



BIBLIOTHÈQUE

CÉGEP DE L'ABITIBI-TÉMISCAMINGUE
UNIVERSITÉ DU QUÉBEC EN ABITIBI-TÉMISCAMINGUE

Mise en garde

La bibliothèque du Cégep de l'Abitibi-Témiscamingue et de l'Université du Québec en Abitibi-Témiscamingue (UQAT) a obtenu l'autorisation de l'auteur de ce document afin de diffuser, dans un but non lucratif, une copie de son œuvre dans [Depositum](#), site d'archives numériques, gratuit et accessible à tous. L'auteur conserve néanmoins ses droits de propriété intellectuelle, dont son droit d'auteur, sur cette œuvre.

Warning

The library of the Cégep de l'Abitibi-Témiscamingue and the Université du Québec en Abitibi-Témiscamingue (UQAT) obtained the permission of the author to use a copy of this document for nonprofit purposes in order to put it in the open archives [Depositum](#), which is free and accessible to all. The author retains ownership of the copyright on this document.



**GESTION ÉNERGÉTIQUE INTELLIGENTE POUR LA
PRÉVISION D'ÉNERGIE PAR L'APPRENTISSAGE MACHINE
ET L'INTELLIGENCE ARTIFICIELLE**

**MÉMOIRE
PRÉSENTÉ COMME EXIGENCE PARTIELLE
DE LA MAITRISE EN INGÉNIERIE
GÉNIE ÉLECTRIQUE**

**PAR
OTHMANE KINANI**

OCTOBRE 2024

Remerciements

Je tiens à exprimer ma profonde gratitude à mon directeur de recherche, le Professeur **Fouad Slaoui Hasnaoui** de l'Université du Québec en Abitibi-Témiscamingue.

Tout au long de ce projet, il a su me guider avec une grande expertise et un sens aigu de la pédagogie. Sa confiance en mes capacités m'a permis de me dépasser et de relever les défis qui se sont présentés. Son encadrement rigoureux, son soutien constant, et sa disponibilité ont été des piliers essentiels dans la réalisation de ce travail.

Je suis particulièrement reconnaissant pour ses précieux conseils, sa patience, et pour le partage de ses vastes connaissances, qui ont grandement enrichi mon parcours académique et personnel.

Je souhaite également remercier ma famille et mes amis pour leur soutien indéfectible et leurs encouragements tout au long de cette période d'étude. Leur présence à mes côtés a été une source inestimable de motivation et de réconfort.

Résumé

Ce projet se concentre sur l'analyse prédictive en utilisant différentes techniques de modélisation, notamment les **modèles de base**, les **modèles hybrides**, et la **fusion tri-modèle**. L'objectif est d'améliorer la précision des prévisions dans des contextes complexes, tels que l'analyse des séries chronologiques et d'autres données temporelles. Les **modèles de base**, tels que **ARIMA**, sont souvent utilisés pour des données avec des relations linéaires simples, mais présentent des limites pour des dynamiques plus complexes. Les **modèles hybrides**, comme **SARIMA-SVR**, combinent plusieurs approches pour capturer à la fois les relations linéaires et non linéaires. La **fusion tri-modèle** est une technique plus avancée qui combine trois types de modèles, comme **SARIMA**, **SVM**, et **Random Forest**, pour maximiser la robustesse et la précision des prévisions. Ce projet insiste également sur la **préparation des données**, en particulier la gestion des données manquantes, la normalisation et l'analyse des tendances. Cette combinaison de techniques avancées et d'une préparation rigoureuse des données vise à fournir des prévisions plus précises et à mieux comprendre les dynamiques sous-jacentes des données étudiées.

Mots-clés

Saisonnière autorégressive intégrée à moyenne mobile (SARIMA), machines à vecteurs de support (SVM), Forêt aléatoire, Modèle prédictif, Modelé hybride.

Abstract

This project focuses on predictive analytics using different modeling techniques, including baseline models, hybrid models, and tri-model fusion. The goal is to improve forecast accuracy in complex contexts, such as analyzing time series and other temporal data. Baseline models, such as ARIMA, are often used for data with simple linear relationships, but have limