paramètres change à chaque fois. Il est possible d'arrêter le processus de recherche de la combinaison qui fonctionne le mieux à n'importe qu'elle moment et plus ce processus roule plus la recherche est meilleur. D'après ces avantages, Bergstra et Bengio dans l'article de la recherche aléatoire [Bergstra 2012], propose qu'elle soit la méthode de benchmark pour d'autres méthodes d'optimisation des hyper-paramètres. Cette méthode a l'inconvénient qu'elle demande un forte capacité computationnelle et les résultats sont difficile à reproduire.

Optimisation Bayésienne

Des travaux plus récents ont été mise sur l'amélioration des approches comme la recherche en grille, en utilisant l'information acquise à partir d'une expérience donnée pour décider comment ajuster les hyper paramètres pour l'expérience suivante. Un exemple de ce travail serait l'optimisation bayésienne proposé par Snoek et al. [Snoek 2012]

2.3 Apprentissage non supervisé

Dans cette section nous allons présenter des RNAs pour l'apprentissage non supervisé, c'est à dire en utilisant seulement les entrées $x(t)$ des données d'apprentissage sans les sorties désirées comme dans le cas de l'apprentissage supervisé. Ce type d'apprentissage peut être utile pour :

- Extraire automatiquement des caractéristiques significatives à partir des données.
- Profiter de la disponibilité abondante des données non étiquetées.
- Ajouter une régularisation dépendant des données dans l'apprentissage.

2.3.1 Machine de Boltzmann restreinte

Une machine de Boltzmann restreinte (RBM pour Restricted Boltzmann Machine) [Hinton 2006] est un type de RNA constitué d'une couche d'unités visibles qui reçoit les entrées et une couche d'unités cachées et peut être représentée comme un graphe biparti. Le terme restreinte est pour dire qu'il n'y a pas de connexion entre

les neurones de la même couche, toutes les unités visibles sont connectées à toutes les unités cachées, et il n'y a pas de connexions visibles-visibles ou cachées-cachées.

Dans une RBM, la distribution de probabilité $p(v, h; \theta)$ sur les unités visibles v et les unités cachées h , étant donné les paramètres du modèle $\theta = (w, b, c)$, est défini en termes d'une fonction d'énergie $E(v, h; \theta)$:

$$E(v, h; \theta) = - \left(\sum_{i=1}^m \sum_{j=1}^n w_{ij} v_i h_j + \sum_{i=1}^m b_i v_i + \sum_{j=1}^n c_j h_j \right) \quad (2.16)$$

Où v_i sont des neurones binaires visibles et h_j sont des neurones cachés binaires. b_i et c_j sont des biais correspondant aux neurones cachés et visibles, respectivement, w_{ij} sont des poids reliant les neurones visibles et cachés. Il est possible de calculer la distribution de probabilité conditionnelle, quand v et h sont données, comme suit :

$$P(h_j | v, \theta) = z \left(\sum_{i=1}^m w_{ij} v_i + c_j \right) \quad (2.17)$$

$$P(v_i | h, \theta) = z \left(\sum_{j=1}^n w_{ij} h_j + b_i \right) \quad (2.18)$$

où z est la fonction sigmoïde.

Nous utilisons la fonction d'énergie (équation 2.16) dans les RBMs de type Bernoulli-Bernoulli, c'est à dire pour des applications avec des caractéristiques binaires comme un texte codé ou une image noir-blanc. Lorsque les entrées sont des valeurs continues, telles que dans la prédiction de la charge électrique, nous utilisons les RBMs de type Gaussian-Bernoulli et la fonction d'énergie est donnée par :

$$E(v, h; \theta) = - \left(\sum_{i=1}^m \sum_{j=1}^n w_{ij} v_i h_j + \sum_{i=1}^m \frac{(v_i - b_i)^2}{2\sigma_i^2} + \sum_{j=1}^n c_j h_j \right) \quad (2.19)$$

Où σ_i est l'écart-type de la distribution gaussienne pour le neurone visible v_i .

Carreira-Perpinan et al. [Carreira-Perpinan 2005] ont proposé l'idée de l'algorithme Divergence Contrastive pour l'apprentissage des RBMs et qui permet l'ajustement des paramètres $\theta = (w, b, c)$, en maximisant la distribution de probabilité représentée par RBM

$\frac{1}{M} \sum_{i=1}^M \log p(v = v^i)$ avec les données d'entrée d'apprentissage M .

2.3.2 Auto-encodeur

L'Auto-Encodeur (AE) [Bengio 2009] est un RNA à propagation avant formé pour reproduire ses entrées à la couche de sortie. Autrement dit, il n'est pas formé pour trouver la distribution d'une classe à la sortie mais pour reproduire l'entrée.

Initialement un auto-encodeur a une seule couche cachée et la couche de sortie a la même taille que la couche d'entrée. Le RNA est entraîné en quelque sorte en comparant les valeurs produites à la sortie par le réseau de neurones avec celles d'entrées. Donc, le RNA est entraîné de reproduire à la couche de sortie aussi parfaitement que possible la valeur qui était dans l'entrée.

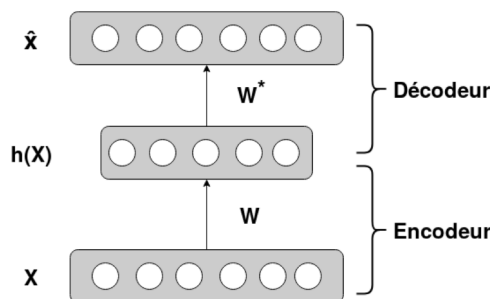


FIGURE 2.5 – Illustration d'un auto-encodeur

Dans la terminologie de l'auto-encodeur, nous faisons référence à la partie du modèle qui calcule la couche cachée $h(x)$ par l'*encodeur*, qui va encoder l'entrée et utilise généralement la sigmoïde de la transformation linéaire comme dans un réseau de neurones régulier. Ainsi, on fait référence à la partie du modèle qui calcule la couche de sortie par le *décodeur* qui prendra la représentation latente de $h(x)$. La sortie notée par \hat{x} va être l'entrée décodée basée sur la représentation latente extraite par l'auto-encodeur.

L'AE est un type du réseau de neurones qui d'abord *encode* les entrées dans une couche cachée, puis *décode* pour obtenir une reconstruction que nous appelons \hat{x} .

$$H = h(x) = \text{sigmode}(Wx + b) \quad (2.20)$$

$$\hat{x} = \text{sigmode}(W^*H + c) \quad (2.21)$$

Une pratique fréquente consiste à lier les poids entre la couche d'entrée W et les poids de la couche de sortie W^* par une égalité, c'est à dire W^* sera défini comme la transposition de W .

Une motivation pour l'apprentissage d'un AE est que la représentation cachée conserve toutes les informations sur l'entrée, par exemple si la couche cachée est beaucoup plus petite que la couche d'entrée, le réseau va compresser l'information, il va ignorer l'entrée qui n'est pas utile pour la reconstruction et concentrer seulement sur ce qui rend les entrées spéciales.

L'apprentissage d'un AE consiste à ajuster les paramètres W et W^* à l'aide d'une descente de gradient afin de minimiser l'erreur de reconstruction L sur l'ensemble d'apprentissage M .

$$L(x, \hat{x}) = \sum_{i=1}^m (x_i - \hat{x}_i)^2 \quad (2.22)$$

Cependant, pour la terminologie de la taille de la couche cachée, nous distinguons deux types. Une couche cachée est dite *incomplète* si elle est plus petite que la couche d'entrée. Dans ce cas, la couche cachée "compresse" l'entrée et les unités cachées seront de bonnes caractéristiques pour l'apprentissage.

Ainsi, si la taille de la couche cachée est supérieure à la couche d'entrée, elle est appelée *sur-complète*. Dans ce cas, pas de compression dans la couche cachée et chaque unité cachée pourrait copier un composant d'entrée différent, et il n'est pas garanti que les unités cachées vont extraire une structure significative. Cependant dans [Bengio 2007a], les résultats pratiques rapportent que la couche cachée sur-complète peut donner des représentations utiles.

Pour faire face à ce problème, des Auto-Encodeur débruiteur ont été proposés, avec ces derniers il est possible d'obtenir un ensemble d'unités cachées qui extrait une structure intéressante de l'ensemble d'apprentissage.

2.3.3 Auto-encodeur débruiteur

Le principe de l'auto-Encodeur débruiteur (DAE pour Denoising AutoEncoders) [Vincent 2010] est au lieu que l'encodeur alimenté de l'entrée originale x on alimente une version bruyante \tilde{x} . c'est une entrée (x) passée à travers un processus de bruit. Généralement le type de ce processus de bruit est de :

- Assignment aléatoire du sous-ensemble d'entrée à 0 avec probabilité v
- Bruit additif Gaussien

Avec du bruit, le comportement de l'auto-encodeur de copier chaque entrée sera éliminer, et on ajoutant un bruit, le RNA est forcé d'apprendre sur la structure de la distribution d'entrée et il est forcé aussi d'apprendre des fonctionnalités plus significatives, où :

- Masque à zéro : reconstruction \hat{x} calculée à partir de l'entrée corrompue \tilde{x}

— La fonction du coût compare la reconstruction de \hat{x} avec l'entrée original x .

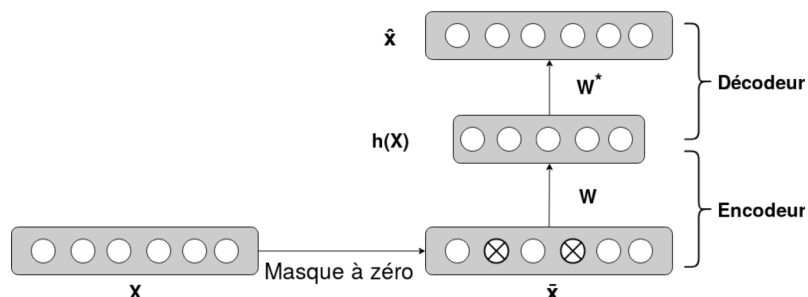


FIGURE 2.6 – Illustration d'un auto-encodeur débruiteur en utilisant un masque à zéro

Dans les type d'auto-encodeur avec ou sans débruiteur, l'apprentissage consiste à optimiser les paramètres W et W^* , itérativement à l'aide d'une descente de gradient afin de minimiser l'erreur de reconstruction sur l'ensemble d'apprentissage, jusqu'à l'atteinte d'un critère d'arrêt.

2.4 Architectures profondes

Historiquement, la faisabilité des RNAs était une source de doute et de désaccord dans la communauté d'intelligence artificiel pour la plupart des années 1970. Jusqu'à la découverte du théorème d'approximation universelle en 1989 par plusieurs chercheurs indépendants qui indique qu'une seule couche cachée de largeur illimitée est probablement suffisante pour qu'un perceptron multicouches (PMC) exprime une fonction continue avec une précision souhaité [Hornik 1989, Cybenko 1989]

Malheureusement, ce théorème ne dit rien sur la façon de construire ou d'ajuster un tel réseau, ni sur son efficacité dans le calcul. Comme il a été révélé, l'expression de fonctions complexes est très coûteuse pour un RNA à deux couches où la couche cachée a besoin de beaucoup plus de neurones. Au lieu de cela, les chercheurs se sont tournés vers des réseaux étroits avec plusieurs couches. Les RNAs profonds sont attrayants en principe : lors de la résolution d'un problème compliqué, il est plus facile de le décomposer en petits morceaux, puis de s'appuyer sur leurs résultats.

Il est plus facile pour l'être humain de décomposer un problème compliqué en petites parties, puis de s'appuyer sur leur résultats. Par exemple, dans la compréhension du langage parlé, il est logique de reconnaître des motifs phonétiques au lieu d'essayer de reconnaître chaque mot directement à partir d'une forme d'onde.

Cette décomposition est attrayantes d'un point de vue calculatoire et en pratique les RNAs modernes ont largement réussi à reproduire ce type de comportement d'après l'analyse de visualisation des couches cachées comme dans [Strobelt 2018].

Dans cette partie nous allons discuter, certains problèmes rencontrés dans l'apprentissage des RNAs profonds. Ces RNAs sont construit d'une architecture dans laquelle il y a plus d'une couche cachée ou de nombreuses couches cachées. Cette profondeur fait référence au nombre de couches ou niveau que le RNA comprend. Chaque couche effectue des calculs non linéaires sur les données entrées afin d'apprendre la fonction désirée.

Les valeurs des poids de neurones changent en fonction des données et, au départ, tous ces poids sont initialisés de manière aléatoire. Il y a beaucoup de travaux sur la distribution exacte de ces valeurs aléatoires [Glorot 2010] et parmi les choix de conception recommandés, elles doivent être petites et non identiques (Section 2.2.5). Pendant longtemps, les chercheurs ont essayé de faire l'apprentissage des RNAs constitués d'une architecture profonde mais souvent ils atteignent une limite au delà de 3 niveaux (avec deux couches cachées) [Bengio 2009].

Des études théoriques ont démontré que les architectures profondes sont nécessaires pour la modélisation des fonctions complexes avec une bonne performance de généralisation [Bengio 2009]. Une tâche qui peut être difficile à gérer pour des architectures peu profondes. Il est démontré aussi que les fonctions qui peuvent être représentées de manière compacte par une architecture de profondeur k peuvent nécessiter un nombre exponentiel d'unités de calcul représentés par une architecture de profondeur $k - 1$ [Bengio 2007b]. Ce nombre d'unités de calcul consomme beaucoup de ressources pour les ajuster. Ainsi, Ces unités dépend du nombre d'exemples d'apprentissage qui doit croître au moins linéairement pour avoir une bonne généralisation [Baum 1989]. Donc, cette dépendance engendre à la fois des problèmes de performance calculatoire et statistique.

Généralement, l'apprentissage d'un RNA profond ou non, consiste à résoudre le problème d'optimisation, où la fonction à minimiser représente la différence entre les résultats réels et les estimations des réseaux. Pour simplifier, supposons que nous avons un ensemble d'exemples d'apprentissage $z = (x, y)$ de taille m et que nous voulons résoudre une tâche d'apprentissage supervisé, mais le principe peut également être étendu à d'autres types d'apprentissage. En utilisant un modèle M avec des paramètres θ initialisé aléatoirement, nous pouvons produire une estimation \hat{y} , qui est la sortie de $M(x)$. Cependant, puisque nous avons maintenant un ensemble de résultats réels et l'estimation correspondante à notre modèle, nous pouvons comparer les deux et voir à quel point notre modèle était. C'est ce que l'on appelle

l'erreur d'apprentissage représenté par la fonction de coût J qui est généralement la moyenne de la fonction d'erreur e :

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m e(M_\theta, z_i) \quad (2.23)$$

Nous utilisons la méthode d'optimisation à base de gradients afin d'ajuster les paramètres du modèle itérativement jusqu'à ce que la différence devienne minimal ou un critère d'arrêt soit satisfait. Cependant, un modèle profond avec une initialisation aléatoire des paramètres donne habituellement des mauvaises performances de généralisation [Bengio 2007b].

En effet, la fonction de coût étant non convexe, rien ne garantit que la descente de gradient ne restera pas coincée dans un mauvais minima local [Auer 1996]. Des résultats de recherche dans [Fukumizu 2000] montrent qu'il existe des minima locaux dans l'espace des paramètres, qui causent des plateaux puisque tous les points autour de l'ensemble des minima locaux convergent vers lui et doivent y échapper par fluctuation aléatoire. Ces résultats ne dépendent pas de la forme spécifique des fonctions d'activation ni des fonctions de coût. Ce qui permet d'expliquer la mauvaises performances de généralisation des réseaux profonds par l'augmentation du nombre de couches cachées dans l'architecture du réseau. Ainsi, initialisation aléatoire des poids synaptiques influence clairement la chance de se situer dans l'un de ces minima locaux.

Dans l'apprentissage automatique, nous parlons des données étiquetées et non étiquetées, la première faisant référence au cas où il y a une cible ou un résultat connu auquel la sortie du RNA peut être comparée. Contrairement au deuxième cas où il n'y a pas de cible. Les algorithmes d'apprentissage en profondeur sont censés de gérer les problèmes suivants [Bengio 2009] :

- Capacité d'apprendre des fonctions complexes et hautement variables, c'est-à-dire avec un nombre de variations bien supérieur au nombre d'exemples d'apprentissage.
- Capacité d'apprendre les abstractions de bas niveau avec peu d'intrants humains, intermédiaires et de haut niveau qui seraient utiles pour représenter le type des fonctions complexes nécessaires.
- Capacité d'apprendre avec un temps de calcul d'apprentissage proportionnelle au nombre d'exemples et être proche du linéaire.
- Capacité d'apprendre dans le cadre semi-supervisé, c'est-à-dire avec des données non étiquetées ou ne sont pas accompagnées des bonnes étiquettes associées.

- Capacité d'exploiter les synergies présentes dans un grand nombre de tâches, c'est-à-dire l'apprentissage multi-tâches.
- Apprentissage fort non supervisé qui capture une grande partie de la structure statistique dans les données observées.

La difficulté d'apprentissage des RNAs profonds revient aussi à la propriété de l'algorithme de rétro-propagation qui propage l'erreur de la dernière couche jusqu'à la première. Le problème est que lorsque le gradient de la fonction de coût se propage vers l'arrière, il est multiplié par les valeurs de poids de chaque couche. Si ces valeurs de poids sont inférieures à un, le gradient rétrécit exponentiellement (il disparaît) ; si elles sont supérieures à un, elles croissent exponentiellement (elles explosent). On appelle ce phénomène par le problème de Dilution du gradient (Vanishing gradient problem) [Hochreiter 2001]. De nombreuses solutions ont été proposées au fil des années. Certaines impliquaient de tenter d'initialiser les poids d'un réseau de sorte que le gradient ne diverge pas rapidement [Glorot 2010]. Des techniques plus modernes comme la normalisation par lots [Ioffe 2015] offrent également des solutions directes. Le passage aussi à l'utilisation de ReLU [Glorot 2011] comme une fonction d'activation peut aider à trouver de meilleurs minima pendant l'apprentissage, car il permet aux gradients de croître au-dessus de 1, contrairement aux fonctions de saturation comme la fonction sigmoïde ou la tangente hyperbolique.

Ainsi, en 2006, Hinton et al. [Hinton 2006] ont montré comment former une architecture de RNA profond (Deep Belief Networks (DBN)) en entraînant une couche à la fois. Le principe est de former les niveaux intermédiaires en utilisant un apprentissage non supervisé, qui est effectué localement à chaque niveau.

Les RNAs profonds sont formés de manière non supervisé, mais dans de nombreuses applications, une fois qu'une bonne représentation a été trouvée à chaque niveau, elle peut être utilisée pour initialiser et former un réseau de neurones profond par une optimisation supervisée basée sur le gradient.

Nous introduisons dans la section suivante l'une des techniques de construction des RNAs profonds et son fonctionnement pour atteindre une meilleure performance de généralisation, cette technique est dite pré-entraînement non-supervisé. Nous allons nous concentrer sur l'empilement d'auto-encodeurs débruiteurs pour construire une architecture profonde qui est le modèle proposé pour notre tâche d'approximation de fonctions pour la prédiction.

2.5 Pré-apprentissage non supervisé

La régularisation dans l'apprentissage automatique est l'un des concepts important qui diminue le problème de sur-apprentissage et améliore la généralisation. L'idée principale derrière la technique de régularisation est de pénaliser des modèles complexes, c'est-à-dire de définir une fonction de pénalité pour quantifier la complexité du modèle. Par exemple dans la régularisation L2, ce terme de pénalité est ajouté à la fonction du coût de problème d'optimisation comme suit, où λ est le paramètre de régularisation ($0 < \lambda \ll 1$) :

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m e(M_{\theta}, z_i) + \frac{\lambda}{m} \sum_{i=1}^m \|\theta\|^2 \quad (2.24)$$

En ajoutant ce terme de régularisation, les valeurs des matrices des poids diminuent, et évitent la configuration idéale en terme d'apprentissage. Cependant, l'application de la régularisation ne représente que très peu d'informations sur ce que devrait être la généralisation. En effet, les chercheurs continuent d'écrire des articles où ils essaient différentes approches de la régularisation, telles que le Décrochage (Dropout) [Srivastava 2014].

La difficulté d'optimisation associée aux modèles profonds a été empiriquement atténuée lorsque Hinton et al. [Hinton 2006] ont publié un article sur un algorithme d'apprentissage basé sur l'utilisation d'un pré-apprentissage non supervisé. Cet article est généralement crédité de fonder le regain d'intérêt pour l'apprentissage des modèles profonds et d'initialiser les poids d'un RNA et produisant souvent de bien meilleurs résultats qu'avec les poids aléatoires.

Dans [Hinton 2006], un modèle génératif profond composé d'une pile de machines de Boltzmann restreintes (RBM) est introduit et appelé Deep Belief Network (DBN). Cette technique comprend deux phases principales d'apprentissage : la phase de pré-apprentissage et la phase de réglage fin (en anglais : fine tuning) qui signifie l'apprentissage selon un critère supervisé, comme illustré dans la figure 2.7. Le protocole se déroule en 4 phases (une par couche cachée, plus la phase de réglage fin), de gauche à droite. Pour les étapes non supervisées, chaque couche (gris foncé) est formée pour apprendre une meilleure représentation de la sortie de la couche précédemment formée (initialement, l'entrée brute). Ces représentations apprises par apprentissage non supervisé forment l'initialisation d'un réseau supervisé profond, qui est ensuite entraîné (affiné) comme d'habitude (dernière phase, droite), tous les paramètres étant libres de changer (gris foncé) [Goodfellow 2016].

Dans la phase de pré-apprentissage, les couches du modèle sont entraînées d'une

manière non-supervisé couche après couche à leur entrée, dans l'ordre de bas en haut, en commençant par la couche qui est en contact direct avec les données d'entrée. Ensuite, la deuxième couche inférieure est entraînée en prenant les activations des unités cachées de la première couche comme entrée pour l'apprentissage non supervisé. Le pré-apprentissage se déroule de la manière suivante : chaque couche s'entraînant sur la sortie ou les activations des unités cachées de la couche inférieure. Après le pré-apprentissage de la dernière couche, une couche supervisée est placée au-dessus, et toutes les couches sont formées par rapport au critère global d'une manière supervisée. En d'autres termes, la pré-apprentissage est seulement utilisé pour initialiser un RNA profond supervisé.

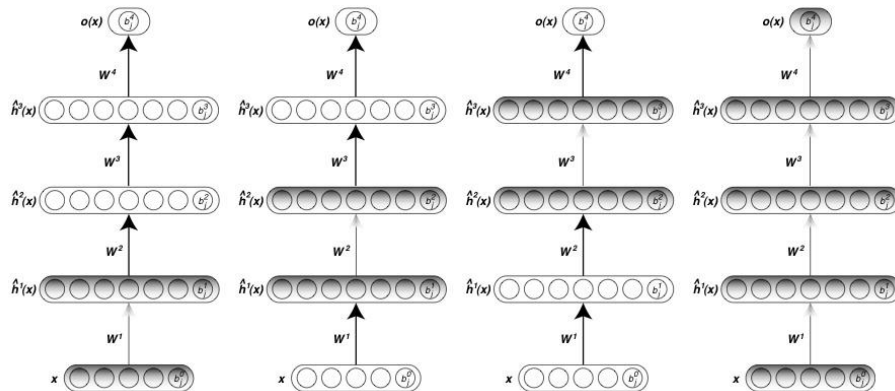


FIGURE 2.7 – Illustration du schéma de pré-apprentissage non supervisé par couches, dans un RNA avec 3 couches cachées. [Goodfellow 2016]

Cette procédure, proposée par Hinton et al. [Hinton 2006] est une technique plus générale de pré-apprentissage non supervisé par couches. Autrement dit, non seulement les RBMs peuvent être empilés pour former des réseaux profonds, mais d'autres types de réseaux d'apprentissage non supervisé peuvent également donner des résultats similaires, tels que : l'empilement d'auto-encodeurs ordinaires (SAE) [Bengio 2007a], l'empilement d'auto-encodeurs débruiteurs (SDAE) [Vincent 2010] ou l'analyse en composantes principales à noyau (ACP Kernel) [Cho 2009].

Dans ce qui suit nous allons présenter deux types d'architectures profondes qui sont : Deep Belief Networks et l'Empilement d'auto-encodeur débruiteur.

2.5.1 Deep Belief Networks

Les Deep Belief Networks (DBN) sont composés de RBMs empilées. L'apprentissage d'un DBN est effectuée en faisant l'apprentissage du premier RBM à partir

des exemples d'entrée, puis les autres RBMs du DBN sont entraînés séquentiellement. Ensuite, les probabilités d'activation de unités cachées du premier RBM sont utilisées comme des données d'entrée pour l'apprentissage du RBM de la couche suivante, et ainsi de suite.

Ce pré-apprentissage génératif est suivi ou combiné avec d'autres types d'apprentissage, typiquement discriminantes qui affinent l'ensemble de tous les poids pour améliorer les performances du réseau. Cet ajustement est effectué en ajoutant une couche finale de variables qui représentent les sorties souhaitées fournies dans les données d'apprentissage. Par la suite, l'algorithme de rétro-propagation peut être utilisé pour ajuster ou affiner les poids du réseau de la même manière que pour le réseau de neurones à propagation avant standard.

2.5.2 Empilement d'auto-encodeur débruiteur

L'empilement d'auto-encodeur débruiteur (SDAE pour Stacked Denoising AutoEncoders) est composé de DAE empilées. L'apprentissage d'un SDAE est effectué en faisant l'apprentissage du premier DAE à partir des exemples d'entrée, avec un processus debruit, puis les autres DAEs du SDAE sont entraînés séquentiellement. Ensuite, les probabilités d'activation des unités cachées du premier DAE sont utilisées comme des données d'entrée pour l'apprentissage du DAE de la couche suivante, etc.

Il est important de noter qu'après l'apprentissage non-supervisé la version des données corrompues utilisées lors de l'apprentissage du DAE, n'est plus utilisée dans la deuxième phase d'apprentissage supervisé. Dans cette phase, une couche finale qui représente les sorties du modèle est ajoutée au réseau et l'ensemble de tous les poids est ajusté pour améliorer les performances du réseau, en utilisant l'algorithme de rétro-propagation.

2.6 Apprentissage approfondi pour la prédiction des séries temporelles

Les RNAs se sont largement étudiés dans la littérature pour les problèmes de prévision des séries temporelles et des bonnes performances ont été rapportées. Traditionnellement, des architectures peu profondes ont été utilisées en raison de problèmes de convergence dans le cas de modèles profonds. Les recherches récentes ont mis l'accent sur l'apprentissage des architectures profondes des RNAs et le sollicitent comme un outil très puissant dans l'apprentissage automatique. Dans ce

2.6. Apprentissage profondi pour la prédiction des séries temporelles

qui suit nous présentons quelques travaux de prédiction des séries temporelles qui utilisent ces architectures dans différents domaines.

Les expérimentations dans [Chao 2011], indiquent que les DBNs sont applicable de prédire des taux de change. Les auteurs ont utilisé le RBM continue (Continuous Restricted Boltzman Machine -CRBM-) introduit par Chen et Murray dans [Chen 2003] au lieu du RBM. Ils ont utilisé six critères pour évaluer la performance du DBN dans la prédiction de taux de change et chaque RNA est entraîné 50 fois pour voir la stabilité du modèle DBN. Pour la sélection de l'architecture du DBN, ils ont utilisé des différents niveaux de couches cachées qui varie de 1 à 10 et le nombre de neurones de chaque couche est : 4, 8, 12, 16 et 20. Ensuite ils ont calculé la moyenne de 10 essais. Les résultats montrent que l'architecture DBN 6-16-8-1 a donné des meilleurs résultats dans la prédiction de l'échange la livre sterling / dollar américain. Ainsi que le modèle DBN dépasse la performance et la stabilité de RNA à propagation en avant dans cinq mesures même sur les données de test.

L'étude de l'apprentissage profondi sur la prédiction de la température intérieur pour réduire la consommation d'énergie liée au chauffage, de ventilation et de climatisation a été réalisée par [Romeu 2013]. Les auteurs ont utilisé différentes configurations des hyper-paramètres et deux critères d'évaluation de performance pour comparer trois types d'apprentissage : le TM-0 (un RNA sans préapprentissage), TM-1 (un SDAE avec un préapprentissage et une rétro-propagation sur toutes les couches) et TM-2 (un RNA-SDAE avec une rétro-propagation seulement dans la dernière couche). Les résultats indiquent que les modèles approfondis notamment les SDAEs donne une meilleure performance de généralisation sur l'ensemble de test et une réduction de sur-apprentissage ont été observées dans la prédiction de la température intérieur.

Une expérimentation en utilisant les données de référence CATS qui représente une série temporelle uni-variée avec 5000 points présentant cinq blocs non consécutifs de 20 points inconnus. Le modèle DBN proposé dans [Kuremoto 2014] a confirmé sa supériorité aux modèles de RNA classiques tels que le perceptron multicouche (MLP), le RNA flous auto-adaptatives (SOFNN) et le modèle ARIMA. Ce DBN se compose de deux RBM avec une structure optimisée par l'Optimisation par Essaims Particulaires (PSO) et les données ont été lissées par un prétraitement pour enlever la tendance. Cette capacité d'approximation et de prédiction du DBN ont également été confirmées en utilisant d'autres séries temporelles chaotiques telles que l'attracteur de Lorenz et la suite logistique, et cela fonctionne aussi mieux que d'autres méthodes conventionnelles bien connues.

Dans l'article de [Romeu 2015] qui a étudié la prédiction de la consommation

d'énergie électrique d'une maison individuelle. Les SDAEs montrent une capacité d'extraction des caractéristiques complexes des données de série temporelle. Par conséquent, ces modèles sont capables de mieux généraliser. La visualisation des poids et l'analyse des probabilités d'activation montrent que les modèles avec pré-apprentissage donnent une meilleure généralisation lorsqu'ils sont utilisés sans co-variables.

L'énergie éolienne est largement utilisée en tant que source d'énergie propre et renouvelable. Cependant, en raison de l'incertitude de la vitesse du vent, il est essentiel de construire un modèle de prévision précis. Q. Hu et al. [Hu 2016] ont bénéficié de la capacité du SDAEs pour extraire les caractéristique globales pour faire la prédiction de la vitesse du vent dans un parc nouvellement construit et dont il n'y a pas suffisamment de données historiques pour former un modèle précis. Ceci en utilisant un modèle de prédiction formé sur les données provenant des plus anciens parcs près de la nouvelle ferme. De cette façon, le SDAE transfère des informations d'une ferme à l'autre et les résultats expérimentaux montrent que les erreurs de prédiction sont réduites de manière significative en utilisant la technique proposée.

Hossain et al. [Hossain 2015] ont utilisé le SDAE pour la prédiction de la température de l'air à partir des données historiques de pression, d'humidité et de température recueillies par des capteurs météorologiques dans le nord-ouest du Nevada. Les résultats de l'apprentissage approfondi ont surpassé la performance d'un simple RNA sans pré-apprentissage.

Également, Liu et al. [Liu 2014] ont utilisé le SDAE pour extraire les caractéristiques des données avant l'utilisation de la méthode SVM afin de faire la prévision météorologique. Les résultats expérimentaux montrent que la technique proposée a surpassé la performance d'un modèle SVR classique et que les caractéristiques obtenues peuvent faciliter le traitement de grand volume de données météorologiques.

Un modèle hybride associant ARIMA et DBN est proposé pour la prévision des marées rouges dans [Qin 2017]. Un modèle ARIMA est construit pour chaque facteur environnemental dans différentes zones côtières pour décrire la corrélation temporelle et l'hétérogénéité spatiale. Le DBN sert à capturer la relation complexe non linéaire entre les facteurs environnementaux et la biomasse de la marée rouge, afin de réaliser l'avertissement de la marée rouge à l'avance. Les résultats expérimentaux démontrent que la méthode proposée réalise une bonne prévision de la marée rouge.

Une méthode basée sur DBN et une méthode de réduction de la dimensionnalité ont été proposées dans [Shao 2018] pour la détection des défauts des paliers à roulement à billes dans le but d'éviter les défauts inattendus pouvant conduire à des accidents catastrophiques. Le prédicteur DBN été utilisé pour modéliser le signal

2.6. Apprentissage profondi pour la prédiction des séries temporelles

de vibration complexe et ces paramètres sont optimisés avec l'algorithme génétique. Les résultats démontrent que la méthode proposée peut capturer le comportement dynamique du système de manière stable et précise, et qu'elle est plus précise et fiable pour la détection des défauts de roulement que les méthodes traditionnelles.

L. Nie et al. [Nie 2017] ont proposé une méthode pour la prédiction du trafic réseau basée sur le modèle DBN et le modèle gaussien. L'architecture DBN est utilisée pour la prédiction de la composante de dépendance à longue portée. La méthode de prédiction proposée a montré une bonne performance par rapport aux trois méthodes largement utilisées pour la prédiction du trafic réseau.

Une approche basée sur le réseau DBN atteint une grande précision avec une réduction d'environ 72% par rapport au prédicteur ARIMA traditionnelle. Dans une étude [Zhang 2017] pour prédire les demandes de ressources dans le cloud afin d'optimiser la planification des tâches et l'équilibre de charge.

[Li 2018] présente une étude sur l'utilisation de l'apprentissage en profondeur pour la modélisation d'un processus de capture du CO_2 post-combustion. Les résultats prouvent que le modèle DBN proposé est capable de prédire avec une capacité de généralisation plus élevée que le modèle RNA avec une seule couche cachées sans préapprentissage, en particulier pour les prédictions multi-étapes. Ce qui montre que l'étape d'apprentissage non supervisé peut éviter que le modèle soit piégé dans les minima locaux et surmonter le problème de sur-apprentissage.

Une approche d'apprentissage profondi basée sur l'utilisation de SAE pour la prévision des inondations dans les six prochaines heures est présentée dans [Liu 2017]. Pour améliorer la capacité de simulation, les auteurs ont classé d'abord toutes les données dans plusieurs catégories par la classification K-means. Ensuite, plusieurs modèles SAE sont utilisés pour simuler leurs catégories de données correspondantes. L'approche proposée a été comparée avec d'autres méthodes comme le SVM, RNA classique, RBF, ELM. Les résultats expérimentaux montrent que l'algorithme basé sur le SAE surpasse les autres approches en termes de d'erreurs moyennes.

Une méthode de diagnostic des pannes basée sur le SDAE est présentée dans [Lv 2017] pour apprendre les corrélations profondes des défauts et réduire la perte d'informations sur les défauts. Les résultats expérimentaux ont montré son gain de performance par rapport aux méthodes existantes, en particulier pour les failles naissantes difficiles à détecter avec les technologies traditionnelles. Ces performances ont été comparées sur une référence industrielle chimique bien connue (Tennessee Eastman) et montrent la capacité de SDAE dans l'approximation de la fonction complexe.

Une prévision à court terme du prix de l'électricité avec des algorithmes basés sur

les SDAEs est étudiée dans [Wang 2017]. Les modèles SDAEs sont comparés à des approches basées les RNAs classiques et les SVMs. Les résultats démontrent que les modèles SDAE sont capables de prévoir avec précision les prix de l'électricité en le comparant avec des résultats industriels et les paramètres de SDAE peuvent affecter ses performances, telles que le nombre de couches cachées et le taux d'apprentissage considéré.

Dans l'article [Lu 2017], une méthode de diagnostic des fautes basée sur SDAE est présentée pour améliorer la gestion de santé des systèmes d'ingénierie rotatifs. Les résultats du diagnostic des expériences comparatives montrent que la méthode d'apprentissage en profondeur proposée est capable d'extraire de manière adaptative les caractéristiques des fautes et d'identifier efficacement les états de santé avec une grande précision du diagnostic surpassant les algorithmes de diagnostic traditionnels.

Un algorithme hybride de SDAE et SVM est proposé dans [Shen 2016] pour la prévision du taux de change et comparé avec les modèles RNAs et SVMs. Les résultats de la simulation indiquent que les performances de l'algorithme intégré SAE-SVM sont bien meilleures que celles des autres benchmarks.

Un DBN composé de plusieurs couches de RBM est utilisé dans [Dedinec 2016] pour la prévision de la charge électrique à court terme sur la base des données macédoniennes de consommation d'électricité horaire. Les résultats obtenus sont comparés avec un RNA à propagation en avant sans pré-apprentissage. Les comparaisons montrent que le modèle proposé n'est pas seulement adapté à la prévision horaire de la charge d'électricité, il fournit aussi des résultats supérieurs à ceux obtenus avec les méthodes traditionnelles.

X. Qiu et al. [Qiu 2017] ont proposé une méthode d'apprentissage approfondi basée sur une décomposition en utilisant l'algorithme Hilbert–Huang et le réseau à plusieurs couches cachées (DBN). La méthode proposée a été évaluée à l'aide de trois ensembles de données sur la demande en électricité en Australie. Les résultats de la simulation ont démontré l'attrait de la méthode proposée par rapport à neuf autres méthodes de prévision.

Référence	Pays	Type d'approche utilisée	Domaine d'application
[Romeu 2013]	Espagne	SDAE	Prédiction de la température intérieur pour réduire la consommation d'énergie liée au chauffage

[Romeu 2015]	Espagne	SDAE	Prédiction de la consommation d'énergie électrique d'une maison individuelle
[Hu 2016]	Chine	SDAE	Prédiction de la vitesse du vent (énergie éolienne)
[Hossain 2015]	États-Unis	SDAE	Prédiction de la température de l'air
[Liu 2014]	Hong Kong	SDAE	Prédiction météorologique
[Wang 2017]	États-Unis	SDAE	Prédiction à court terme du prix de l'électricité
[Dedinec 2016]	Macédoine	DBN	Prédiction de la charge électrique à court terme
[Qiu 2017]	Australie	DBN	Prédiction de la charge électrique à court terme

TABLE 2.1 – Des études sur l'utilisation de l'apprentissage approfondi pour la prédiction des séries temporelles dans secteur énergétique et météorologique

Dans la majorité des travaux présentés ci-dessus, la méthode d'apprentissage non supervisé DAE ou RBM pour l'initialisation des poids d'un RNA profond a montré sa supériorité par rapport aux autres méthodes classiques en terme de performance d'approximation de la fonction de série temporelle dans plusieurs domaines, spécialement le secteur d'énergie et le secteur météorologique comme décrit dans la table 2.1 qui classe certaines études sur l'utilisation des réseaux de neurones profonds par : type d'approche utilisée, domaine d'application et pays.

2.7 Conclusion

Ce chapitre a présenté la théorie des RNAs et les motivations des recherches récentes sur les architectures profondes. Ces dernières années, l'apprentissage profond est apparu comme un outil performant dans l'apprentissage automatique dans plusieurs domaines de l'intelligence artificielle. D'autre part, nous avons présenté les recherches récentes sur l'approximation de fonctions , type séries temporelles en utilisant les réseaux de neurones profonds.

Analyse de données pour la prévision

Sommaire

3.1	Introduction	61
3.2	Données de la charge électrique Algérienne	62
3.3	Caractéristiques de la consommation électrique en Algérie	63
3.3.1	Tendance	64
3.3.2	Saisonnalité	65
3.3.3	Cycle Hebdomadaire	66
3.3.4	Cycle Journalier	66
3.4	Procédures de la prédiction de charge	68
3.4.1	Caractère cyclique de la consommation	68
3.4.2	Les données de prévision météorologique	69
3.4.3	L'activité économique	69
3.4.4	Les événements exceptionnels	70
3.5	Choix des paramètres du modèle	70
3.6	La relation entre la température et la consommation	71
3.7	Évaluation de la qualité de prédiction	73
3.8	Conclusion	75

3.1 Introduction

Tout d'abord dans ce chapitre, nous présentons les données fournies de la compagnie chargée de la production, du transport et de la distribution de l'électricité et du gaz en Algérie (SONELGAZ). Ensuite, nous allons voir en détails les échelles de temps et le type de données disponibles pour la construction d'un modèle de prédiction. Puis, nous montrons une analyse introductive de la charge électrique en Algérie, cette analyse est importante pour la conception d'un modèle de prédiction et elle nous permet d'extraire ses caractéristiques pour mieux comprendre son évolution.

3.2 Données de la charge électrique Algérienne

La population de l'Algérie est évaluée à 42,2 millions d'habitants en 2018 selon l'Office national des statistiques, une population qui est très inégalement répartie sur le territoire national ; 87% de la population concentrée sur le Nord du pays sur 11% de la superficie, principalement sur la côte.

Les grandes agglomérations sont les wilayas du Nord. Principalement, Alger, Oran, Constantine et Annaba avec des densités très fortes qui dépassent les 40 000 habitants/ km^2 . Cette répartition est principalement due au climat. En générale, le climat méditerranéen est caractérisé par des étés relativement chauds et secs et des hivers doux et humides, pendant que les climats des régions un peu plus dans le sud comme les zones montagneuses de l'Atlas Saharien et l'Atlas Tellien jusqu'aux zones du sud saharien sont caractérisée par par une saison sèche s'étendant sur la plus grande partie de l'année et une saison humide, avec de faibles précipitations. Du coup, la densité dans certaines des régions des wilayas du sud comme les wilayas de Djelfa, Adrar et Tamenrasset ne dépasse même pas les 20 habitants au km^2 .

Chargée de l'approvisionnement de toutes ces régions par de l'énergie électrique est la société SONELGAZ, un groupe industriel composé de 16 filiales, ces filiales assurent la production, le transport et la distribution de l'électricité ainsi que le transport et la distribution du gaz par canalisations. Parmi ces filiales, l'Opérateur Système électrique (OS) est chargée de la conduite du système Production / Transport de l'électricité. Notamment, elle assure l'équilibre offre-demande des unités de production.

Les données de la charge électrique nationale, de l'année 2000 à 2016 ainsi que la consommation régionale de cinq différentes régions du pays, à savoir, Oran, Alger, Sétif, Annaba et le SUD durant deux derniers années sont fournies par la compagnie nationale d'électricité Sonelgaz. Les régions d'Oran, Alger, Sétif et Annaba sont situées respectivement au nord du pays réparties de l'ouest vers l'est, comme illustré dans la figure 3.1. Le sud a une densité de population largement inférieure par rapport au nord du pays.

Il est important de bien considérer que chaque région n'a pas le même poids en terme de consommation. Le nord représente 90% de la consommation nationale d'électricité en Algérie. Plus particulièrement, la région Est représente 35% , le Centre 30%, Ouest 25%. Par contre, le Sud ne représente que 10% de la consommation.

Ces données représente la Puissance Maximale Appelée (PMA) qui définit la demande maximale de l'ensemble des consommateurs connectés au réseau. Cette

mesure est communément appelé "point de consommation" en référence au 'pic' de la courbe de charge.

Ces PMA sont échantillonnés en Quart horaire, c'est à dire la consommations chaque $\frac{1}{4}$ (un quart) d'heure. Généralement, les distributeurs d'électricité reviennent à cette mesure pour la tarification de l'électricité de haute tension afin de rémunérer l'investissement matériel consenti pour fournir à leurs clients l'énergie demandée.

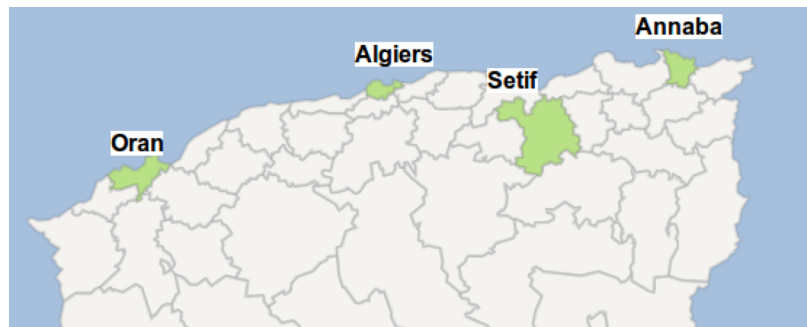


FIGURE 3.1 – Position des villes couvertes par les données de consommation régionale de charge électrique.

La production en énergie en Algérie est principalement issue des centrales électriques de la Société de Production de l'Électricité de Sonelgaz avec des équipements de type turbines à gaz et turbines à vapeurs et cycles combinés. Le gaz naturel représente le combustible principal de ces centrales qui est une ressource primaire plus propre que d'autres ressources fossiles. Sonelgaz utilise aussi d'autres types d'équipement dans ses centrales électriques mais à moindre degré que les équipements qui nécessite des ressources fossiles. Nous citons : les équipements hydrauliques, les éoliens et les photovoltaïques.

3.3 Caractéristiques de la consommation électrique en Algérie

Avant d'entamer la procédure de développement d'un modèle de la charge électrique et il est primordiale de comprendre en profondeur les caractéristiques temporelles telles que la tendance et la saisonnalité. Cette connaissance du comportement de charge est acquise à partir des observations intuitives, mais aussi par l'analyse statistique. Nous présentons dans ce qui suit quelques caractéristiques importantes et spécifiques à la consommation en Algérie.

3.3.1 Tendances

La figure 3.2 montre les charges de consommation d'électricité en Algérie au cours des années 2000 à 2016. Pour avoir un aperçu initial sur la nature des tendances de consommation, une formule polynomiale de degré quatre est ajustée aux données.

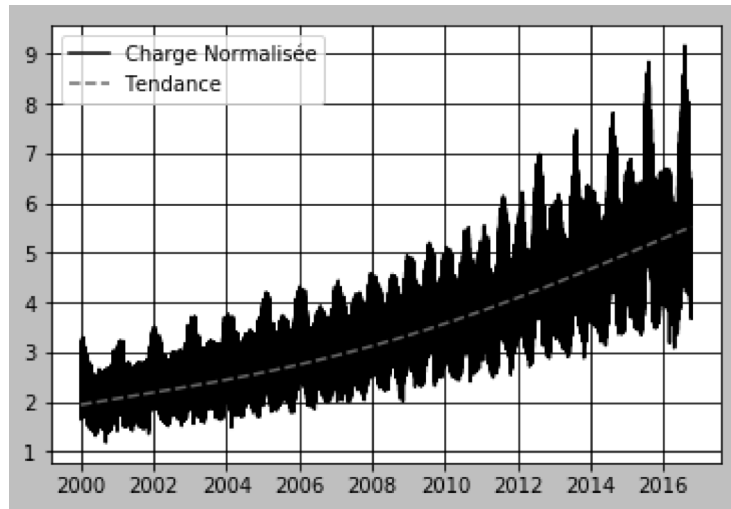


FIGURE 3.2 – Courbes de la consommation électrique en Algérie et la tendance estimée entre 2000 et 2016

La visualisation de la figure 3.2 montre que la courbe de la charge électrique algérienne sur la période de 2000 à 2016 suit une tendance croissante due à une croissance économique notable, liée à des différents facteurs économiques et sociologiques, plus particulièrement, comme illustré dans la figure 3.3, ceux de la croissance démographique et de l'évolution du produit intérieur brut (PIB) en Algérie durant cette période.

Le PIB est un indicateur économique utilisé pour mesurer la production dans un pays donné. Il définit la production économique annuelle ou la production dans un pays au cours d'une année par les agents économiques et il est généralement considéré comme la mesure du revenu provenant de la production dans un pays donné. La croissance économique en Algérie comme illustré dans la figure 3.3a peut conduire à la disponibilité des nouveaux produits qui engendrent de nouveaux besoins de consommations d'électricité. Ainsi, l'augmentation du PIB induit une croissance économique qui reflète l'amélioration du niveau de vie et le pouvoir d'achat d'une population qui a augmenté selon les derniers recensements (voir figure 3.3b). Ce dernier a un impact sur l'acquisition des divers moyens technologiques consommant

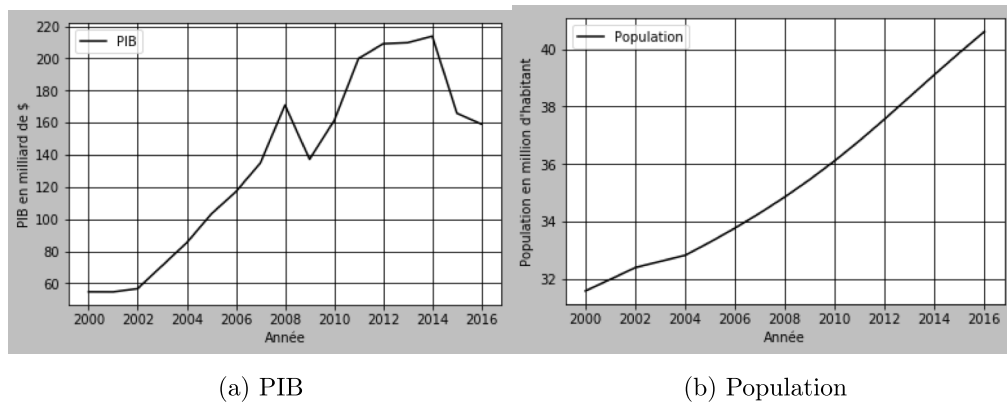


FIGURE 3.3 – Produit intérieur brut et population en Algérie entre 2000 et 2016. (Banque mondiale 2018)

ainsi plus d'énergie électrique.

3.3.2 Saisonnalité

La visualisation de la moyenne de consommation électrique mensuelle en Algérie dans la figure 3.4 montre une variation et une irrégularité d'une période à une autre pendant une même année, mais aussi une similarité entre les mêmes périodes dans des différentes années. Ceci représente une variation cyclique annuelle reflétant un changement saisonnier qui se répète chaque année, ce qui montre que le changement des conditions météorologiques influence le comportement de consommation de la charge électrique. Nous remarquons aussi que la période de consommation maximale a changé en cours de ces dernières années. La consommation estivale en 2012 et en 2016 étant très élevée par rapport à la consommation dans les mois d'hivers des mêmes années, par contre, la consommation était moins variable dans les années 2000 et 2014, où la consommation maximale était souvent atteinte en hivers et non pas en été. On peut noter aussi que le pic annuel dans les mois d'été se manifestant depuis l'an 2008 peut être expliqué par le développement économique et l'utilisation massive des dispositifs de climatisations.

Cette croissance économique et le changement de saisonnalité annuelles en terme d'amplitude de consommation rend les anciennes données moins appropriées pour suivre les tendances récentes. Ces conclusions sont très significatives pour la bonne modélisation de la charge électrique Algérienne à court terme.

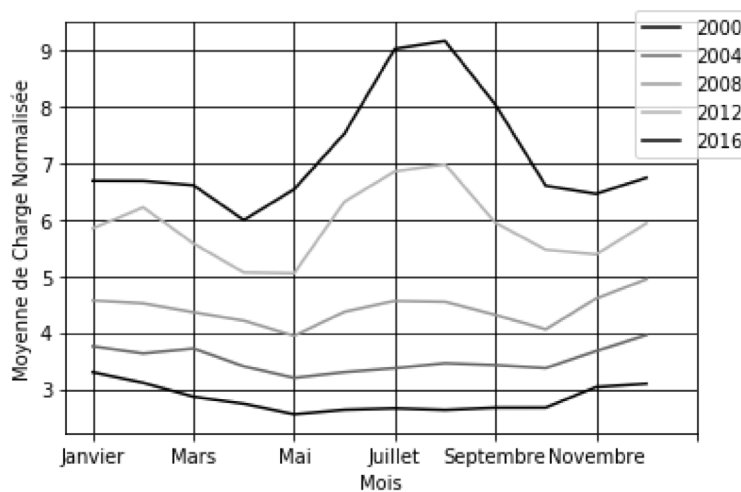


FIGURE 3.4 – La moyenne de consommation mensuelles d’électricité en Algérie de 2000 à 2016 avec un pas de 4 ans

3.3.3 Cycle Hebdomadaire

La moyenne de consommation horaires journalières dans une journées groupés par jour de semaines en 2015 illustrée dans la figure 3.5b, montre une claire différence de consommation entre les jours ouvrables et les jours non ouvrables (de week-end). Ceci est expliqué par la diminution de l’activité économique pendant les week-end et la prière hebdomadaire du Vendredi. La baisse est très apparente au vendredi par rapport aux autres jours de semaine. Notons que l’Algérie a connu un changement des journées de week-end en 2009 du Jeudi et Vendredi au Vendredi et Samedi. Ce changement a eu un impact sur la consommation d’électricité tel que visualisé dans la figure 3.5a.

La modélisation de la charge doit prendre en considération le type de jours de semaines. Comme on peut remarquer dans la figure 3.5, la consommation moyenne en 2008 est plus régulière qu’on 2015, et la consommation a la veille du week-end (Jeudi) en 2015 a une moyenne de consommation moins élevée que les autres jours ouvrables dans les heures de pointe le soir, et une moyenne de consommation similaire pendant les veilles des jours ouvrable (Samedi) après ces heures de pointes.

3.3.4 Cycle Journalier

Afin de déterminer la similarité entre les heures d’un même jour, la corrélation de consommation entre ces heures est calculée. Le résultat est présenté dans la figure

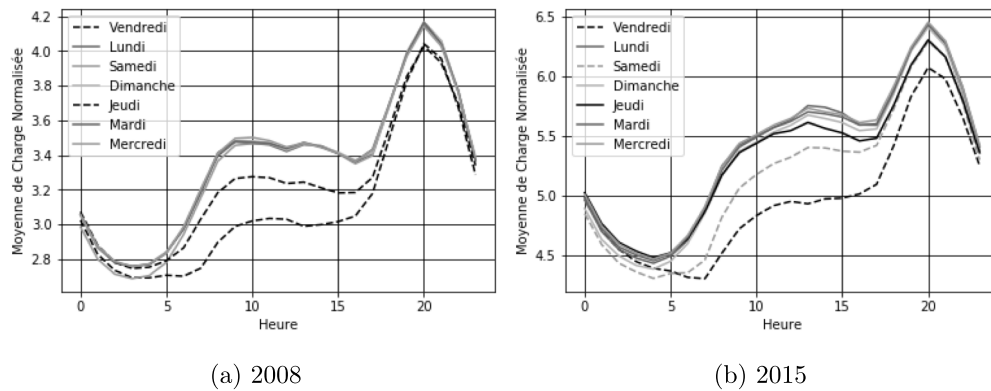


FIGURE 3.5 – La moyenne de consommation horaires par jour de semaines en Algérie en 2008 et 2015.

TABLE 3.1 – Intervalle de temps du jours des 5 modèles RNA parallèles

Numéro de modèle	Intervalle de temps du jours
Modèle RNA 1	00 - 06
Modèle RNA 2	07 - 09
Modèle RNA 3	10 - 17
Modèle RNA 4	18 - 20
Modèle RNA 5	21 - 23

3.6 où les plus fortes corrélations sont données par des zones plus sombres, tandis que les plus faibles corrélations sont données par des zones plus claires.

Cette analyse de corrélation vient de notre étude comparative des approches RNA en série et en parallèle pour prédire 168 heures d'avance de consommation [Farfar 2015]. À partir de cette analyse de corrélation, cinq groupes d'heures similaires sont définis. Cinq modèles RNA pour la méthode parallèle sont donc formés indépendamment avec les heures de données détaillées dans la Table 3.1. Regrouper des heures similaires d'un même jour implique réduire la complexité de la fonction non linéaire à approximer. Les résultats ont montrés que la prédiction en utilisant une méthode RNA en parallèle surpasse les résultats de performance obtenus en utilisant l'approche RNA en série. Ceci est expliqué par le nombre d'étape nécessaires pour la prédiction en utilisant l'approche parallèle, qui est inférieur, c'est à dire, une étape pour la prédiction d'un jour et sept étape pour la prédiction hebdomadaire.

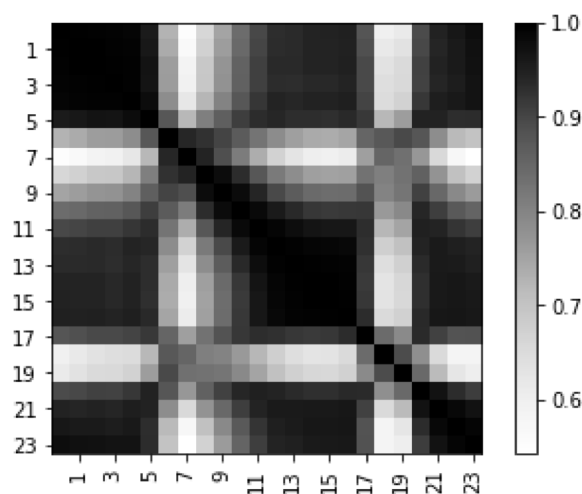


FIGURE 3.6 – Image de corrélation de la consommation d’électricité en Algérie entre 2013 et 2016 dans les heures quotidiennes

3.4 Procédures de la prédiction de charge

La prévision de la consommation Algérienne d’électricité est une des missions essentielles du département programmation de l’Opérateur du Système Électrique (OS société Spa du groupe Sonelgaz). Chaque jour, le département programmation effectue une prévision de consommation pour le lendemain. Elle est constituée de 96 valeurs de puissance, correspondant aux 24 heures d’une journée. Cette prévision sert de référence pour estimer les besoins et adapter le programme de marche du lendemain de l’ensemble des unités de production afin de satisfaire à l’équilibre offre-demande.

3.4.1 Caractère cyclique de la consommation

Comme expliqué un peu plus haut, la prédiction doit tenir compte trois cycles temporels, à savoir :

- Un cycle annuel ou saisonnier dont la consommation augmente dans les saisons d’hiver (mois de janvier et février) et d’été (mois de juillet), et diminue en saisons de printemps et d’automne, notamment.
- Un cycle hebdomadaire, c’est à dire, 5 jours ouvrés avec une consommation globalement stable et le week-end où la consommation diminue.
- Un cycle journalier où le profil de la consommation sur une journée est caractérisé par 4 points : le creux de nuit, qui correspond au minimum de la

consommation sur les 24 heures de la journée, la pointe du matin, le creux d'après-midi et la pointe du soir. Le maximum de consommation est atteint à la pointe du soir.

3.4.2 Les données de prévision météorologique

Le centre national de la météorologie fourni à l'OS.spa les prévisions météorologiques qui sont les premiers éléments influençant la consommation d'électricité. L'évolution des conditions météorologiques tout au long de l'année est pour une bonne part, à l'origine des variations cycliques annuelles de la consommation. Pour établir une prévision journalière, nous nous intéressons en particulier à deux mesures météorologiques :

- Les températures prévisionnelles régional *max*, *min*. Ces variations se traduisent par des modifications de l'utilisation du chauffage électrique en hiver et des climatiseurs en été.
- La nébulosité représentant le taux de couverture nuageuse.

Ces données météorologiques font l'objet d'un examen attentif pour l'ensemble des températures minimales et maximales pour la journée à venir, où il est important de bien considérer les prévisions météorologique, région par région. Ensuite, on recherche dans le passé des journées présentant des similitudes avec la situation actuelle.

Les références les plus courantes, pour un jour ouvrable, sont la courbe de consommation de la veille, du jour correspondant la semaine précédente, du même jour de l'année passée, etc. Ces courbes de référence sont toujours corrigées suite aux aléas climatiques : la consommation est recalculée pour des conditions météorologiques normales. Mais il arrive aussi que les références soient éloignées dans le passé ou très peu nombreuses. Dans ce cas, la prévision réalisée sera entachée d'une plus forte incertitude. Le choix de ces références est très important et conditionnera fortement la qualité de la prévision.

3.4.3 L'activité économique

L'activité économique des entreprises influence également la consommation. Ces effets sont directement observables sur les courbes de consommation hebdomadaire (voir figure 3.5). Où les jours fériés modifient profondément le profil de la consommation et influent considérablement sur la consommation nationale.

3.4.4 Les événements exceptionnels

Certains événements exceptionnels peuvent perturber le profil de la consommation, En voici quelques exemples :

- Le mois de Ramadan ou la forme de la courbe de la consommation change par rapport aux autres mois.
- Un événement sportif tel qu'une finale de coupe de football où la consommation augmente pendant la mi-temps.
- Les intempéries, en entraînant d'importantes coupures, perturbe la consommation d'électricité. L'impact de tels événements est difficile à prévoir, du fait de leur caractère exceptionnel.

3.5 Choix des paramètres du modèle

La charge future sur un système électrique est prédite en extrapolant une relation prédéterminée entre la charge et ses variables influentes, à savoir, le temps et les conditions météorologiques. Ceci nécessite l'identification et la quantification de cette relation.

Les méthodes de prédiction quantitatives impliquent l'analyse des données historiques de la variable d'intérêt. Il y a deux types de ces modèles de prévision, les modèles univariés et les modèles multivariés.

Les modèles univariés prédisent les valeurs futures d'une série temporelle uniquement sur la base de ces valeurs passées. Les données historiques de la charge électrique dépendent fortement des variables calendaires comme le mois de l'année et le jour de la semaine comme rapporté dans la section 3.3.

Les expérimentations que nous avons fait sur la prédiction de la charge électrique Algérienne montre que l'utilisation des valeurs récentes de l'heure à prédire, les valeurs du jour passé et la semaine passé à la même heure, aide clairement dans les prédictions à court terme. Par exemple, la figure 3.7 montre l'erreur de prédiction des expérimentations par rapport au nombre d'entrées utilisées dans un modèle univarié. On peut voir que l'inclusion de la valeur de charge du jour précédant et de la semaine précédente permet de diminuer considérablement le taux d'erreur dans la prédiction. Par contre, les autres valeurs qui sont entre 30 heures et environ 160 heures présente une performance stable et n'ont pas une grande influence sur les résultats de prédiction.

L'utilisation des modèles multivariés de prédiction implique l'identification de d'autres variables liées à la variable d'intérêt. Pour le cas de la modélisation de la

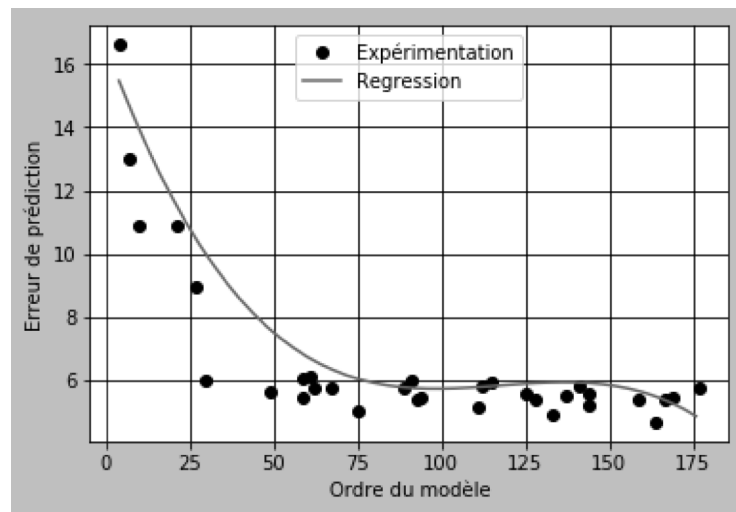


FIGURE 3.7 – Erreur de prédiction par rapport au nombre d'entrées (heure précédent).

charge, les paramètres météorologiques comme la température, la vitesse du vent, la nébulosité et l'humidité affectent le comportement des clients des systèmes électrique, dont la température représente un facteur principal dans la majorité des cas. Cependant, les situations anormales telles que les orages peuvent effectuer de manière significative ou catastrophique la charge dans ces systèmes. Ces conditions météorologiques anormales sont difficile à prédire et ont une grande incertitude, c'est pour cela que généralement elle ne sont pas prises en compte par les modèles de prévision.

3.6 La relation entre la température et la consommation

Le nuage de points qui permet de mettre en évidence le degré de corrélation entre la pointe de consommation journalière d'électricité et la moyenne de la température maximum de la région du nord en Algérie entre 2013 et 2016 est illustré dans la figure 3.8. On peut voir clairement que la consommation d'électricité atteint ses niveaux supérieur lorsque la température est élevé. La consommation baisse pour une atmosphère douce et connaît une augmentation dans les climats froids.

Pour ceci, la température est généralement considéré comme l'information la plus importante pour la prédiction de charge à court terme parmi les autres variables météorologiques. Les chercheurs introduit souvent dans leurs modèles d'approximation un certain d'autre variables afin d'avoir plus de précision, comme l'humidité, la cou-

verture nuageuse et la vitesse du vent. Mais cette inclusions de paramètres nécessite la disponibilité des données de ces paramètres, relativement ce n'est pas toujours le cas.

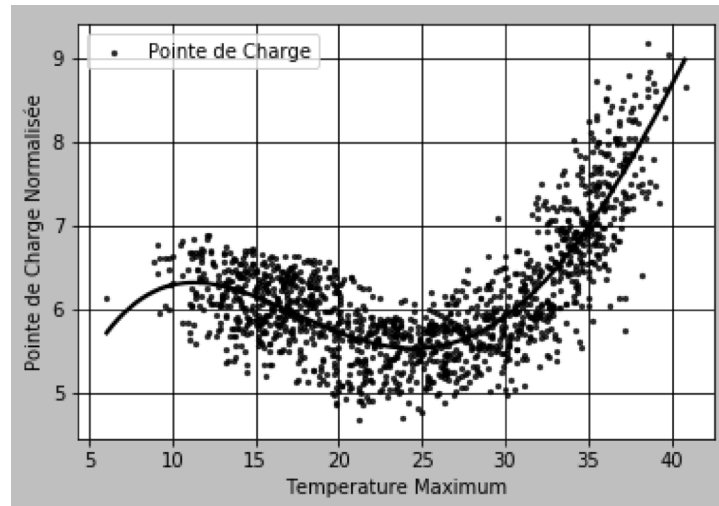


FIGURE 3.8 – Pointe de consommation journalière d'électricité en Algérie entre 2013 et 2016 par rapport à la température maximum

L'humidité relative est la quantité qui varie de 0 à 100% qui représente la vapeur d'eau contenue dans un volume d'air donné par rapport au maximum qu'il pourrait contenir à une température et une pression données. On dit l'air est humide à plus de 65% d'humidité relative, l'air est moyennement humide entre 35 et 65%, et l'air est sec lorsque l'humidité est inférieure à 35% [Naud 1995]. En résumé, l'humidité relative peut devenir importante lorsque la température est élevée en été et que la vitesse du vent peut devenir non négligeable lorsque la température est basse en hiver.

Un indice de confort humain basé sur la température mesurée, du vent et de l'humidité relative sera un bon paramètre pour la modélisation, parmi ces indices on trouve la température ressentie. Cet indice permet aux autorités et aux personnes de prendre des mesures précoces pour prévenir les effets des météos extrêmes.

Pour calculer le froid ressentie ou le chaud ressentie on utilise généralement les indices mis au point historiquement au Canada et aux Etats-unis. Ces indices sont basés sur des formules empiriques qui permettent d'évaluer la sensation d'une personne se trouvant dans un champ ouvert. Dans la saison hivernale et la saison estivale on peut évaluer la température ressentie comme suit :

Saison hivernale

La formule qui permet de calculer la température ressentie en fonction de la température ambiante et de la vitesse du vent, nommé l'indice du refroidissement éolien est la suivante [Osczevski 2005] :

$$T_R = 13,12 + 0,6215 T_a - 11,37 V^{+0.16} + 0,3965 T_a V^{+0.16} \quad (3.1)$$

Où :

- V : La vitesse du vent en km/h ;
- T_a : La température ambiante en degrés Celsius.
- T_R : La température ressentie équivalente en degrés Celsius.

Saison estivale

La formule calcule la température ressentie en fonction de la température ambiante, de pression de vapeur saturante de l'eau et de la vitesse du vent, elle inclut aussi le taux important d'humidité [Steadman 1994] :

$$T_R = T_a + 0,33 e - 0,70 V - 4,00 \quad (3.2)$$

Où :

- T_a : Température ambiante (°C).
- e : Pression de vapeur saturante de l'eau (hPa).
- V : La vitesse du vent en (m/s).

La pression de vapeur saturante de l'eau peut être calculée de l'humidité relative et de la température en utilisant l'équation :

$$e = \frac{RH}{100} \cdot 6,105 \cdot \exp\left(\frac{17,27 \cdot T_a}{237,7 + T_a}\right) \quad (3.3)$$

- T_a est la température ambiante (°C).
- RH est l'humidité relative.
- exp représente la fonction exponentielle.

3.7 Évaluation de la qualité de prédiction

Malheureusement, toutes les situations de prédiction impliquent un certain degré d'incertitude rend l'erreur inévitable. Le choix de cette mesure d'exactitude est très important pour l'évaluation de la méthode proposé. Ainsi, ces mesures d'erreurs doivent être faciles à comprendre et liées au besoins des décideurs du système

électrique où les grandes erreurs peuvent avoir des conséquences désastreuses. Dans les tâches de prédiction, l'erreur de prévision est le décalage entre une prévision à l'instant t et sa valeur réel, et afin d'éviter les erreurs positives et négatives, les mesures basées sur l'erreur absolue sont largement utilisé dans littérature, c'est-à-dire des valeurs absolues de différence entre les valeurs réel A_t et prédites F_t . Par exemple, la somme des erreurs absolues ou l'erreur absolue moyenne (MAE pour Mean Absolute Error) défini comme suit :

$$MAE = \frac{1}{m} \sum_{i=1}^m |A_t - F_t| \quad (3.4)$$

La majorité des travaux sur la prédiction de la charge électrique à court terme utilise une autre mesure d'exactitude nommé l'erreur absolue moyenne en pourcentage (MAPE pour Mean Absolute Percentage Error), qui exprime l'erreur en terme de pourcentage. MAPE est mesure préféré par les experts du domaines vu que parfois la différence relative ou en pourcentage est plus informative que l'erreur absolu lorsque ils comparent des résultats pour deux ensembles de données distincts. Cette mesure prend la forme :

$$MAPE = \frac{1}{m} \sum_{i=1}^m \frac{|A_t - F_t|}{A_t} \quad (3.5)$$

Le choix de mesures d'erreur pour aider à comparer les méthodes de prédiction dans le secteur d'approvisionnement en électricité a été beaucoup discuté à la suite des nombreuses compétitions qui ont débuté dans les années 1980. Certains articles rapportent que les services publics préfèrent évaluer les systèmes de prévision par les erreurs absolues produites ce qui suggère que les MAEs pourraient être utiles. Parfois aussi l'erreur quadratique est suggérée car elle pénalisent les erreurs importantes [Hippert 2001].

La MAPE apparaît adéquate dans la prévision de charge puisque les valeurs de charge réelles sont plutôt importantes, malgré qu'elle n'est pas définie lorsque la valeur réelle est zéro ($A_t = 0$), ce qui est rare d'apparaître dans les données de consommation. Mais, il faut noter que les mesures d'erreurs utilisés dans la littérature varient beaucoup vu qu'il n'y a pas une standard industrielle. Par conséquent, les résultats de prévisions d'un article à l'autre n'est aussi claire, même si les données utilisés représentent le même ensembles.

Parmi les mesures d'exactitude suggérées dans la littérature on peut cité l'erreur quadratique moyenne (RMSE pour Root Mean Square Error) qui permet d'agréger les amplitudes des erreurs car elle est sensible aux valeurs aberrantes, où l'effet de

chaque erreur est proportionnel à la taille de l'erreur quadratique et les erreurs plus importantes ont un effet disproportionné [Pontius 2008]. RMSE est défini par :

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (A_t - F_t)^2}{m}} \quad (3.6)$$

3.8 Conclusion

Tout au long de ce chapitre, nous avons présenté les caractéristiques de la charge électrique en Algérie au fil des dernière années. Cette charge suit une tendance croissante et consommation cyclique. Dans le chapitre suivant, nous allons présenter notre contribution effectués pour la prédiction de la charge électrique pour l'horizon court terme.

Contribution

Sommaire

4.1	Introduction	77
4.2	Méthode proposée	80
4.2.1	Pré-traitement	80
4.2.2	Clustering	81
4.2.3	Estimation des profils de température	82
4.2.4	Normalisation des données	83
4.2.5	Stacked Denoising Auto-encoders	84
4.3	Expérimentations	85
4.4	Résultats	87
4.5	PREVOS-DZ	95
4.5.1	Prévision de la charge électrique	97
4.5.2	Technologies utilisées	97
4.5.3	Fonctionnalités de PREVOS-DZ	98
4.6	Conclusion	102

4.1 Introduction

Le système d'énergie électrique est une infrastructure critique et complexe, à partir de la génération et la transmission de l'énergie aux circuits de distribution. La principale préoccupation de ce système est de fournir en permanence aux clients une énergie électrique fiable et de bonne qualité. En raison du manque de capacité de stockage d'énergie, au moins pour l'instant, les compagnies d'électricité doivent prévoir la consommation de charge pour assurer un bon équilibre de charge entre l'offre et la demande instantanément, tout le temps. Par conséquent, les scénarios de prévision sont réalisés sur différents horizons temporels : court terme (d'une heure à une semaine), moyen terme (une semaine à plusieurs mois ou quelques années) et long terme (plus d'un an à l'avance), chacun étant essentiel pour différentes tâches

d'aspect de planification, allant des opérations du système d'énergie électrique à la sécurité ainsi que la gestion économique de l'approvisionnement en énergie.

La prévision de charge électrique à court terme (STLF pour Short Term Load Forecasting) implique une planification économique de la capacité des générateurs pour minimiser les coûts de fonctionnement quotidiens, affectant la planification de la transmission en répartissant la charge produite entre les clients et les opérations du système en temps réel afin d'assurer que le système continue de fonctionner de manière fiable et sécurisée.

Plusieurs méthodes et techniques ont été appliquées aux données de différents pays pour effectuer des prévisions de charge électrique. Ces techniques sont généralement divisées en deux catégories principales :

D'une part, les approches traditionnelles qui incluent une certaine variation des modèles de séries temporelles classiques comme les méthodes de régression [Ramanathan 1997], le modèle autorégressive et moyenne mobile (ARMA) comme dans [Huang 2003] où ils ont inclus le concept de cumulatif et bispectre pour valider la Gaussianité des données de charge. S. Sargunaraaj et al. [Sargunaraaj 1997] a proposé une approche en deux étapes basée sur le filtrage de Kalman [Zarchan 2005] où la moyenne est prédite en premier, puis le filtre est appliqué en temps réel en utilisant un retour d'erreur des heures précédentes. Une discussion de ces approches peut être trouvée dans [Kyriakides 2007].

D'un autre côté, des techniques basées sur l'intelligence artificielle sont appliquées. On peut citer les réseaux neuronaux artificiels (RNA) [Haykin 2009] utilisés dans de nombreuses expériences réussies et des tests pratiques [Hippert 2001]. Machine à vecteurs de support [Vapnik 1995] utilisé dans [Yuancheng 2003] pour interpoler les données de charge et de température. D'autres chercheurs ont appliqué des ondelettes [Shensa 1992] pour extraire des informations redondantes de la courbe de charge [Bashir 2009], tandis que d'autres ont essayé des techniques de logique floue pour gérer des situations anormales dans la demande de charge [Mamlook 2009]. Le modèle de réseau neuronal à carte auto adaptatives est également été utilisé dans le but d'obtenir la courbe de charge de certaines heures du jour suivant [Senabre 2010]. D'autres chercheurs ont utilisé des systèmes experts [Gonzalez 1993] comme proposé dans [Rahman 1988] où les règles sont développées par l'examen des relations logiques et syntaxiques entre les conditions météorologiques, la charge et les formes de charges journalières habituelles. Diverses approches hybrides de ces techniques ont également été appliquées pour augmenter l'efficacité du modèle [Fan 2006, Amjady 2009, El Desouky 2000].

Un grand nombre d'études montrent des résultats admirables en utilisant les

techniques mentionnées ci-dessus comme noté dans les revues de la littérature [Tzafestas 2001, Alfares 2002, Feinberg 2005]. Pour améliorer la précision des prévisions, ils regroupent souvent les données d'entrée par des jours similaires [Chen 2010], et également par des facteurs influençant comme le type de jour et les week-ends [Çevik 2015] où les vacances sont généralement traitées comme week-ends [Fan 2012] ou traiter séparément dans [Song 2005].

Récemment, les réseaux neuronaux profonds ont montré des promesses pour la modélisation des données statiques à travers les domaines de classification des patterns, et leur application à des données de série temporelles attire de plus en plus l'attention [Långkvist 2014]. Puisque les Machines de Boltzmann restreintes étaient la première méthode utilisée dans la phase de préapprentissage de la méthode d'apprentissage Greedy Layer-Wise Training [Bengio 2007a], la méthode SDAE (Stacked Denoising Autoencoders) [Vincent 2010] a été peu utilisée pour les problèmes de prédiction de séries temporelles, dont les secteurs de l'énergie et de la météorologie sont les bénéficiaires de la majorité des expérimentations telles que la prédiction de la température interne dans [Romeu 2013] et [Romeu 2015].

En plus de ces méthodes et techniques, certaines approches traitent le problème comme une prévision univariée, où la charge est modélisée comme une fonction en utilisant seulement ses données historiques. D'autres, le considèrent comme un problème de prévision multivarié, utilisant des variables supplémentaires comme facteurs exogènes, essentiellement la température, l'humidité, le type de jour et les variables sociales. Dans la plupart des modèles de prévision horaires utilisent des données météorologiques, chaque valeur de charge prévue est estimée en fonction de la température à cette heure. Cependant, les services météorologiques ne proposent pas ces profils et fournissent seulement la prévision de la température comme valeur minimum et maximum de ce jour-là. De plus, l'influence des erreurs de prévision météorologique peut aller de 17% à 60% de l'imprécision sur la prévision de charge [Fay 2010].

Une interpolation linéaire ou spline cubique entre les valeurs de température extrêmes pour estimer la température pendant les heures quotidiennes pourrait être utilisée comme dans [Taylor 2000]. En cas de manque de mesures horaires, l'interpolation linéaire ne représente pas les chutes brusques de température après le coucher du soleil et l'augmentation de la température après le lever du soleil, montrant des pics anormaux. Et la méthode cubique montre des plateaux atypiques lorsque les températures minimales et maximales sont atteintes. Hippert et Pedreira [Hippert 2004] ont proposé un modèle de réseau de neurones à sorties multiples, entraîné à une série d'observation horaires de la température sèche de l'air, où la mé-

thode neuronale surpassait les méthodes traditionnelles. Cependant, cela nécessite la disponibilité des profils historiques.

Dans cette thèse, nous proposons une méthode d'estimation des profils de température principalement établie par les bio-météorologues, basée sur le calcul de la durée du jour et des températures minimales et maximales prévues, surmontant l'indisponibilité des profils quotidiens mesurés. De plus, les courbes de charge quotidiennes de la charge électrique Algérienne sont regroupées à l'aide de l'algorithme de clustering K-Means pour créer des classes de profils de charge. Ensuite, nous proposons une architecture profonde basée sur des autoencodeurs débruiteurs empilés pour les différents modèles de charge afin de prévoir la charge électrique à court terme, bénéficiant des avancées récentes dans les architectures de réseaux neuronaux profonds. Les résultats obtenus sont évalués et comparés avec ceux obtenus par une régression linéaire multiple (MLR) et des modèles de réseaux de neurones artificiels traditionnels.

4.2 Méthode proposée

Le but de la présente étude est de construire un modèle STL_F afin d'aider les opérateurs de dispatching à réaliser des prévisions effectives, ce qui permettrait d'optimiser la production et la consommation d'énergie. L'approche proposée repose sur deux étapes principales. La première est une combinaison d'une estimation de la température horaire, étant la principale variable exogène, ainsi que d'un regroupement des courbes de charge quotidiennes. Et la seconde est une régression en utilisant un réseaux de neurones artificiels est effectuée pour prévoir la charge électrique à court terme pour chaque type de jour indépendamment, avec des poids initiaux obtenus à partir des autoencodeurs débruiteurs empilés, comme illustré dans la figure 4.1, les étapes suivantes peuvent être distingué :

4.2.1 Pré-traitement

Afin d'analyser séparément l'ampleur de la variation de la consommation entre les heures journalières, un prétraitement des données est effectué en supprimant la tendance qui caractérise la charge électrique algérienne. Cette tendance reflète l'expansion économique que les Algériens ont connue ces dernières années [Khadir 2010]. Pour ce faire, nous avons appliqué le premier opérateur de différence à la série temporelle x , où la valeur au temps t est la différence entre $x(t+1)$ et $x(t)$, obtenant une nouvelle série x' dont les valeurs sont normalisées autour de 0.

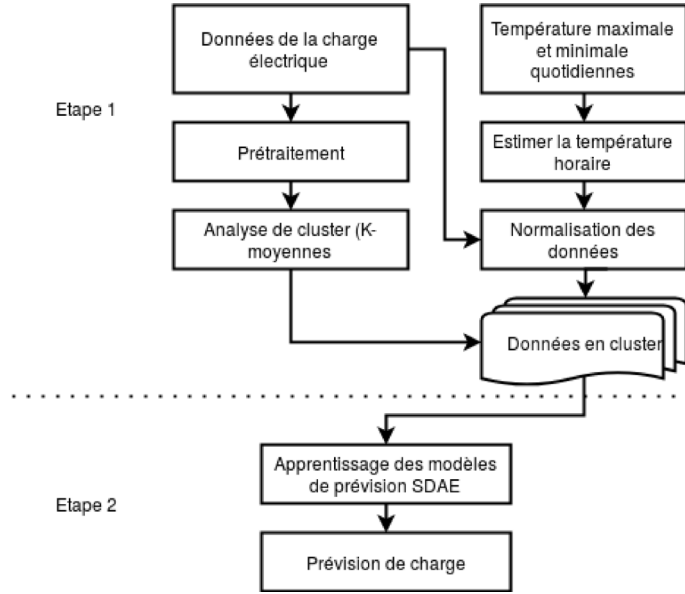


FIGURE 4.1 – Organigramme de l'approche.

4.2.2 Clustering

L'objectif principal de cette étape est de trouver le regroupement intrinsèque des profils de charge quotidiens. Pour ce faire, nous avons appliqué le clustering K-means qui est basé sur le centroïde pour déterminer cette structure de type jour en utilisant l'algorithme de Lloyd [Lloyd 1982]. L'approche consiste en un algorithme itératif avec l'objectif J de minimiser la somme des distances au carré entre chaque point de données $x^{(i)}$ à son centroïde de cluster $\mu_{c^{(i)}}$ comme exprimé dans l'équation (4.1).

$$J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K) = \sum_{i=1}^m d^2(x^{(i)} - \mu_{c^{(i)}}) \quad (4.1)$$

Avec K indiquant le nombre de clusters devant être spécifié à l'avance, μ_k est le centroïde du cluster de k où $k \in \{1, \dots, K\}$, $c^{(i)}$ est l'indice du cluster $(1, 2, \dots, K)$ auquel est affecté l'exemple $x^{(i)}$. m est le nombre d'observations de données, et $d^2(\dots)$ est le carré de la distance euclidienne entre une observation de données $x^{(i)}$ et son centroïde de cluster correspondant $\mu_{c^{(i)}}$.

Après l'initialisation aléatoire des centroïdes de cluster, la boucle interne de l'algorithme itère sur les deux étapes suivantes jusqu'à la convergence :

1. Chaque point de données x est affecté à son centroïde de cluster le plus

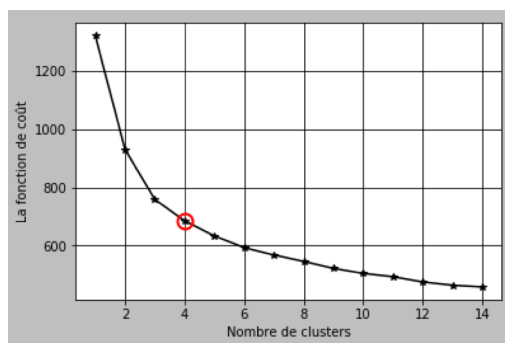


FIGURE 4.2 – La méthode du coude pour Clustering Kmeans

proche :

$$c^{(i)} = \arg \min_k d^2(x^{(i)} - \mu_k) \quad (4.2)$$

2. Mettre à jour chaque centroïde à la moyenne des points qui lui sont assignés comme suit :

$$\mu_k = \frac{1}{m_k} \sum_{x \in C_k} x \quad (4.3)$$

Où C_k est le cluster k^{th} et m_k est le nombre d'observations de données dans ce cluster.

Pour sélectionner un nombre approprié de clusters, nous avons effectué un clustering pour un nombre différent de clusters à la lumière des travaux précédents sur les types de jours de la charge électrique Algérienne [Khadir 2010, Benabbas 2011] et examiné la fonction de coût dans le cluster. De plus, nous avons effectivement essayé la méthode du coude comme illustré dans la figure 4.2 pour mieux choisir la valeur K de K-means et l'augmentation du nombre de clusters ne donne pas un changement notable de la présence du type de jour dans les données. Les regroupements générés par l'algorithme de clustering sont utilisés pour modéliser chaque type de jour indépendamment pendant le processus de régression dans la deuxième étape de cette approche.

4.2.3 Estimation des profils de température

Le cycle diurne de la climatologie exprime un phénomène qui résulte d'une rotation complète de la Terre par rapport au Soleil. Où, la température oscille entre la température maximale et la température minimale. Cet aspect de la variation de température diurne peut varier considérablement d'un jour à l'autre, car un certain

nombre de conditions peuvent élargir et réduire cette plage, telles que : longueur du jour, nébulosité, humidité, vitesse du vent et hauteur au-dessus du niveau moyen de la mer.

À cet étape, les courbes de température journalières avec une résolution horaire sont estimées en utilisant les températures minimales quotidiennes T_{min} et T_{max} températures maximales fournies par le bureau météorologique Algérien.

Cette estimation est faite sur les données des quatre région du nord qui sont : Oran, Alger, Setif et Annaba. Sachant que le nord de l'Algérie représente 90% de la consommation national. Pour ce faire, nous avons utilisé la courbe de température idéalisée selon les équations de Linvill [Linvill 1990] qui sont basées sur l'hypothèse que la température quotidienne maximale est à 2 heures après midi solaire et que le cycle de température après lever du soleil suit une courbe sinusoïdale :

$$T_{ih} = (T_{maxi} - T_{mini}) \cdot \sin\left(\frac{\pi \cdot t_a}{DL_i + 4}\right) + T_{mini} \quad (4.4)$$

Où T_{ih} est la température à l'heure h et T_{maxi} et T_{mini} sont les températures maximale et minimale au jour i , DL_i est la durée du jour (en heures) du lever au coucher du soleil, et t_a est le temps en heures (≥ 0) compté à partir du lever du soleil. Cependant, pour le moment après le coucher du soleil, la température est supposée suivre un déclin logarithmique :

$$T_{ih} = T_s - \frac{T_s - T_{mini+1}}{\ln(24 - DL_i)} \cdot \ln(t_b) \quad (4.5)$$

T_s est la température du coucher du soleil obtenue à partir du cas après le lever du soleil comme déterminé par l'équation (4.4), et T_{mini+1} est la température minimale du jour suivant $i + 1$, qui est remplacé par la température minimale du même jour pour les heures avant le lever du soleil. L'heure t_b indique l'heure en heures après le coucher du soleil plus une heure ($t_b \geq 1$). Les équations de Linvill nécessitaient des entrées telles que le coucher et le lever du soleil ainsi que la durée du jour, ces paramètres étaient calculés après les procédures [Spencer 1971] et [Almorox 2005] basées sur la latitude géographique comme expliqué dans l'annexe A.

4.2.4 Normalisation des données

Eviter la saturation de la fonction d'activation, de sorte que la fonction d'activation sigmoïde [Han 1995] (équation 4.9) est essentiellement activée dans sa région linéaire, nécessite une coordination entre la normalisation et l'initialisation des poids. L'utilisation d'une approche appropriée pour normaliser les données de séries

temporelles est essentielle avant d'alimenter les modèles de réseaux neuronaux, en particulier dans des systèmes évolutifs comme la prévision de charge.

Dans cette étude, nous avons utilisé des informations spécifiques au domaine, où la valeur maximale est sélectionnée sur la base des travaux antérieurs sur la prédiction à moyen terme [Nezzar 2016]. Par conséquent, une valeur de X est normalisée à X' en calculant :

$$X' = \frac{X}{X_{mt}} \quad (4.6)$$

où X' est la nouvelle valeur, X est l'ancienne valeur et X_{mt} est la plus grande valeur basée sur la prédiction à moyen terme.

4.2.5 Stacked Denoising Auto-encoders

Habituellement, les réseaux de neurones multicouches avec des valeurs de poids de connexion initiales aléatoires semblent se bloquer à l'optimum local [Bengio 2007a]. Hinton et al. [Hinton 2006] a présenté une approche de pré-apprentissage par couche comme stratégie pour réduire cette sensibilité initiale des poids. Où chaque couche est traînée pour abstraire la distribution des entrées produites par la couche précédente et ces poids initiaux appropriés obtenus en formant le réseau couche par couche de manière non supervisée.

Un Auto-Encoder (AE) [Larochelle 2007] est un type de réseau neurone avec la particularité d'avoir le même nombre de nœuds dans les couches d'entrée et de sortie, dans le but de la reconstruction de ses entrées au lieu de certaines valeurs cibles. Lorsque son activation de couche cachée calcule une version codée des entrées g , cette transformation différente des données d'entrée z est ensuite recopiée avec un décodeur pour une reconstruction des entrées à g' comme suit :

$$z = \sigma(Wg + b) \quad (4.7)$$

$$g' = \sigma(W'z + b') \quad (4.8)$$

où b, b' est le biais, W, W' sont les poids, et σ est la fonction d'activation sigmoïdale :

$$\text{sigmoid}(v) = \frac{1}{1 + e^{-v}} \quad (4.9)$$

L'empilement de plusieurs couches d'AE est utile pour obtenir une erreur de reconstruction plus faible et l'introduction d'un critère de débruitage sur les données

d'entrée de la tâche non supervisée aide à obtenir des représentations de niveau supérieur [Vincent 2010]. De cette manière, le DAE (Denoising Auto-Encoder) est entraîné à reconstruire une version claire de ses entrées. Par conséquent, à cette phase, chaque couche cachée DAE de l'approche proposée calcule une version codée de son entrée, où un correcteur gaussien est appliqué aux entrées en ajoutant un bruit blanc gaussien de moyenne nulle comme suit :

$$\dot{x} = x + b\mathcal{N}(0, 1) \quad (4.10)$$

\dot{x} étant l'entrée corrompue, \mathcal{N} indiquant une distribution normale et $b > 0$ le facteur de bruit. Chaque DAE est entraîné en utilisant l'algorithme de réduction de gradient Mini-batch [Bengio 2012] afin de minimiser l'erreur quadratique moyenne (MSE) exprimée dans l'équation (4.11) entre l'entrée propre x et sa version reconstruite x' , étant $x_{i,j}$ le composant i du motif j , N est la taille de Mini-batch et M est le nombre de composants (entrées).

$$MSE = \sum_{j=1}^N \sum_{i=1}^M (x_{i,j} - x'_{i,j})^2 \quad (4.11)$$

Les poids obtenus dans ces étapes sont utilisés pour initialiser le réseau de neurones à propagation avant de chaque modèle. Ensuite, chaque réseau de neurones est entraîné de manière supervisée dans l'étape finale vers la valeur cible, qui représente la prévision de charge horaire suivante, où la fonction d'activation sigmoïde est utilisée dans la couche de sortie, qui est ensuite dénormalisée par une transformation linéaire.

4.3 Expérimentations

Dans cette section, l'approche proposée est appliquée aux données qui représentent la charge électrique algérienne horaire pour une période de quatre ans (2013-2016). Les données de charge précédentes sont omises de l'ensemble d'apprentissage car un changement de comportement de la charge électrique est remarqué à partir de 2013 avec l'adoption généralisée des machines de conditionnement d'air. Afin de valider et d'évaluer l'amélioration de la performance donnée par les modèles RNA et SDAE proposés, une technique traditionnelle d'analyse des séries temporelles a été introduite sous la forme d'une régression linéaire multiple.

En tant qu'entrées pour chaque modèle, certaines valeurs de charge retardée sont choisies sur la base d'une analyse d'autocorrélation et d'une variable exogène non

TABLE 4.1 – Résumé des entrées et sorties

Entrées				Sorties
Charge électrique	Jour de la semaine	Température	Type de jour	L_t
L_{t-1} , L_{t-2} , L_{t-3} , L_{t-24} , L_{t-168}	Encodage one-hot	T_{min} , T_{max} , T_{estim}	$\begin{cases} 1 & \text{si jour spécial} \\ 0 & \text{sinon} \end{cases}$	

linéaire, c'est-à-dire que l'estimation de température est sélectionnée. Par conséquent, la charge électrique L à l'heure t est expliquée par plusieurs valeurs passées sélectionnées, en plus des estimations de température : L_{t-1} , L_{t-2} , L_{t-3} , L_{t-24} et L_{t-168} représentant respectivement les trois charges horaires précédentes, la charge de dernier jour et la charge de la dernière semaine à la même heure du jour. De plus, le contexte du type de jour est représenté en utilisant l'encodage one-hot pour chaque jour de la semaine ainsi qu'une entrée supplémentaire pour les jours spéciaux, reprenant à un total de 16 variables d'entrée comme décrit dans le tableau 4.1.

Après répartition des données sur les clusters obtenus dans la première étape, les données sont ensuite séparées en trois partitions : l'ensemble d'apprentissage (60 %) pour la création du modèle, l'ensemble de validation (20 %) pour l'optimisation des hyper-paramètres et éviter le sur-apprentissage, et le dernier ensemble de données pour les tests (20 %), afin d'évaluer la qualité du modèle. Où dans nos expériences, nous avons utilisé l'erreur absolue moyenne (MAE) et l'erreur moyenne absolue de pourcentage (MAPE) [Hyndman 2006] pour obtenir une évaluation complète de la performance de notre approche proposée. Les formules correspondantes sont données dans l'équation (4.12) et (4.13), soit A la valeur de chargement réelle, F représente la valeur prédire et m est le nombre d'échantillons.

$$MAE = \frac{1}{m} \sum_{t=1}^m |A_t - F_t| \quad (4.12)$$

$$MAPE = \frac{100}{m} \sum_{t=1}^m \left| \frac{A_t - F_t}{A_t} \right| \quad (4.13)$$

L'application de réseaux de neurones, sur les données créées, nécessite le réglage de différents paramètres, donc de décider du nombre d'unités du SDAE et le RNA sans pré-apprentissage et des valeurs appropriées de taux d'apprentissage, la

TABLE 4.2 – Paramètres utilisés dans différentes configurations expérimentales de SDAE et RNA séparément. Le symbole U signifie uniformément échantillonné.

Paramètre	Intervalle	Commentaire
Nombre de nerones cachés	$U(5, 40)$	Le nombre de neurones cachés est non fixé sur les couches cachées.
Le pas d'apprentissage pour le pré-apprentissage non supervisé de SDAE	$U(10^{-3}, 0.3)$	
Le pas d'apprentissage pour la phase supervisé	$U(10^{-3}, 0.3)$	

recherche aléatoire est utilisée, car elle est connue pour être plus efficace dans la recherche d'un espace de configuration plus grand et moins prometteur que la recherche de grille [Bergstra 2012]. La tableau 4.2 montre différentes combinaisons de paramètres que nous avons utilisées dans une configuration expérimentale différente pour SDAE et RNA.

Pour mesurer la capacité de DAE en tant que méthode pour l'extraction de caractéristiques et son effet sur notre approche, le modèle sans pré-apprentissage est utilisé comme référence pour être comparé à l'approche proposée. 30 expérimentations ont été effectuées pour chaque méthode séparément. Le nombre de couches cachées varie de 1 à 2 et le nombre d'époques pour la phase de réglage supervisé est fixé à 200, alors que pour le pré-apprentissage non supervisée de SDAE, 30 epochs ont été trouvées suffisantes. Le pré-apprentissage s'arrête si jamais le MSE ne passe pas en dessous de 10^{-2} durant les dernières 5 epochs et la taille du mini-batch a été fixée à 32 dans toutes les expérimentations.

4.4 Résultats

Dans la première étape, pour le nombre donné de 4 clusters, la visualisation de dix jours aléatoires de chaque cluster dans Fig. 4.3 montre une courbe qui présente une similitude entre les jours qui sont dans le même cluster et la dissemblance avec jours appartenant à d'autres clusters.

Pour identifier chaque groupe, les jours de chaque groupe sont regroupés par

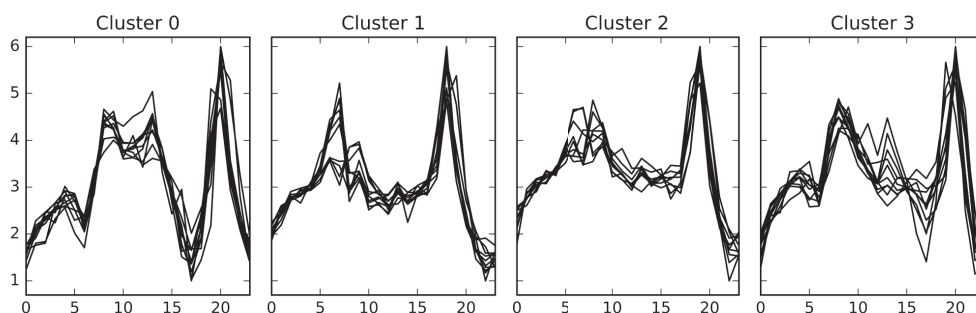


FIGURE 4.3 – Visualisation de dix observations aléatoires de chaque cluster.

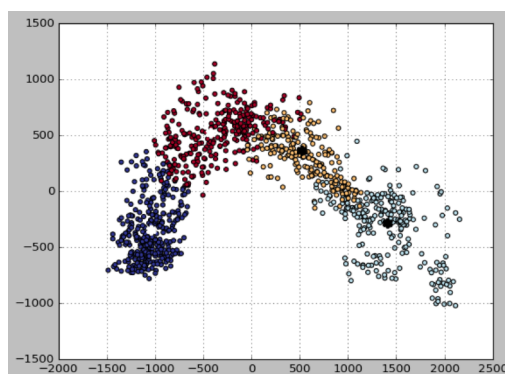


FIGURE 4.4 – Les clusters de K-means de la charge électrique algérienne (données réduit par ACP)

semaine et par mois. Le regroupement mensuel (Fig. 4.5b) montre que le groupe 3 couvre presque toutes les observations de mai et de juin, et que le groupe 2 contient la plupart des jours de mars. De plus, le groupe 0 regroupe les mois d'été (juillet et août) et le groupe 1 regroupe les mois d'hiver (novembre, décembre, janvier). Ces observations montrent une relation entre les variations de type diurne et saisonnière. Par conséquent, le regroupement par jour de la semaine dans Fig. 4.5a ne montre aucune dominance d'un jour particulier.

De plus on a appliqué l'Analyse en composantes principales ACP pour réduire la dimension des données dans le but de faire la visualisation en 2D comme illustré dans la figure 4.4. On peut remarquer que le centre des groupe des mois d'hiver (Groupe 1) est loin du centre des mois d'été (Groupe 0)

Afin d'établir un modèle dédié à chaque groupe, une analyse basée sur la date (1er, 2e, jusqu'au 30 ou 31) pour chaque jour des mois éparpillés entre les groupes

TABLE 4.3 – Taille des données pour l'apprentissage, la validation et le test.

Cluster	Heures d'apprentissage (60%)	Heures de validation (20%)	Heures de test (20%)	Total
Cluster 0	3888	1296	1296	6480
Cluster 1	6278	2093	2093	10464
Cluster 2	4968	1656	1656	8280
Cluster 3	3614	1205	1205	6024

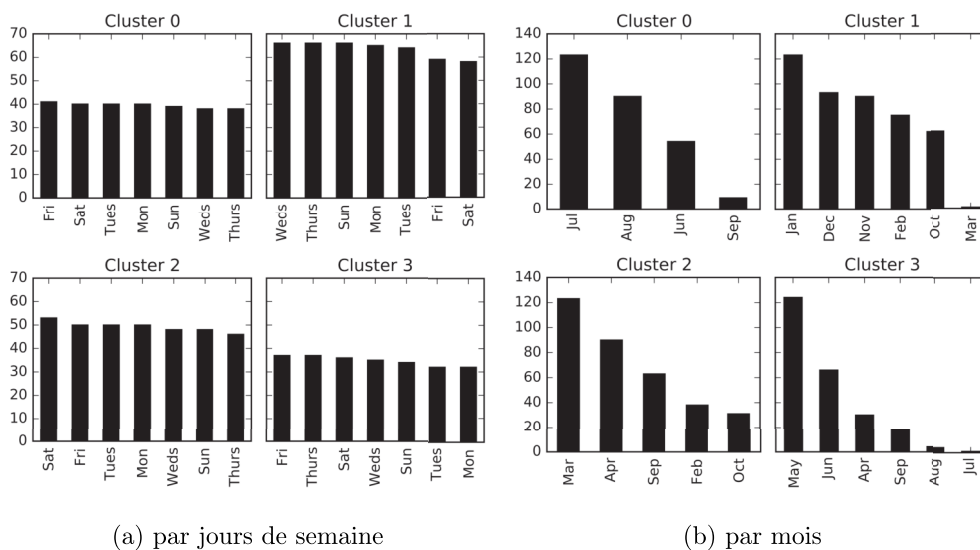


FIGURE 4.5 – Nombre d'observations par jour de la semaine et par mois dans chaque classe.

est effectuée et a été trouvée généralement liée au comportement social ; comme la reprise du travail après les vacances. Par conséquent, les jours de février qui suivent le 20 sont affectés au groupe 2, ce qui fait 30 observations à partir de 38 observations, alors que les jours précédant le 20 sont attribués au groupe 1 (68 à partir de 75 observations). De la même manière, le reste des mois est divisé par rapport au jour de saisie concerné.

Ensuite, les données de chaque cluster sont divisées en trois partitions (train, validation et test) dans la deuxième étape. La tableau 4.3 a décrit la longueur des données dans chaque cluster. Où, l'ordre chronologique des données dans chaque cluster est préservé et la dernière division est utilisée comme partition de test.

En plus de l'analyse des clusters, des profils quotidiens d'estimation des tem-

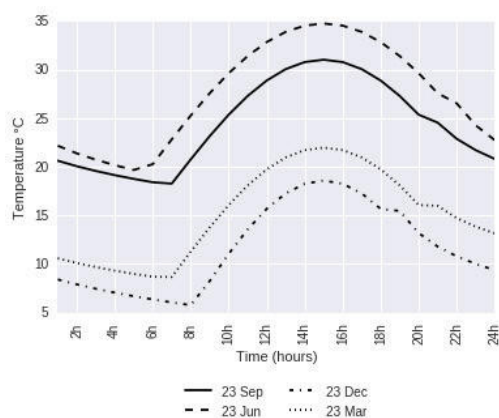


FIGURE 4.6 – La variations de la température horaires estimées sur une période de 24 heures.

pératures ont été générés. Dans la Fig. 4.6, nous remarquons la variation entre les saisons dans le temps approximatif de la température minimale et maximale. Ainsi la variation entre les saisons dans l’augmentation après le lever du soleil et la diminution après le coucher du soleil. En outre, la température minimale se produit toujours au lever du soleil en raison de deux mécanismes. Premièrement, la surface de la Terre arrête de chauffer car le rayonnement solaire n’est pas absorbé après le coucher du soleil. Deuxièmement, la conduction et la convection transportent l’énergie thermique dans l’atmosphère et l’air chaud qui était à la surface est remplacé par de l’air plus frais par le haut à cause du mélange atmosphérique [Pidwirny 2006].

Ces résultats montrent que la méthode proposée représente mieux les tendances de refroidissement mensuelles par rapport aux méthodes linéaires et cubiques. L’estimation montre également une meilleure représentation de l’augmentation et de la diminution lorsque la température maximale ou minimale quotidienne est atteinte.

Dans la deuxième étape, le premier résultat notable est que les estimations des profils de température aident les modèles à obtenir de meilleures performances que ceux utilisant seulement la température maximale et minimale, comme indiqué dans le tableau 4.4. Ensuite, la charge d’heures prévue est répétée afin d’obtenir une prévision d’un jour. Nous avons observé que l’estimation de la température donne une amélioration de 0.1% MAPE pour l’ensemble complet de données.

Les résultats du modèle de régression linéaire multiple sont donnés dans le tableau 4.5, où une prévision MAPE = 3,28% de 24 heures est obtenue. Le modèle a régression linéaire multiple est ici considéré comme un modèle de référence et de

TABLE 4.4 – Effet d’estimation du profil de température sur MAE pour chaque cluster

Modèle de cluster	MAE sans profils de température quotidien estimés	MAE avec profils de température quotidien estimés
Custer 0	0.0167	0.0157
Custer 1	0.0135	0.0134
Custer 2	0.0144	0.0143
Custer 3	0.0137	0.0132

TABLE 4.5 – MAE et MAPE de chaque modèle de cluster en utilisant la régression linéaire multiple

	MAPE	MAE
Cluster 0	1.96%	0,1725
Cluster 1	1.89%	0.1334
Cluster 2	2.16%	0.1451
Cluster 3	2.25%	0.1618

comparaison à l’approche proposée.

Le meilleur modèle pour chaque cluster est choisi en fonction de l’erreur de validation et l’évaluation de ces modèles est donnée dans la tableau 4.6. Où, les deux types de modèles de réseaux de neurones (avec et sans pré-apprentissage) ont donné des résultats presque similaires, SDAE montrant une meilleure performance de généralisation sur l’ensemble de test et une réduction du sur-apprentissage, surtout en été (Cluster 0 et 3).

Les résultats des modèles RNA sont moyennés et comparés avec un modèle RNA global qui utilise les données de tous les clusters pour l’entraînement où une amélioration bénéfique est donnée par la technique proposée (RNA séparées pour chaque cluster) comme indiqué dans le tableau 4.7. De plus, cela permet aux opérateurs d’analyser différents scénarios de prévision séparément pour chaque type de jour, où chaque type de jour est différent en termes de comportement du consommateur.

La meilleure topologie de système, en ce qui concerne les performances sur les ensembles de validation, est donnée dans Table 4.8 pour les modèles SDAE et Table 4.9 pour les modèles RNA, où nous pouvons voir qu’un modèle SDAE avec deux couches

TABLE 4.6 – Moyenne et l'écart type pour MAE et MAPE de chaque modèle de Cluster sur les ensembles de tests calculés sur les 10 meilleures erreurs de validation.

	MAE($\mu \pm \sigma$)		MAPE($\mu \pm \sigma$)	
	RNA	SDAE	RNA	SDAE
Cluster 0	0.1629 \pm 0.0133	0.1565 \pm 0.0287	1.9496 \pm 0.0012	1.9112 \pm 0.0036
Cluster 1	0.1156 \pm 0.0070	0.1150 \pm 0.0058	1.7522 \pm 0.0017	1.7456 \pm 0.0013
Cluster 2	0.1275 \pm 0.0120	0.1225 \pm 0.0212	1.9769 \pm 0.0034	1.9046 \pm 0.0062
Cluster 3	0.1152 \pm 0.0114	0.1116 \pm 0.0119	1.7635 \pm 0.0020	1.7228 \pm 0.0030

TABLE 4.7 – MAPE du modèle global et moyenne pondérée des modèles RNAs sur les ensembles de test calculés sur les 10 meilleures erreurs de validation.

Modèle	MAPE
RNA Global	1.9073 \pm 0.0279
Moyenne pondérée RNAs	1.8548 \pm 0.0021

TABLE 4.8 – Paramètres utilisés pour obtenir les résultats des modèles SDAE sélectionnés.

	Pas d'apprentissage de pré-apprentissage	Pas d'apprentissage	Tailles de couche cachée
Cluster 0	0.044	0.012	23 - 35
Cluster 1	0.034	0.012	30 - 6
Cluster 2	0.068	0.051	30 - 23
Cluster 3	0.083	0.064	31 - 31

TABLE 4.9 – Paramètres utilisés pour obtenir les résultats des modèles RNA sélectionnés.

	Pas d'apprentissage	Tailles de couche cachée
Cluster 0	0.095	15 - 25
Cluster 1	0.016	6 - 9
Cluster 2	0.019	22 - 9
Cluster 3	0.037	19 - 7

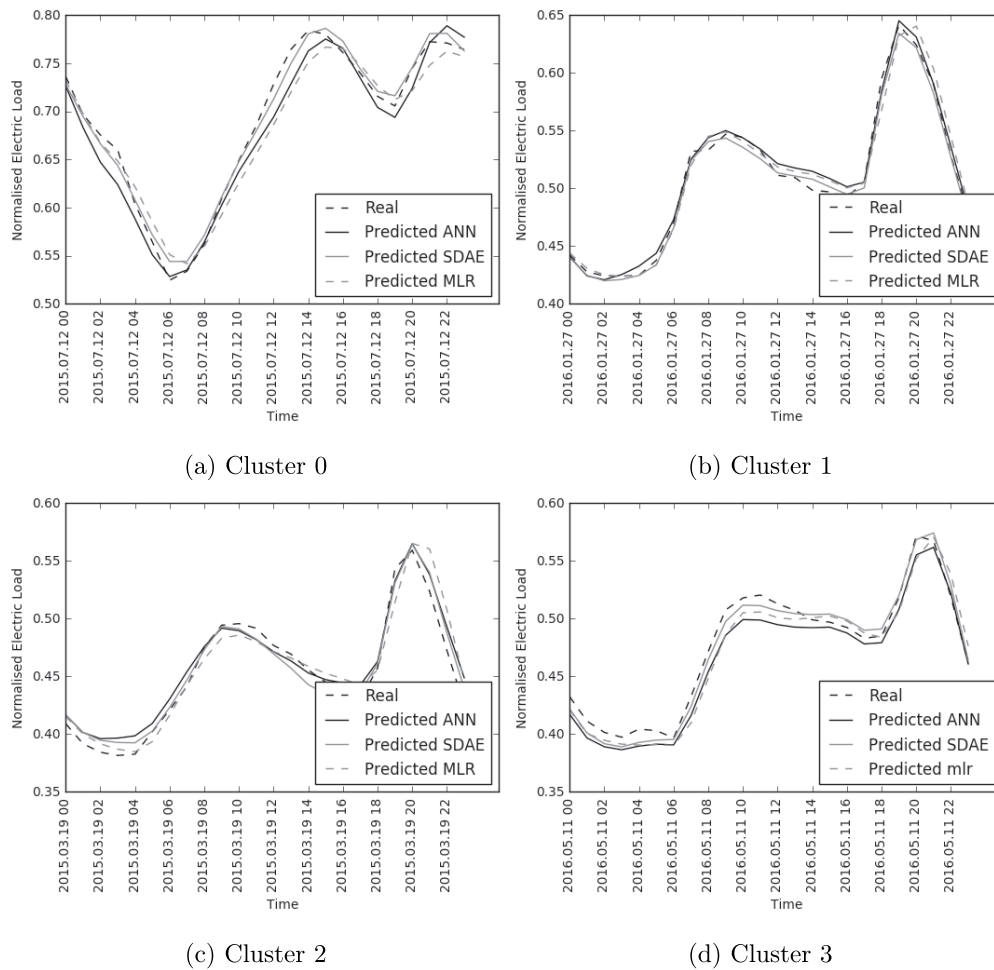


FIGURE 4.7 – Résultats de prévision de 24 heures en utilisant les modèles sélectionnés RNA et SDAE pour chaque Clusteur.

cachées donne une meilleure performance avec plus de neurones cachés par rapport aux topologies RNA sans pré-apprentissage. Les modèles sélectionnés pour chaque clusteurs sont donnés en utilisant un pas d'apprentissage pour un pré-apprentissage non supervisé supérieur au pas d'apprentissage pour un ajustement supervisé. Cette combinaison de paramètres montre la capacité de DAE à extraire des caractéristiques avec un niveau de corruption suffisant au niveau de la couche d'entrée.

Fig. 4.7 présente les résultats de la prédiction de 24 pas en avant par rapport aux charges réelles. On peut voir que les modèles RNA et SDAE s'adaptent bien à la forme quotidienne de la charge globale, avec une faiblesse prédisant les pics et les creux. Ceci est compréhensible en raison de la nature non linéaire élevée de

TABLE 4.10 – Evaluation du Modèle Saisonnier (été) pour les deux mois Juillet-Aout 2016.

Prévision 1H							
Palier <	%	Nb de jour	Arrondi	Palier >	%	Nb de jours	Arrondi jours sup
% <1%	33,87	20,66	21 jours	% >1%	66,13	40,34	40 jours
% <2%	69,35	42,30	42 jours	% >2%	30,65	18,70	19 jours
% <3%	87,10	53,131	53 jours	% >3%	12,9	7,87	8 jours
% <4%	85,55	52,18	52 jours	% >4%	14,45	8,81	9 jours
% <5%	96,77	59,03	59 jours	% >5%	3,23	1,97	2 jours
Prévision 24H							
% <1%	0	0		% >1%	100	61	61 jours
% <2%	3,33	2,03	2 jours	% >2%	96,67	58,97	59 jours
% <3%	41,94	25,58	26 jours	% >3%	58,06	35,41	35 jours
% <4%	67,74	41,32	42 jours	% >4%	32,26	19,68	20 jours
% <5%	91,94	56,08	56 jours	% >5%	8,06	4,92	5 jours
% <6%	95,16	58,04	58 jours	% >6%	4,84	2,96	3 jours

ces phénomènes. De plus, l'erreur absolue correspondante de prévision d'un jour d'avance est de 2,65% pour les modèles RNA et de 2,62% pour les modèles SDAE.

Toutefois, l'évaluation prend en compte toutes les journées, incluant des journées à événement spéciaux nécessitant d'autres données exogènes que nous n'avons pas. Exemple des feux de forêts, impossible à prédire et qui affecteront la température, des événements sociaux et sportifs, etc.

Dans le cas de journées conventionnelles, les modèles développées sont plus performants et l'erreur diminue même en dessous de 1%. Le tableau (4.10) donne une évaluation pour les mois de Juillet Août 2016, et montre que pour la plupart des journées l'approche est très performante. Le % est alors tiré vers le bas par des journées non conventionnelles.

Les résultats présentés dans le tableau 4.6 sont encourageants. En observant les résultats d'erreur, il est clair que le SDAE a de meilleurs résultats que ceux obtenus par le RNA sans pré-apprentissage.

Enfin, la table 4.9 rapporte une comparaison de certaines performances de littérature. Gardant à l'esprit la spécificité de la charge algérienne héritée d'un grand pays avec des différences importantes en termes de météo, de traditions, d'industrialisation etc. induisant une non-régularité dans la série de temps de chargement,

conduisant à un problème de prévision plus difficile.

Les performances de l'approche proposée sont proches ou supérieures aux modèles de référence. De plus, l'estimation du profil journalier de température pourrait être utilisée pour améliorer les performances au lieu d'utiliser la variable de température moyenne quotidienne comme dans [Çevik 2015]. En plus de la température, la vitesse du vent et l'humidité, qui permettront le calcul ou l'estimation de la température ressentie, pourraient également être utilisées pour améliorer les résultats du système de prévision.

Cependant, les avantages ont été jugés minimes et une meilleure initialisation du réseau n'améliore pas drastiquement les résultats en termes de mesure d'erreur. Les résultats sont assez modérés par rapport aux améliorations impressionnantes et à la réalisation de l'apprentissage profond à surpassé l'état de l'art dans diverses tâches difficiles [Vincent 2010], même si cela donne une meilleure performance de généralisation. La relation entre les séries temporelles de charge électrique et le comportement social reflété par les variables du calendrier, et la faible dimensionnalité de la tâche (prévisions en utilisant 16 entrées) diminuent l'avantage d'utiliser les modèles SDAE comme indiqué dans [Vincent 2010] avec des tâches de dimensionnalités supérieures (comme dans le traitement d'image, la reconnaissance de formes, etc.)

4.5 PREVOS-DZ

Les résultats de cette étude ont été mis en œuvre par le Laboratoire de Gestion Électronique de Documents (LabGED) dans la conception et le développement d'une application de prévisions nommé « PREVOS-DZ » qui intègre les modèles créés dans ces expérimentations.

PREVOS-DZ est un logiciel de prévisions conçu spécialement pour les prévisions de la charge électrique à court terme, à la demande ou en temps réel. En permettant l'utilisation des modèles dans plusieurs aspects du logiciel afin de garantir des résultats qui augmentent la productivité et la fiabilité des services assurés par l'OS.

La conception de PREVOS-DZ, se devait de répondre à plusieurs demandes, comme la disponibilité des données de consommation et des températures, et le pouvoir de faire des prévisions à tout moment et en temps réel.

TABLE 4.11 – Performances de certaines études sur la prévision de charge électrique à court terme

Référence	Entrées utilisées	Région étudiée	Modèles	MAPE
Çevik and Çunkaş [Çevik 2015]	Historique de la charge, Variables calendaires, Température	Turquie	Système d'inférence adaptative neuro-flou	1.85
Fan and Hyndman [Fan 2012]	Historique de la charge, Variables calendaires, Température	Australie	RNA et Modèle additif	2.08
Chen et al. [Chen 2010]	Historique de la charge, Variables calendaires, Température, Vitesse du vent, Humidité, Précipitations	Nouvelle-Angleterre	Décomposition d'ondelettes et RNA	1.63
Farfar et al. [Farfar 2015]	Historique de la charge	Algérie	Parallèle RNAs	2.68
Notre étude	Historique de la charge, Variables calendaires, Température	Algérie	SDAE	1.84

4.5.1 Prévision de la charge électrique

La prévision de la charge électrique est l'objectif principale de PREVOS-DZ, le logiciel donnera aux utilisateurs la possibilité d'avoir deux alternatives pour faire des prévisions, la prévision RLM et la prévision RNA, la première utilisera la méthode de régression linéaire multiple, elle sera moins efficace et moins précise que la deuxième méthode qui elle utilise les réseaux de neurones artificiels.

PREVOS-DZ permet de faire des prévisions de la charge électrique pour chaque région du pays (Est, Ouest, Sud, Centre), ainsi qu'au niveau national, cela en offrant des différents modèles qui sont cités ci-dessous :

- **Modèles Globaux** : Ces modèles ont été créés en utilisant toutes les données sans se soucier d'une période donnée.
- **Modèles saisonniers** : Les données ont été divisées en 4 groupes chaque groupe contient les données de la même saison, ensuite chaque groupe à été utilisé pour l'apprentissage de 4 modèles différents (été, hiver, printemps, automne).
- **Modèles journaliers** : Les ont été divisées en 3 groupes selon le jour de la semaine (Jour férié : vendredi, jour veille de jour ouvrable : samedi, jour ouvrable : dimanche ... jeudi), ensuite un modèle à été entraîné pour chaque groupe (JF, JVO, JO).
- **Modèles mensuels** : même chose que les modèles saisonniers et journaliers sauf qu'ici un modèle pour chaque mois a été crée.
- **Modèles du Ramadan** : Les modèles Ramadan ont créés en utilisant les données de ce dernier.

4.5.2 Technologies utilisées

PREVOS-DZ a été réalisé avec le langage de programmation C# en utilisant l'environnement de développement Microsoft Visual Studio 2017 sous .NET Framework 5.2.1. PREVOS-DZ utilise une base de données SQLite pour la gestion des données et l'API « NNSHARP » qui permet d'utiliser des modèles d'apprentissage automatique implémentés à l'aide de Keras (une bibliothèque de python) sur C#, en sérialisant les modèles Keras en JSON, puis en les dé-sérialisant de JSON vers un Objet spécifique a « NNSHARP » qui pourra fournir toutes les informations sur le modèle.

4.5.3 Fonctionnalités de PREVOS-DZ

L'interface de PREVOS-DZ se divise en huit volets, où chaque volet permet d'effectuer plusieurs opérations.

4.5.3.1 Volet accueil

Ce volet est affiché dès l'exécution du logiciel (Figure 4.8). Il est composé de 3 sous-sections qui permettent respectivement, la gestion des études (créer une nouvelle étude, l'ouvrir, ou la supprimer), de définir les paramètres (Réinitialisation des paramètres, Langue) ou la gestion de connexion (gestion des utilisateurs, connexion, déconnexion et quitter l'application)

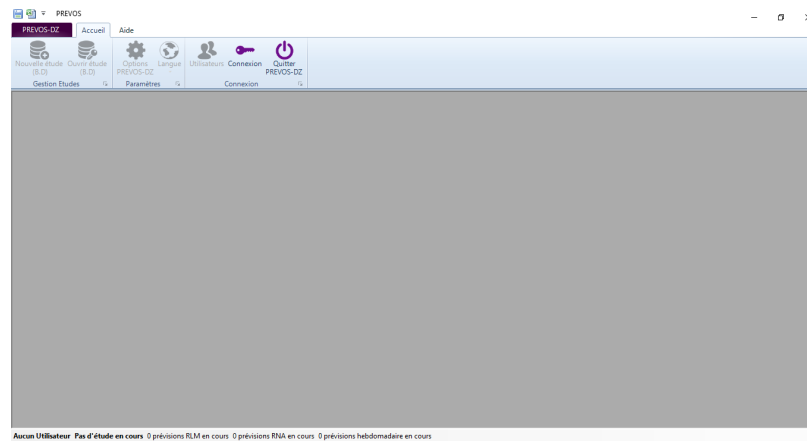


FIGURE 4.8 – Volet accueil de PREVOS-DZ

4.5.3.2 Volet données

Le volet *données* (Figure 4.9) permet de consulter et manipuler toutes les données présentes sur l'application, il est divisé en 4 parties distinctes (Courbes de charges et températures, RLM et Modèles par Défaut), chacune contenant plusieurs options.

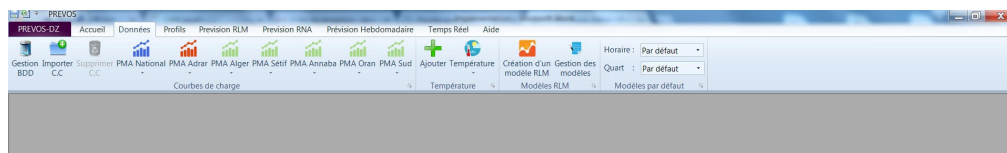


FIGURE 4.9 – Volet données de PREVOS-DZ

Courbes de charge

La section *courbes de charge* (CC), contient les courbes de charge PMA de tout l'historique national et régional ainsi que des boutons pour l'importation des données dans la base de données de PREVOS-DZ à partir d'un fichier Excel et la suppression d'une CC de types PMA. L'utilisateur peut consulter et visualiser toutes les données de consommation appelées PMA, disponible dans la base de données.

Température

Ce volet permet d'ajouter une nouvelle température par région en important un fichier Excel. Ainsi, L'utilisateur peut consulter les températures de chaque région. Il offre une liste déroulante qui permet de choisir l'année, et une fois l'année choisie, les températures de cette année s'afficheront. L'utilisateur pourra choisir la région dont il veut afficher les températures.

Modèles RLM

Ce pavé contient deux boutons, le premier *Création d'un modèle RLM* permet à l'utilisateur de créer son propre modèle de régression linéaire multiple, et le deuxième *Gestion des modèles* qui permet de lister tous les modèles disponibles et les modifiés.

4.5.3.3 Volet profils

Le volet *profils* (Figure 4.10) n'est visible et actif que lorsque une étude à été choisie au préalable par l'utilisateur. Ce volet permet de créer, de consulter, de modifier et de visualiser différents types de profils, le profil dans ce cas est le comportement des données par rapport à une période prédéfinie ou flottante que nous définissons, les périodes que nous avons prédéfinies sont le journalier en prenant en considération trois types de jours (jours ouvrables, jours fériés, et veille de jours ouvrables), le profil hebdomadaire, le profil annuel, et le profil flottant définie entre deux périodes de temps.

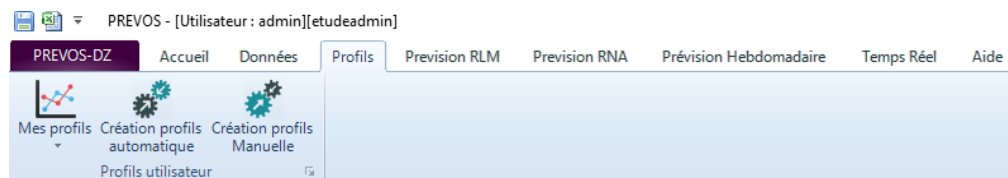


FIGURE 4.10 – Volet profils de PREVOS-DZ

4.5.3.4 Volet prévisions RLM et RNA

Ces deux volets (Figure 4.11) permet de faire des prévisions en utilisant des modèles de RLM et des modèles RNA. Ils contiennent six (06) sections, chaque section est expliquée dans le tableau 4.12.

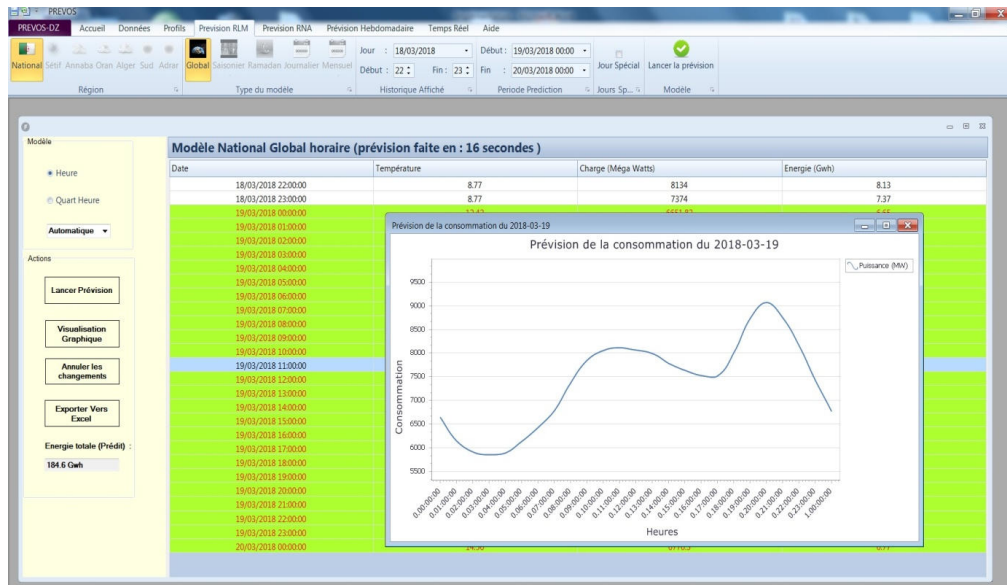


FIGURE 4.11 – Volet prévisions RLM de PREVOS-DZ

4.5.3.5 Volet prévisions hebdomadaires

Ce volet (Figure 4.12) se focalise sur la prévision à long terme, allant jusqu'à une année complète, par pas hebdomadaire.

4.5.3.6 Volet temps réel

La prévision à court terme, utilise les deux modèles vus précédemment RLM et RNA, elle est caractérisée dans PREVOS-DZ par deux fonctionnalités.

La prévision à la demande donne à l'utilisateur le pouvoir de choisir la région pour laquelle la prévision doit être effectuée, le type du modèle à utiliser, la plage horaire qu'il veut prédire, ainsi que le format des données (horaire ou quart-horaire), les températures (max, min et actuel).

La prévision en temps réel utilise le même algorithme que la prévision à la demande, à l'exception qu'elle commence la prévision à partir de l'heure actuelle et qu'elle dure 10 jours. L'utilisateur aura accès à un outil qui lui permettra de choisir

TABLE 4.12 – Fonctionnalités du volet RLM et RNA

Fonction		Action
Région	National	Choisit les données Nationales
	Annaba	Choisit les données d'Annaba
	Alger	Choisit les données d'Alger
	Oran	Choisit les données d'Oran
	Sétif	Choisit les données de Sétif
	Sud	Choisit les données du Sud
Type du modèle	Global	Choisit un modèle RLM / RNA global
	Saisonnier	Choisit un modèle RLM / RNA saisonnier au choix
	Ramadan	Choisit un modèle RLM / RNA Ramadan
	Mensuel	Choisit un modèle RLM / RNA mensuel au choix
	Journalier	Choisit un modèle RLM / RNA journalier
Historique affiché	Jour	Choisir le jour de l'historique
	Début	Choisir l'heure du début de l'historique
	Fin	Choisir l'heure de fin de l'historique
Période Prédiction	Début	Choisir l'heure de début de la prévision
	Fin	Choisir l'heure de fin de la prévision
Jours Spéciaux	Jour spécial	Fixer les jours spéciaux
Modèle	Lancer la prévision	Lancer une prévision

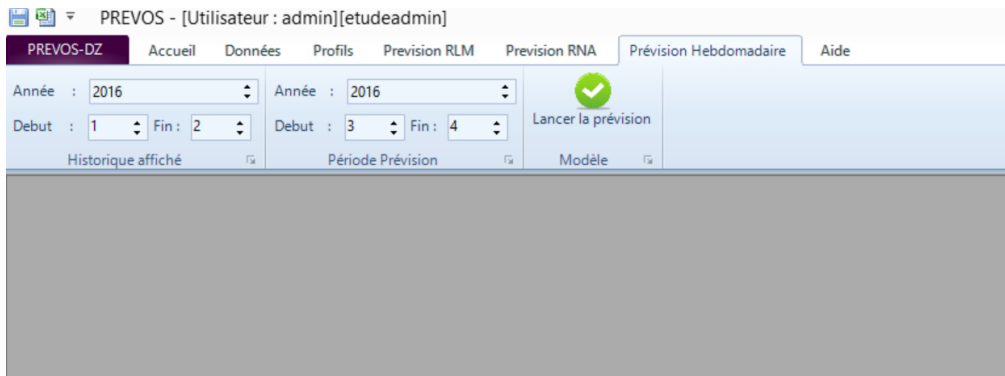


FIGURE 4.12 – Volet prévisions hebdomadaires de PREVOS-DZ

quel modèle de prévision assigné à chaque journée et quel sont les jours spéciaux parmi les 10 jours prédits.

Le volet temps réel permet donc de faire des prévisions qui commencent à partir de l'heure actuelle, et s'étendent sur 10 jours.

4.5.3.7 Volet aide

Le volet Aide contient deux sections « Manuel d'aide » et « A propos.. », ils contiennent les directives proposées aux personnels de l'OS.

4.6 Conclusion

Dans notre contributions, nous avons proposé une approche à deux étapes pour la prévision de charge électrique à court terme, où l'algorithme de classification K-Means est utilisé pour définir un ensemble de classes de jour de profils. Le pré-apprentissage des auto-encodeurs débruiteurs empilés est utilisée pour réaliser un apprentissage profond de réseaux neuronaux artificiels capables de prévoir indépendamment la charge électrique à court terme pour différents jours. Les expériences montrent que les profils quotidiens de température estimés aident les modèles à obtenir de meilleures performances. De plus, les modèles SDAE sont capables d'extraire des caractéristiques et leurs performances de généralisation sont meilleures que celles des RNAs sans pré-apprentissage. Cependant, l'avantage d'utiliser des modèles SDAE n'est pas écrasant comparé aux améliorations de l'apprentissage profond dans d'autres tâches, mais c'est un résultat préliminaire prometteur.

L'approche proposée dans ce chapitre a été concrétisée par le développement

d'un outils nommé PREVOS-DZ. Il propose différents fonctionnalités qui permis de faciliter la gestion des données de consommation, de pouvoir exploiter ces données en offrant la possibilité de dresser des profils à l'aide de ces dernières, PREVOS-DZ a aussi bien répondu a son principal objectif qui consistait à prodiguer de bonnes prévisions à court terme à l'aide des modèles RLM et RNA, mais l'application va encore plus loin en permettant de créer des modèles hebdomadaires qui peuvent donner des prévisions horaires journalières cela en appliquant des profils à ces prévision, Enfin la prévision en temps réel vient parachever l'application en offrant une prévision continue qui permet à l'OS de mieux gérer la production en temps réel.

Les recherches futures se concentreront sur le STLF qui implique les résultats de cette approche, soit en incluant la température apparente comme une entrée, soit en incluant d'autres variable exogènes provenant des paramètres atmosphériques qui affectent les sources d'énergie renouvelables telles que l'énergie éolienne et solaire.

Conclusion

Cette thèse porte sur l'étude et la proposition d'une méthodologie pour la prédiction de la charge électrique à court terme. Nous avons mis l'accent sur l'utilisation des réseaux de neurones profonds et plus particulièrement la méthode de pré-apprentissage non-supervisé par couche utilisant une architecture à deux couches cachées. Ceci pour l'approximation de fonction de la série temporelle de la consommation électrique horaire Algérienne. En tant que benchmark de comparaison et évaluation, nous avons utilisé les résultats obtenus à partir d'un RNA à propagation avant sans pré-apprentissage et une autre méthode traditionnelle de prédiction des séries temporelles qui est la régression linéaire multiple.

La prédiction de la charge électrique est très importante pour la bonne gestion des ressources d'énergie et le pilotage du système électrique en temps réel. Les conditions météorologiques telles que la température, l'humidité, la nébulosité et la vitesse du vent jouent un rôle vital, elles sont considérées comme un facteur exogène dans la modélisation de la charge électrique. Cependant, les services météorologiques ne servent pas souvent des profils horaires de prévision de la température et fournissent seulement la prévision des valeurs minimums et maximums de la température.

Dans notre travail, nous avons proposé d'utiliser une technique d'estimation de profile de température basée sur les valeurs de température minimum et maximum prévue. Cette méthode d'estimation est particulièrement utile en cas de manque des données historiques de la température régulière.

L'approche proposé dans cette thèse passe par deux étapes. Premièrement, une identification d'un ensemble de type des jours est établie en utilisant l'algorithme de classification non supervisé K-Means, l'analyse des résultats de ce partitionnement de données a permet de grouper les périodes qui ont une similarité dans le profil de consommation au cours d'une journée. Dans la deuxième étape, les données de chaque type de jours sont utilisées pour construire un modèle de RNA profond pour la prédiction de la consommation horaire.

L'analyse de la consommation électrique Algérienne montre que cette dernière suit une un cycle saisonnier et un cycle hebdomadaire. Pour cela, les données d'apprentissage ont été renforcées par l'estimation de la température prédite et des variables calendaires qui sont : les jours de semaine et si le jour à prédire est un jour férié ou non. Ainsi, une normalisation de données basée sur la prédiction à moyen terme est proposée pour éviter la saturation de la fonction d'activation vu que la consommation suit une tendance croissante.

Le RNA profond est construit par l'empilement d'auto-encodeur débruiteur. La théorie de ce type de RNA est présentée dans le deuxième chapitre qui peut être résumée par une procédure à deux étapes. La première étape consiste à appliquer un algorithme d'apprentissage non-supervisé par couche cachée en utilisant un RNA de type auto-encodeur sur des entrées bruitées, par la suite les poids de l'encodeur sont utilisés comme une initialisation des poids des couches cachées, et les sorties du décodeur sont utilisées comme une entrée de la couche cachée suivante. La deuxième étape consiste à appliquer l'algorithme de rétro-propagation afin d'affiner l'ensemble des poids du réseau.

Pour choisir les hyper-paramètres, une combinaison de la recherche en grille et la recherche aléatoire est utilisée pour tester les différentes configurations telles que le nombre de couche cachée, le nombre de neurones par couche, le pas d'apprentissage et le pas de pré-apprentissage. Ainsi, nous avons testé différentes combinaison d'entrées pour notre modèle, avec et sans entrée exogène particulièrement la température.

Pour assurer la bonne analyse de performance des résultats présentés dans ce travail, les données enregistrées par la compagnie Sonelgaz dans la période du 2013 au 2016 ont été partitionnées comme suit : 60 % ont été utilisées pour la création du modèle, 20 % pour l'optimisation des hyper-paramètres et les 20% restantes ont été utilisées pour former l'ensemble de test afin d'évaluer la qualité du modèle.

Les expériences établies au cours de ce travail montrent que les profils quotidiens de température estimés aident les modèles à obtenir de meilleures performances. De plus, La performance de généralisation des modèles SDAEs est meilleure que celle des RNAs sans pré-apprentissage. Cependant, l'utilisation des modèle SDAE ne permettent pas d'obtenir des améliorations drastiquement importantes par rapport aux RNAs profonds quand ces derniers sont utilisées pour des problèmes de reconnaissances.

Toutefois, les résultats obtenus avec ces données ont été très intéressants. L'erreur de test était 1.84% et elle est considérée comme une mesure de la puissance prédictive de notre modèle à des données non vues. Élargir le pas de prédiction vers 24 heures itérativement a aussi donné des résultats encourageants, où l'erreur moyenne de MAPE était 2.62%.

Perspectives

Notre travail a donné des résultats intéressants ainsi que des pistes sur la façon dont nous pouvons profiter des recherches récentes de l'apprentissage profond des

RNAs pour l'approximation des séries temporelles, particulièrement la prédiction de la charge électrique à court terme. Nous croyons que des meilleurs résultats peuvent être obtenus si nous incluons plus des facteurs qui influencent la consommation électrique et exploiter les avancées de l'apprentissage profond dans la modélisation des données statiques à travers les domaines d'intelligence artificielle tels que la reconnaissance d'objet.

Les travaux effectués dans cette thèse dégagent plusieurs améliorations envisageables et plusieurs aspects méritant d'être explorés dans le cadre des travaux futurs, telles que :

- Construire un modèle basé sur une décomposition contenant une composante de saisonnalité, tendance et une composante non linéaires.
- Utiliser les données des événements susceptible d'affecter la consommation telles que les fêtes annuelles.
- Utiliser d'autres variables météorologique telles que la vitesse du vent et l'humidité.
- Utiliser un indice de confort humain basé sur la température mesurée, la vitesse du vent et de l'humidité relative comme expliquer dans le troisième chapitre.
- Utiliser d'autres types de RNA comme les réseaux récurrents.
- Utiliser d'autres algorithmes pour l'identification des types des jours.

Calcul de la durée du jour

A.1 Calcul de la durée du jour

L'orbite de la planète Terre est un long chemin qui représente son mouvement autour du Soleil dans le sens inverse des aiguilles d'une montre, ceci, comme le montre le schéma de la figure A.1. La position de la Terre sur l'orbite change d'un jour à un autre et plusieurs travaux dans la littérature permettent son estimation. Parmi ces travaux nous nous sommes servis de celui par Almorox et al. [Almorox 2005] qui rapport la validation statistique des définitions de la longueur du jour pour l'estimation du rayonnement solaire global dans une ville en Espagne.

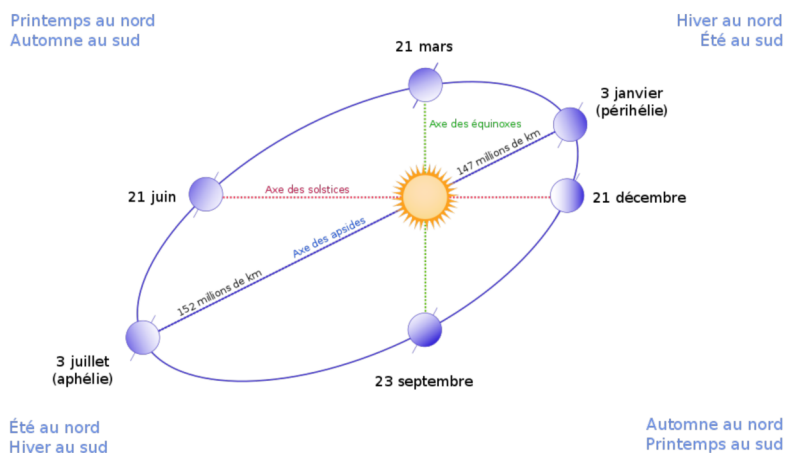


FIGURE A.1 – Schéma de l'orbite de la Terre. [Wikipedia 2018a]

En raison de l'inclinaison de l'axe de la Terre, la forme de son orbite autour du soleil change tout au long de l'année. Lorsque le pôle nord est incliné vers le soleil, le jour dure plus longtemps et le soleil apparaît plus haut dans le ciel, ce qui conduit à une température moyenne plus élevée. Lorsque le pôle nord est plus loin du soleil, où l'hémisphère nord ne fait plus directement face à la lumière, la température est généralement moins élevée, et c'est donc la saison d'hiver. En raison de l'inclinaison,

les régions polaire peuvent même ne recevoir aucun rayon de soleil pendant l'hiver. La durée d'un jour dépend donc de deux paramètres :

- La date, sous la forme de la position angulaire δ dans l'orbite elliptique ; δ s'appelle la déclinaison solaire.
- La latitude σ .

L'angle du jour τ (radians) est donné par :

$$\tau = \frac{2 * \pi}{365} * (N\text{Jour} - 1) \quad (\text{A.1})$$

Où $N\text{Jour}$ est le numéro du jour de l'année, à compter du 1er janvier.

La déclinaison solaire δ est calculée selon le modèle de Spencer [Spencer 1971] comme suit :

$$\delta = \frac{180}{\pi} * (0.006918 - 0.399912 * \cos(\tau) + 0.070257 * \sin(\tau) - 0.006758 * \cos(2 * \tau) + 0.000907 * \sin(2 * \tau) - 0.002697 * \cos(3 * \tau) + 0.00148 * \sin(3 * \tau)) \quad (\text{A.2})$$

L'angle géométrique moyen des heures sur une surface horizontale représentant un lever ou coucher de soleil théorique se produit quand le centre du Soleil est sur un angle d'élévation égal à zéro. Il peut être calculé en degrés à partir de :

$$\omega_0 = \frac{\sin(\frac{-0.8333}{360 * 2 * \pi}) - \sin(\frac{\lambda}{360 * 2 * \pi}) * \sin(\frac{\delta}{360 * 2 * \pi})}{\cos(\frac{\lambda}{360 * 2 * \pi}) * \cos(\frac{\delta}{360 * 2 * \pi})} \quad (\text{A.3})$$

La durée du jour ou la longueur du jour est une variable météorologique importante pour plusieurs modèles écologiques et études agro-environnementales. Cette variable qui varie au cours de l'année et mesure la durée en heures entre le lever et le coucher du soleil peut être estimé par [Almorox 2005] :

$$DL = \frac{2 * \arccos(\omega_0)}{\frac{15}{360 * 2 * \pi}} \quad (\text{A.4})$$

$$\text{Leve de soleil} = \text{Sunrise} = 12 - \frac{\cos(\omega_0)}{\frac{15}{360 * 2 * \pi}} \quad (\text{A.5})$$

$$\text{Coucher du soleil} = \text{Sunset} = 12 + \frac{\cos(\omega_0)}{\frac{15}{360 * 2 * \pi}} \quad (\text{A.6})$$

Bibliographie

- [Abdel-Aal 2005] RE Abdel-Aal. *Improving electric load forecasts using network committees*. Electric Power Systems Research, vol. 74, no. 1, pages 83–94, 2005. (Cité en pages 19 et 29.)
- [Alfares 2002] Hesham K Alfares et Mohammad Nazeeruddin. *Electric load forecasting : literature survey and classification of methods*. International Journal of Systems Science, vol. 33, no. 1, pages 23–34, 2002. (Cité en page 79.)
- [AlFuhaid 1997] AS AlFuhaid, MA El-Sayed et Magdi S Mahmoud. *Cascaded artificial neural networks for short-term load forecasting*. IEEE Transactions on Power Systems, vol. 12, no. 4, pages 1524–1529, 1997. (Cité en pages 18 et 28.)
- [Almorox 2005] J Almorox, C Hontoria et M Benito. *Statistical validation of day-length definitions for estimation of global solar radiation in Toledo, Spain*. Energy Conversion and Management, vol. 46, no. 9, pages 1465–1471, 2005. (Cité en pages 83, 109 et 110.)
- [Amjady 2001] N. Amjady. *Short-term hourly load forecasting using time-series modeling with peak load estimation capability*. IEEE Transactions on Power Systems, vol. 16, no. 4, pages 798–805, 2001. (Cité en pages 13 et 28.)
- [Amjady 2009] Nima Amjady et Farshid Keynia. *Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm*. Energy, vol. 34, no. 1, pages 46–57, 2009. (Cité en page 78.)
- [Auer 1996] Peter Auer, Mark Herbster et Manfred K Warmuth. *Exponentially many local minima for single neurons*. In Advances in neural information processing systems, pages 316–322, 1996. (Cité en page 50.)
- [Azadeh 2013] A Azadeh, R Babazadeh et SM Asadzadeh. *Optimum estimation and forecasting of renewable energy consumption by artificial neural networks*. Renewable and Sustainable Energy Reviews, vol. 27, pages 605–612, 2013. (Cité en page 21.)
- [Bakirtzis 1996] AG Bakirtzis, V Petridis, SJ Kiartzis, MC Alexiadis et AH Maisis. *A neural network short term load forecasting model for the Greek power system*. IEEE Transactions on power systems, vol. 11, no. 2, pages 858–863, 1996. (Cité en pages 18 et 28.)

- [Barakat 1990] E. H. Barakat, M. A. Qayyum, M. N. Hamed et S. A. Al Rashed. *Short-term peak demand forecasting in fast developing utility with inherit dynamic load characteristics. I. Application of classical time-series methods. II. Improved modelling of system dynamic load characteristics.* IEEE Transactions on Power Systems, vol. 5, no. 3, pages 813–824, 1990. (Cité en pages 13 et 28.)
- [Barakat 1992] EH Barakat, JM Al-Qassim et SA Al Rashed. *New model for peak demand forecasting applied to highly complex load characteristics of a fast developing area.* In IEE Proceedings C (Generation, Transmission and Distribution), volume 139, pages 136–140. IET, 1992. (Cité en pages 12 et 27.)
- [Bashir 2009] ZA Bashir et ME El-Hawary. *Applying wavelets to short-term load forecasting using PSO-based neural networks.* IEEE Transactions on Power Systems, vol. 24, no. 1, pages 20–27, 2009. (Cité en page 78.)
- [Baum 1989] Eric B Baum et David Haussler. *What size net gives valid generalization ?* In Advances in neural information processing systems, pages 81–90, 1989. (Cité en page 49.)
- [Benabbas 2008] Farouk Benabbas, Mohamed Tarek Khadir, Damien Fay et Ahmed Boughrira. *Kohonen map combined to the K-means algorithm for the identification of day types of Algerian electricity load.* In Computer Information Systems and Industrial Management Applications, 2008. CISIM'08. 7th, pages 78–83. IEEE, 2008. (Cité en pages 19 et 29.)
- [Benabbas 2011] Farouk Benabbas et Mohamed Tarek Khadir. *Fuzzy C-Means clustering and kohonen maps for the identification of regional electricity load day types.* International Journal of Hybrid Intelligent Systems, vol. 8, no. 2, pages 81–92, 2011. (Cité en pages 19, 29 et 82.)
- [Bengio 2007a] Yoshua Bengio, Pascal Lamblin, Dan Popovici et Hugo Larochelle. *Greedy layer-wise training of deep networks.* In P. B. Schölkopf, J. C. Platt et T. Hoffman, éditeurs, Advances in neural information processing systems, volume 19, pages 153–160. MIT Press, 2007. (Cité en pages 47, 53, 79 et 84.)
- [Bengio 2007b] Yoshua Bengio, Yann LeCun et al. *Scaling learning algorithms towards AI.* Large-scale kernel machines, vol. 34, no. 5, pages 1–41, 2007. (Cité en pages 49 et 50.)
- [Bengio 2009] Yoshua Bengio et al. *Learning deep architectures for AI.* Foundations and trends® in Machine Learning, vol. 2, no. 1, pages 1–127, 2009. (Cité en pages 46, 49 et 50.)

- [Bengio 2012] Yoshua Bengio. *Practical recommendations for gradient-based training of deep architectures*. In *Neural networks : Tricks of the trade*, pages 437–478. Springer, 2012. (Cité en page 85.)
- [Bergstra 2012] James Bergstra et Yoshua Bengio. *Random search for hyperparameter optimization*. *Journal of Machine Learning Research*, vol. 13, pages 281–305, 2012. (Cité en pages 44 et 87.)
- [Bianchi 2015] Filippo Maria Bianchi, Enrico De Santis, Antonello Rizzi et Alireza Sadeghian. *Short-term electric load forecasting using echo state networks and PCA decomposition*. *Ieee Access*, vol. 3, pages 1931–1943, 2015. (Cité en pages 22 et 29.)
- [Box 2015] George EP Box, Gwilym M Jenkins, Gregory C Reinsel et Greta M Ljung. *Time series analysis : forecasting and control*. John Wiley & Sons, 2015. (Cité en page 11.)
- [Brown 1999] Richard E Brown, Andrew P Hanson et David L Hagan. *Long range spatial load forecasting using non-uniform areas*. In *Transmission and Distribution Conference, 1999 IEEE*, volume 1, pages 369–373. IEEE, 1999. (Cité en page 24.)
- [Carreira-Perpinan 2005] Miguel A Carreira-Perpinan et Geoffrey E Hinton. *On contrastive divergence learning*. In *Aistats*, volume 10, pages 33–40. Citeseer, 2005. (Cité en page 45.)
- [Cecati 2015] Carlo Cecati, Janusz Kolbusz, Paweł Różycki, Pierluigi Siano et Bogdan M Wilamowski. *A novel RBF training algorithm for short-term electric load forecasting and comparative studies*. *IEEE Transactions on industrial Electronics*, vol. 62, no. 10, pages 6519–6529, 2015. (Cité en pages 22 et 29.)
- [Çevik 2015] Hasan Hüseyin Çevik et Mehmet Çunkaş. *Short-term load forecasting using fuzzy logic and ANFIS*. *Neural Computing and Applications*, vol. 26, no. 6, pages 1355–1367, 2015. (Cité en pages 26, 30, 79, 95 et 96.)
- [Chao 2011] Jing Chao, Furaio Shen et Jinxi Zhao. *Forecasting exchange rate with deep belief networks*. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 1259–1266. IEEE, 2011. (Cité en page 55.)
- [Charytoniuk 1998] W Charytoniuk, MS Chen et P Van Olinda. *Nonparametric regression based short-term load forecasting*. *IEEE transactions on Power Systems*, vol. 13, no. 3, pages 725–730, 1998. (Cité en pages 15 et 28.)
- [Chatfield 2000] Chris Chatfield. *Time-series forecasting*. CRC Press, 2000. (Cité en page 7.)

- [Chen 1992] S-T Chen, David C Yu et Alireza R Moghaddamjo. *Weather sensitive short-term load forecasting using nonfully connected artificial neural network*. IEEE Transactions on Power Systems, vol. 7, no. 3, pages 1098–1105, 1992. (Cité en pages 17 et 28.)
- [Chen 1995] Jiann-Fuh Chen, Wei-Ming Wang et Chao-Ming Huang. *Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting*. Electric Power Systems Research, vol. 34, no. 3, pages 187–196, 1995. (Cité en pages 12 et 27.)
- [Chen 2003] Hsin Chen et Alan F Murray. *Continuous restricted Boltzmann machine with an implementable training algorithm*. IEE Proceedings-Vision, Image and Signal Processing, vol. 150, no. 3, pages 153–158, 2003. (Cité en page 55.)
- [Chen 2004] Bo-Juen Chen, Ming-Wei Chang et al. *Load forecasting using support vector machines : A study on EUNITE competition 2001*. IEEE transactions on power systems, vol. 19, no. 4, pages 1821–1830, 2004. (Cité en pages 27 et 30.)
- [Chen 2010] Ying Chen, Peter B Luh, Che Guan, Yige Zhao, Laurent D Michel, Matthew A Coolbeth, Peter B Friedland et Stephen J Rourke. *Short-term load forecasting : similar day-based wavelet neural networks*. IEEE Transactions on Power Systems, vol. 25, no. 1, pages 322–330, 2010. (Cité en pages 21, 29, 79 et 96.)
- [Chiu 1997] Chih-Chou Chiu, Ling-Jing Kao et Deborah F Cook. *Combining a neural network with a rule-based expert system approach for short-term power load forecasting in Taiwan*. Expert Systems with Applications, vol. 13, no. 4, pages 299–305, 1997. (Cité en pages 25 et 30.)
- [Cho 2009] Youngmin Cho et Lawrence K Saul. *Kernel methods for deep learning*. In Advances in neural information processing systems, pages 342–350, 2009. (Cité en page 53.)
- [Chow 1996] TWS Chow et CT Leung. *Neural network based short-term load forecasting using weather compensation*. IEEE Transactions on Power Systems, vol. 11, no. 4, pages 1736–1742, 1996. (Cité en pages 18 et 28.)
- [Cybenko 1989] George Cybenko. *Approximation by superpositions of a sigmoidal function*. Mathematics of control, signals and systems, vol. 2, no. 4, pages 303–314, 1989. (Cité en page 48.)

- [Dash 1995a] PK Dash, AC Liew et S Rahman. *Peak load forecasting using a fuzzy neural network*. Electric Power Systems Research, vol. 32, no. 1, pages 19–23, 1995. (Cité en pages 25 et 30.)
- [Dash 1995b] PK Dash, AC Liew et G Ramakrishna. *Power-demand forecasting using a neural network with an adaptive learning algorithm*. IEE Proceedings-Generation, Transmission and Distribution, vol. 142, no. 6, pages 560–568, 1995. (Cité en pages 17 et 28.)
- [Dedinec 2016] Aleksandra Dedinec, Sonja Filiposka, Aleksandar Dedinec et Ljupco Kocarev. *Deep belief network based electricity load forecasting : An analysis of Macedonian case*. Energy, vol. 115, pages 1688–1700, 2016. (Cité en pages 58 et 59.)
- [Ding 2016] Ni Ding, Clémentine Benoit, Guillaume Foggia, Yvon Bésanger et Frédéric Wurtz. *Neural network-based model design for short-term load forecast in distribution systems*. IEEE Transactions on Power Systems, vol. 31, no. 1, pages 72–81, 2016. (Cité en pages 22 et 29.)
- [Drucker 1997] Harris Drucker, Christopher JC Burges, Linda Kaufman, Alex J Smola et Vladimir Vapnik. *Support vector regression machines*. In Advances in neural information processing systems, pages 155–161, 1997. (Cité en page 26.)
- [El Desouky 2000] AA El Desouky et MM Elkateb. *Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA*. IEE Proceedings-Generation, Transmission and Distribution, vol. 147, no. 4, pages 213–217, 2000. (Cité en page 78.)
- [El-Hawary 1990] ME El-Hawary et GAN Mbamalu. *Short-term power system load forecasting using the iteratively reweighted least squares algorithm*. Electric power systems research, vol. 19, no. 1, pages 11–22, 1990. (Cité en pages 15 et 28.)
- [Elrazaz 1989] Z. S. Elrazaz et A. A. Mazi. *Unified weekly peak load forecasting for fast growing power system*. IEE Proceedings C - Generation, Transmission and Distribution, vol. 136, no. 1, pages 29–34, 1989. (Cité en pages 13 et 27.)
- [Espinoza 2005] Marcelo Espinoza, Johan AK Suykens et Bart De Moor. *Load forecasting using fixed-size least squares support vector machines*. In International Work-Conference on Artificial Neural Networks, pages 1018–1026. Springer, 2005. (Cité en pages 27 et 30.)

- [Fan 1994] JY Fan et JD McDonald. *A real-time implementation of short-term load forecasting for distribution power systems*. IEEE Transactions on Power Systems, vol. 9, no. 2, pages 988–994, 1994. (Cit  en pages 12 et 27.)
- [Fan 2006] Shu Fan et Luonan Chen. *Short-term load forecasting based on an adaptive hybrid method*. IEEE Transactions on Power Systems, vol. 21, no. 1, pages 392–401, 2006. (Cit  en page 78.)
- [Fan 2012] Shu Fan et Rob J Hyndman. *Short-term load forecasting based on a semi-parametric additive model*. IEEE Transactions on Power Systems, vol. 27, no. 1, pages 134–141, 2012. (Cit  en pages 79 et 96.)
- [Farfar 2015] Kheir Eddine Farfar, Mohamed Tarek Khadir et Oussama Laib. *Comparison of serial and parallel approaches using artificial neural networks for Algerian short-term load forecasting*. In Third International Conference on Advances in Computing, Electronics and Electrical Technology. The Institute of Research Engineers and Doctors, 2015. (Cit  en pages 67 et 96.)
- [Fay 2010] Damien Fay et John V Ringwood. *On the influence of weather forecast errors in short-term load forecasting models*. IEEE transactions on power systems, vol. 25, no. 3, pages 1751–1758, 2010. (Cit  en page 79.)
- [Feinberg 2005] Eugene A Feinberg et Dora Genethliou. Load forecasting, pages 269–285. Springer US, 2005. (Cit  en page 79.)
- [Fidalgo 2007] J Nuno Fidalgo et Manuel A Matos. *Forecasting Portugal global load with artificial neural networks*. In International Conference on Artificial Neural Networks, pages 728–737. Springer, 2007. (Cit  en pages 20 et 29.)
- [Fukumizu 2000] Kenji Fukumizu et Shun-ichi Amari. *Local minima and plateaus in hierarchical structures of multilayer perceptrons*. Neural networks, vol. 13, no. 3, pages 317–327, 2000. (Cit  en page 50.)
- [Glorot 2010] Xavier Glorot et Yoshua Bengio. *Understanding the difficulty of training deep feedforward neural networks*. In Proceedings of the thirteenth international conference on artificial intelligence and statistics, pages 249–256, 2010. (Cit  en pages 49 et 51.)
- [Glorot 2011] Xavier Glorot, Antoine Bordes et Yoshua Bengio. *Deep sparse rectifier neural networks*. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 315–323, 2011. (Cit  en page 51.)
- [Gonzalez 1993] Avelino J Gonzalez et Douglas D Dankel. The engineering of knowledge-based systems : theory and practice. Prentice-Hall, NJ, USA, 1993. (Cit  en page 78.)

- [Goodfellow 2016] Ian Goodfellow, Yoshua Bengio, Aaron Courville et Yoshua Bengio. Deep learning, volume 1. MIT press Cambridge, 2016. (Cité en pages ix, 52 et 53.)
- [Gottschalk 1994] H Gottschalk, S Heine, B Fox et I Neumann. *Economic operation of a power system with a significant amount of controllable load*. In Proceedings of 29th Universities Power Engineering Conference UPEC, University of Galway, pages 673–675, 1994. (Cité en page 25.)
- [Hagan 1987] Martin T Hagan et Suzanne M Behr. *The time series approach to short term load forecasting*. IEEE Transactions on Power Systems, vol. 2, no. 3, pages 785–791, 1987. (Cité en pages 13 et 28.)
- [Haida 1994] T. Haida et S. Muto. *Regression based peak load forecasting using a transformation technique*. IEEE Transactions on Power Systems, vol. 9, no. 4, pages 1788–1794, 1994. (Cité en pages 14 et 28.)
- [Hamid 2010] MB Abdul Hamid et TK Abdul Rahman. *Short term load forecasting using an artificial neural network trained by artificial immune system learning algorithm*. In Computer Modelling and Simulation (UKSim), 2010 12th International Conference on, pages 408–413. IEEE, 2010. (Cité en pages 21 et 29.)
- [Han 1995] Jun Han et Claudio Moraga. *The influence of the sigmoid function parameters on the speed of backpropagation learning*. From Natural to Artificial Neural Computation, pages 195–201, 1995. (Cité en page 83.)
- [Haykin 2009] Simon O Haykin. Neural networks and learning machines, volume 3. Pearson, New York, 2009. (Cité en pages 34, 35, 37, 38, 42 et 78.)
- [Hernández 2014] Luis Hernández, Carlos Baladrón, Javier M Aguiar, Belén Carro, Antonio Sánchez-Esguevillas et Jaime Lloret. *Artificial neural networks for short-term load forecasting in microgrids environment*. Energy, vol. 75, pages 252–264, 2014. (Cité en pages 21 et 29.)
- [Hinton 2006] Geoffrey E Hinton, Simon Osindero et Yee-Whye Teh. *A Fast Learning Algorithm for Deep Belief Nets*. Neural computation, vol. 18, no. 7, pages 1527–1554, 2006. (Cité en pages 44, 51, 52, 53 et 84.)
- [Hippert 2001] Henrique Steinherz Hippert, Carlos Eduardo Pedreira et Reinaldo Castro Souza. *Neural networks for short-term load forecasting : A review and evaluation*. IEEE Transactions on power systems, vol. 16, no. 1, pages 44–55, 2001. (Cité en pages 17, 74 et 78.)

- [Hippert 2004] Henrique Steinhilber Hippert et Carlos Eduardo Pedreira. *Estimating temperature profiles for short-term load forecasting : neural networks compared to linear models*. IEE Proceedings-Generation, Transmission and Distribution, vol. 151, no. 4, pages 543–547, 2004. (Cité en page 79.)
- [Hippert 2005] HS Hippert, DW Bunn et RC Souza. *Large neural networks for electricity load forecasting : Are they overfitted?* International Journal of forecasting, vol. 21, no. 3, pages 425–434, 2005. (Cité en pages 19 et 29.)
- [Ho 1990] Ku-Long Ho, Yuan-Yih Hsu, Chuan-Fu Chen, Tzong-En Lee, Chih-Chien Liang, Tsau-Shin Lai et Kung-Keng Chen. *Short term load forecasting of Taiwan power system using a knowledge-based expert system*. IEEE Transactions on Power Systems, vol. 5, no. 4, pages 1214–1221, 1990. (Cité en pages 24 et 29.)
- [Hochreiter 2001] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, Jürgen Schmidhuber et al. *Gradient flow in recurrent nets : the difficulty of learning long-term dependencies*, 2001. (Cité en page 51.)
- [Hornik 1989] Kurt Hornik, Maxwell Stinchcombe et Halbert White. *Multilayer feedforward networks are universal approximators*. Neural networks, vol. 2, no. 5, pages 359–366, 1989. (Cité en page 48.)
- [Hossain 2015] Moinul Hossain, Banafsheh Rekabdar, Sushil J Louis et Sergiu Dascalu. *Forecasting the weather of Nevada : A deep learning approach*. In Neural Networks (IJCNN), 2015 International Joint Conference on, pages 1–6. IEEE, 2015. (Cité en pages 56 et 59.)
- [Hu 2016] Qinghua Hu, Rujia Zhang et Yucan Zhou. *Transfer learning for short-term wind speed prediction with deep neural networks*. Renewable Energy, vol. 85, pages 83–95, 2016. (Cité en pages 56 et 59.)
- [Huang 1997] SR Huang. *Short-term load forecasting using threshold autoregressive models*. IEE Proceedings-Generation, Transmission and Distribution, vol. 144, no. 5, pages 477–481, 1997. (Cité en pages 11 et 27.)
- [Huang 2003] Shyh-Jier Huang et Kuang-Rong Shih. *Short-term load forecasting via ARMA model identification including non-Gaussian process considerations*. IEEE Transactions on Power Systems, vol. 18, no. 2, pages 673–679, 2003. (Cité en page 78.)
- [Hyde 1997] O Hyde et PF Hodnett. *An adaptable automated procedure for short-term electricity load forecasting*. IEEE Transactions on Power Systems, vol. 12, no. 1, pages 84–94, 1997. (Cité en pages 15 et 28.)

- [Hyndman 2006] Rob J Hyndman et Anne B Koehler. *Another look at measures of forecast accuracy*. International Journal of Forecasting, vol. 22, no. 4, pages 679–688, 2006. (Cité en page 86.)
- [Ioffe 2015] Sergey Ioffe et Christian Szegedy. *Batch normalization : Accelerating deep network training by reducing internal covariate shift*. arXiv preprint arXiv :1502.03167, 2015. (Cité en page 51.)
- [Jabbour 1988] Kamal Jabbour, Jose Fernando Vega Riveros, David Landsbergen et Walter Meyer. *ALFA : Automated load forecasting assistant*. IEEE Transactions on Power Systems, vol. 3, no. 3, pages 908–914, 1988. (Cité en pages 24 et 29.)
- [Jang 1997] Jyh-Shing Roger Jang, Chuen-Tsai Sun et Eiji Mizutani. *Neuro-fuzzy and soft computing ; a computational approach to learning and machine intelligence*. Prentice Hall, 1997. (Cité en page 25.)
- [Juberias 1999] G Juberias, R Yunta, J Garcia Moreno et C Mendivil. *A new ARIMA model for hourly load forecasting*. In Transmission and Distribution Conference, 1999 IEEE, volume 1, pages 314–319. IEEE, 1999. (Cité en pages 13 et 27.)
- [Kandel 1991] Abraham Kandel. *Fuzzy expert systems*. CRC press, 1991. (Cité en page 23.)
- [Khadir 2010] Mohamed Tarek Khadir, Farouk Benabbas et Sofiane Khdairia. *Kohonen maps combined to k-means in a two level strategy for time series clustering application to meteorological and electricity load data*. InTech, 2010. (Cité en pages 80 et 82.)
- [Khadir 2011] Mohamed Tarek Khadir et Farouk Benabbas. *Kohonen Maps Combined to Fuzzy C-means, a Two Level Clustering Approach. Application to Electricity Load Data*. In Self Organizing Maps-Applications and Novel Algorithm Design. InTech, 2011. (Cité en pages 19 et 29.)
- [Kiartzis 1997] SJ Kiartzis, CE Zoumas, JB Theocharis, AG Bakirtzis et V Petridis. *Short-term load forecasting in an autonomous power system using artificial neural networks*. IEEE Transactions on Power Systems, vol. 12, no. 4, pages 1591–1596, 1997. (Cité en pages 18 et 28.)
- [Kim 1995] Kwang-Ho Kim, Jong-Keun Park, Kab-Ju Hwang et Sung-Hak Kim. *Implementation of hybrid short-term load forecasting system using artificial neural networks and fuzzy expert systems*. IEEE Transactions on Power Systems, vol. 10, no. 3, pages 1534–1539, 1995. (Cité en pages 25 et 30.)

- [Kim 2000] Kwang-Ho Kim, Hyoung-Sun Youn et Yong-Cheol Kang. *Short-term load forecasting for special days in anomalous load conditions using neural networks and fuzzy inference method*. IEEE Transactions on Power Systems, vol. 15, no. 2, pages 559–565, 2000. (Cit  en pages 25 et 30.)
- [Kling 1985] John L Kling et David A Bessler. *A comparison of multivariate forecasting procedures for economic time series*. International Journal of Forecasting, vol. 1, no. 1, pages 5–24, 1985. (Cit  en page 8.)
- [Kriebol 1999] Stefan Kriebol. *A combined parametric and nonparametric approach to time series analysis*. Ios Press, 1999. (Cit  en pages 6 et 7.)
- [Kuremoto 2014] Takashi Kuremoto, Shinsuke Kimura, Kunikazu Kobayashi et Masanao Obayashi. *Time series forecasting using a deep belief network with restricted Boltzmann machines*. Neurocomputing, vol. 137, pages 47–56, 2014. (Cit  en page 55.)
- [Kyriakides 2007] Elias Kyriakides et Marios Polycarpou. *Short term electric load forecasting : A tutorial*. In Trends in Neural Computation, pages 391–418. Springer, 2007. (Cit  en pages 10, 14 et 78.)
- [Langkvist 2014] Martin Langkvist, Lars Karlsson et Amy Loutfi. *A review of unsupervised feature learning and deep learning for time-series modeling*. Pattern Recognition Letters, vol. 42, pages 11–24, 2014. (Cit  en page 79.)
- [Larochelle 2007] Hugo Larochelle, Dumitru Erhan, Aaron Courville, James Bergstra et Yoshua Bengio. *An empirical evaluation of deep architectures on problems with many factors of variation*. In Proceedings of the 24th International Conference on Machine Learning, pages 473–480. ACM, 2007. (Cit  en pages 43 et 84.)
- [LeCun 1993] Yann LeCun. *Efficient learning and second-order methods*. A tutorial at NIPS, vol. 93, page 61, 1993. (Cit  en page 42.)
- [Li 2016] Song Li, Peng Wang et Lalit Goel. *A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection*. IEEE Transactions on power systems, vol. 31, no. 3, pages 1788–1798, 2016. (Cit  en pages 22 et 29.)
- [Li 2018] Fei Li, Jie Zhang, Chao Shang, Dexian Huang, Eni Oko et Meihong Wang. *Modelling of a post-combustion CO2 capture process using deep belief network*. Applied Thermal Engineering, vol. 130, pages 997–1003, 2018. (Cit  en page 57.)

- [Linville 1990] Dale E Linville. *Calculating chilling hours and chill units from daily maximum and minimum temperature observations*. HortScience, vol. 25, no. 1, pages 14–16, 1990. (Cité en page 83.)
- [Liu 2014] James NK Liu, Yanxing Hu, Jane Jia You et Pak Wai Chan. *Deep neural network based feature representation for weather forecasting*. In Proceedings on the International Conference on Artificial Intelligence (ICAI), page 1. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2014. (Cité en pages 56 et 59.)
- [Liu 2017] Fan Liu, Feng Xu et Sai Yang. *A flood forecasting model based on deep learning algorithm via integrating stacked autoencoders with BP neural network*. In Multimedia Big Data (BigMM), 2017 IEEE Third International Conference on, pages 58–61. IEEE, 2017. (Cité en page 57.)
- [Lloyd 1982] Stuart P Lloyd. *Least squares quantization in PCM*. IEEE Transactions on Information Theory, vol. 28, no. 2, pages 129–137, 1982. (Cité en page 81.)
- [Lu 2017] Chen Lu, Zhen-Ya Wang, Wei-Li Qin et Jian Ma. *Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification*. Signal Processing, vol. 130, pages 377–388, 2017. (Cité en page 58.)
- [Lv 2017] Feiya Lv, Chenglin Wen, Meiqin Liu et Zhejing Bao. *Weighted time series fault diagnosis based on a stacked sparse autoencoder*. Journal of Chemometrics, vol. 31, no. 9, 2017. (Cité en page 57.)
- [Mamlook 2009] Rustum Mamlook, Omar Badran et Emad Abdulhadi. *A fuzzy inference model for short-term load forecasting*. Energy Policy, vol. 37, no. 4, pages 1239–1248, 2009. (Cité en page 78.)
- [Mbamalu 1993] GAN Mbamalu et ME El-Hawary. *Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation*. IEEE Transactions on Power Systems, vol. 8, no. 1, pages 343–348, 1993. (Cité en pages 11 et 27.)
- [Mena 2014] R Mena, F Rodriguez, M Castilla et Manuel R Arahal. *A prediction model based on neural networks for the energy consumption of a bioclimatic building*. Energy and Buildings, vol. 82, pages 142–155, 2014. (Cité en page 21.)
- [Merritt 2012] Dennis Merritt. Building expert systems in prolog. Springer Science & Business Media, 2012. (Cité en page 23.)

- [Moghram 1989] Ibrahim Moghram et Saifur Rahman. *Analysis and evaluation of five short-term load forecasting techniques*. IEEE Transactions on power systems, vol. 4, no. 4, pages 1484–1491, 1989. (Cité en pages 16 et 17.)
- [Mohamad 1996] EA Mohamad, MM Mansour, S El-Debeiky, KG Mohamad, ND Rao et G Ramakrishna. *Results of Egyptian unified grid hourly load forecasting using an artificial neural network with expert system interface*. Electric power systems research, vol. 39, no. 3, pages 171–177, 1996. (Cité en pages 24 et 30.)
- [Naud 1995] Colette Naud. *Introduction-la conservation préventive dans les musées*. Université du Québec à Montréal, 1995. (Cité en page 72.)
- [Nezzar 2016] Reda Mohamed Nezzar, Nadir Farah, Mohamed Tarek Khadir et Lakhdar Chouireb. *Mid-Long Term Load Forecasting using Multi-Model Artificial Neural Networks*. International Journal on Electrical Engineering and Informatics, vol. 8, no. 2, page 389, 2016. (Cité en pages 13, 28 et 84.)
- [Nie 2017] Laisen Nie, Dingde Jiang, Shui Yu et Houbing Song. *Network traffic prediction based on deep belief network in wireless mesh backbone networks*. In Wireless Communications and Networking Conference (WCNC), 2017 IEEE, pages 1–5. IEEE, 2017. (Cité en page 57.)
- [Osczevski 2005] Randall Osczevski et Maurice Bluestein. *The new wind chill equivalent temperature chart*. Bulletin of the American Meteorological Society, vol. 86, no. 10, pages 1453–1458, 2005. (Cité en page 73.)
- [Palit 2006] Ajoy K Palit et Dobrivoje Popovic. *Computational intelligence in time series forecasting : theory and engineering applications*. Springer Science & Business Media, 2006. (Cité en pages 7 et 8.)
- [Papalexopoulos 1990] Alex D Papalexopoulos et Timothy C Hesterberg. *A regression-based approach to short-term system load forecasting*. IEEE Transactions on Power Systems, vol. 5, no. 4, pages 1535–1547, 1990. (Cité en pages 15 et 28.)
- [Papalexopoulos 1994] Alex D Papalexopoulos, Shangyou Hao et Tie-Mao Peng. *An implementation of a neural network based load forecasting model for the EMS*. IEEE transactions on Power Systems, vol. 9, no. 4, pages 1956–1962, 1994. (Cité en pages 16 et 17.)
- [Park 1991] JH Park, YM Park et KY Lee. *Composite modeling for adaptive short-term load forecasting*. IEEE Transactions on Power Systems, vol. 6, no. 2, pages 450–457, 1991. (Cité en pages 16 et 28.)

- [Peng 1992] TM Peng, NF Hubele et GG Karady. *Advancement in the application of neural networks for short-term load forecasting*. IEEE Transactions on Power Systems, vol. 7, no. 1, pages 250–257, 1992. (Cité en pages 17 et 28.)
- [Perez 2008] LP Perez et al. *Application of support vector machines and ANFIS to the short-term load forecasting*. In Transmission and Distribution Conference and Exposition : Latin America, 2008 IEEE/PES, pages 1–5. IEEE, 2008. (Cité en pages 27 et 30.)
- [Pidwirny 2006] Michael Pidwirny. *Fundamentals of Physical Geography*, 2006. Accessed January 2017. (Cité en page 90.)
- [Platon 2015] Radu Platon, Vahid Raissi Dehkordi et Jacques Martel. *Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis*. Energy and Buildings, vol. 92, pages 10–18, 2015. (Cité en page 21.)
- [Pontius 2008] Robert Gilmore Pontius, Olufunmilayo Thontteh et Hao Chen. *Components of information for multiple resolution comparison between maps that share a real variable*. Environmental and Ecological Statistics, vol. 15, no. 2, pages 111–142, 2008. (Cité en page 75.)
- [Qin 2017] Mengjiao Qin, Zhihang Li et Zhenhong Du. *Red tide time series forecasting by combining ARIMA and deep belief network*. Knowledge-Based Systems, vol. 125, pages 39–52, 2017. (Cité en page 56.)
- [Qiu 2017] Xueheng Qiu, Ye Ren, Ponnuthurai Nagarathnam Suganthan et Gehan AJ Amaratunga. *Empirical mode decomposition based ensemble deep learning for load demand time series forecasting*. Applied Soft Computing, vol. 54, pages 246–255, 2017. (Cité en pages 58 et 59.)
- [Rahman 1988] Saifur Rahman et Rahul Bhatnagar. *An expert system based algorithm for short term load forecast*. IEEE Transactions on Power Systems, vol. 3, no. 2, pages 392–399, 1988. (Cité en pages 23, 29 et 78.)
- [Rahman 1993] Saifur Rahman et Ossama Hazim. *A generalized knowledge-based short-term load-forecasting technique*. IEEE Transactions on power systems, vol. 8, no. 2, pages 508–514, 1993. (Cité en pages 23 et 29.)
- [Rahman 1996] S Rahman et O Hazim. *Load forecasting for multiple sites : development of an expert system-based technique*. Electric Power Systems Research, vol. 39, no. 3, pages 161–169, 1996. (Cité en pages 23 et 29.)
- [Ramanathan 1997] Ramu Ramanathan, Robert Engle, Clive WJ Granger, Farshid Vahid-Araghi et Casey Brace. *Short-run forecasts of electricity loads and*

- peaks*. International Journal of Forecasting, vol. 13, no. 2, pages 161–174, 1997. (Cité en pages 15, 28 et 78.)
- [Ranaweera 1995] DK Ranaweera, NF Hubele et AD Papalexopoulos. *Application of radial basis function neural network model for short-term load forecasting*. IEE Proceedings-Generation, Transmission and Distribution, vol. 142, no. 1, pages 45–50, 1995. (Cité en pages 19 et 29.)
- [Ranaweera 1996] Damitha K Ranaweera, George G Karady et Richard G Farmer. *Effect of probabilistic inputs on neural network-based electric load forecasting*. IEEE Transactions on Neural networks, vol. 7, no. 6, pages 1528–1532, 1996. (Cité en pages 18 et 28.)
- [Ringwood 2001] John V Ringwood, D Bofelli et Fiona T Murray. *Forecasting electricity demand on short, medium and long time scales using neural networks*. Journal of Intelligent and Robotic Systems, vol. 31, no. 1-3, pages 129–147, 2001. (Cité en pages 20 et 29.)
- [Romeu 2013] Pablo Romeu, Francisco Zamora-Martínez, Paloma Botella-Rocamora et Juan Pardo. *Time-series forecasting of indoor temperature using pre-trained deep neural networks*. In International Conference on Artificial Neural Networks, pages 451–458. Springer, 2013. (Cité en pages 55, 58 et 79.)
- [Romeu 2015] Pablo Romeu, Francisco Zamora-Martínez, Paloma Botella-Rocamora et Juan Pardo. *Stacked Denoising Auto-Encoders for Short-Term Time Series Forecasting*. In Artificial Neural Networks, pages 463–486. Springer, 2015. (Cité en pages 55, 59 et 79.)
- [Rumelhart 1986] David E Rumelhart, Geoffrey E Hinton et Ronald J Williams. *Learning representations by back-propagating errors*. nature, vol. 323, no. 6088, page 533, 1986. (Cité en page 39.)
- [Ruzic 2003] Slobodan Ruzic, Aca Vuckovic et Nikola Nikolic. *Weather sensitive method for short term load forecasting in electric power utility of Serbia*. IEEE Transactions on Power Systems, vol. 18, no. 4, pages 1581–1586, 2003. (Cité en pages 15 et 28.)
- [Sargunaraaj 1997] S Sargunaraaj, DP Sen Gupta et S Devi. *Short-term load forecasting for demand side management*. IEE Proceedings-Generation, Transmission and Distribution, vol. 144, no. 1, pages 68–74, 1997. (Cité en pages 16, 28 et 78.)
- [Senabre 2010] C Senabre, S Valero et J Aparicio. Using a self organizing map neural network for short-term load forecasting, analysis of different input

- data patterns, pages 397–400. Springer Berlin Heidelberg, 2010. (Cité en page 78.)
- [Senjyu 2002] Tomonobu Senjyu, Hitoshi Takara, Katsumi Uezato et Toshihisa Funabashi. *One-hour-ahead load forecasting using neural network*. IEEE Transactions on power systems, vol. 17, no. 1, pages 113–118, 2002. (Cité en pages 18 et 29.)
- [Senjyu 2004] T Senjyu, P Mandal, K Uezato et T Funabashi. *Next day load curve forecasting using recurrent neural network structure*. IEE Proceedings-Generation, Transmission and Distribution, vol. 151, no. 3, pages 388–394, 2004. (Cité en pages 20 et 29.)
- [Shao 2018] Haidong Shao, Hongkai Jiang, Xingqiu Li et Tianchen Liang. *Rolling bearing fault detection using continuous deep belief network with locally linear embedding*. Computers in Industry, vol. 96, pages 27–39, 2018. (Cité en page 56.)
- [Shen 2016] Hua Shen et Xun Liang. *A Time Series Forecasting Model Based on Deep Learning Integrated Algorithm with Stacked Autoencoders and SVR for FX Prediction*. In International Conference on Artificial Neural Networks, pages 326–335. Springer, 2016. (Cité en page 58.)
- [Shensa 1992] Mark J Shensa. *The discrete wavelet transform : wedding the a trous and Mallat algorithms*. IEEE Transactions on signal processing, vol. 40, no. 10, pages 2464–2482, 1992. (Cité en page 78.)
- [Snoek 2012] Jasper Snoek, Hugo Larochelle et Ryan P Adams. *Practical bayesian optimization of machine learning algorithms*. In Advances in neural information processing systems, pages 2951–2959, 2012. (Cité en page 44.)
- [Soliman 1997] SA Soliman, S Persaud, K El-Nagar et ME El-Hawary. *Application of least absolute value parameter estimation based on linear programming to short-term load forecasting*. International Journal of Electrical Power & Energy Systems, vol. 19, no. 3, pages 209–216, 1997. (Cité en pages 13 et 28.)
- [Soliman 2010] Soliman Abdel-hady Soliman et Ahmad Mohammad Al-Kandari. *Electrical load forecasting : modeling and model construction*. Elsevier, 2010. (Cité en pages 9, 11 et 12.)
- [Song 2005] Kyung-Bin Song, Young-Sik Baek, Dug Hun Hong et Gilsoo Jang. *Short-term load forecasting for the holidays using fuzzy linear regression method*. IEEE transactions on power systems, vol. 20, no. 1, pages 96–101, 2005. (Cité en page 79.)

- [Spencer 1971] JW Spencer. *Fourier series representation of the position of the sun*. Search, vol. 2, no. 5, pages 172–172, 1971. (Cit  en pages 83 et 110.)
- [Srinivasan 1994] D Srinivasan, AC Liew et CS Chang. *Forecasting daily load curves using a hybrid fuzzy-neural approach*. IEE Proceedings-Generation, Transmission and Distribution, vol. 141, no. 6, pages 561–567, 1994. (Cit  en pages 25 et 30.)
- [Srinivasan 1995] Dipti Srinivasan et MA Lee. *Survey of hybrid fuzzy neural approaches to electric load forecasting*. In Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on, volume 5, pages 4004–4008. IEEE, 1995. (Cit  en page 25.)
- [Srivastava 2014] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever et Ruslan Salakhutdinov. *Dropout : A simple way to prevent neural networks from overfitting*. The Journal of Machine Learning Research, vol. 15, no. 1, pages 1929–1958, 2014. (Cit  en page 52.)
- [Steadman 1994] Robert G Steadman. *Norms of apparent temperature in Australia*. Australian Meteorological Magazine, vol. 43, no. 1, pages 1–16, 1994. (Cit  en page 73.)
- [Strobelt 2018] Hendrik Strobelt, Sebastian Gehrmann, Hanspeter Pfister et Alexander M Rush. *Lstmvis : A tool for visual analysis of hidden state dynamics in recurrent neural networks*. IEEE transactions on visualization and computer graphics, vol. 24, no. 1, pages 667–676, 2018. (Cit  en page 49.)
- [Taylor 2000] James W Taylor et Shanti Majithia. *Using combined forecasts with changing weights for electricity demand profiling*. Journal of the Operational Research Society, vol. 51, no. 1, pages 72–82, 2000. (Cit  en page 79.)
- [Taylor 2002] James W Taylor et Roberto Buizza. *Neural network load forecasting with weather ensemble predictions*. IEEE Transactions on Power systems, vol. 17, no. 3, pages 626–632, 2002. (Cit  en pages 19 et 29.)
- [Tian 2009] Shu Tian et Liu Tuanjie. *Short-term load forecasting based on RBFNN and QPSO*. In Power and Energy Engineering Conference, 2009. APPEEC 2009. Asia-Pacific, pages 1–4. IEEE, 2009. (Cit  en pages 20 et 29.)
- [Tran 2006] Chung Nguyen Tran, Dong-Chul Park et Hwan-Soo Choi. *Short-term load forecasting using multiscale bilinear recurrent neural network with an adaptive learning algorithm*. In International Conference on Neural Information Processing, pages 964–973. Springer, 2006. (Cit  en pages 20 et 29.)

- [Trudnowski 2001] Dan J Trudnowski, Warren L McReynolds et Jeffery M Johnson. *Real-time very short-term load prediction for power-system automatic generation control*. IEEE Transactions on Control Systems Technology, vol. 9, no. 2, pages 254–260, 2001. (Cité en pages 16 et 28.)
- [Tzafestas 2001] Spyros Tzafestas et Elpida Tzafestas. *Computational Intelligence Techniques for Short-Term Electric Load Forecasting*. Journal of Intelligent and Robotic Systems, vol. 31, no. 1-3, pages 7–68, 2001. (Cité en page 79.)
- [Vapnik 1995] Vladimir N Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995. (Cité en page 78.)
- [Vincent 2010] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio et Pierre-Antoine Manzagol. *Stacked denoising autoencoders : Learning useful representations in a deep network with a local denoising criterion*. Journal of Machine Learning Research, vol. 11, pages 3371–3408, 2010. (Cité en pages 47, 53, 79, 85 et 95.)
- [Wang 2017] Long Wang, Zijun Zhang et Jieqiu Chen. *Short-Term Electricity Price Forecasting with Stacked Denoising Autoencoders*. IEEE Transactions on Power Systems, vol. 32, no. 4, pages 2673–2681, 2017. (Cité en pages 58 et 59.)
- [Weron 2007] Rafal Weron. *Modeling and forecasting electricity loads and prices : A statistical approach*, volume 403. John Wiley & Sons, 2007. (Cité en page 14.)
- [Wikipedia 2018a] Wikipedia. *Orbite de la Terre — Wikipedia, The Free Encyclopedia*, 2018. [Online ; accessed 21-March-2018]. (Cité en pages x et 109.)
- [Wikipedia 2018b] Wikipedia. *Time series — Wikipedia, The Free Encyclopedia*, 2018. [Online ; accessed 15-March-2018]. (Cité en page 8.)
- [Wood 2012] Allen J Wood et Bruce F Wollenberg. *Power generation, operation, and control*. John Wiley & Sons, 2012. (Cité en pages xi, 9 et 10.)
- [Ying 2008] Li-Chih Ying et Mei-Chiu Pan. *Using adaptive network based fuzzy inference system to forecast regional electricity loads*. Energy Conversion and Management, vol. 49, no. 2, pages 205–211, 2008. (Cité en pages 26 et 30.)
- [Yoo 1999] Hyeonjoong Yoo et Russell L Pimmel. *Short term load forecasting using a self-supervised adaptive neural network*. IEEE transactions on Power Systems, vol. 14, no. 2, pages 779–784, 1999. (Cité en pages 18 et 29.)

- [Yuancheng 2003] Li Yuancheng, Fang Tingjian et Yu Erkeng. *Study of support vector machines for short-term load forecasting*. Proceedings of the CSEE, vol. 23, no. 6, pages 55–59, 2003. (Cité en page 78.)
- [Zarchan 2005] Paul Zarchan et Howard Musoff. *Fundamentals of kalman filtering : A practical approach*, volume 208. American Institute of Aeronautics and Astronautics, 2005. (Cité en page 78.)
- [Zhang 2017] Weishan Zhang, Pengcheng Duan, Laurence T Yang, Feng Xia, Zhongwei Li, Qinghua Lu, Wenjuan Gong et Su Yang. *Resource requests prediction in the cloud computing environment with a deep belief network*. Software : Practice and Experience, vol. 47, no. 3, pages 473–488, 2017. (Cité en page 57.)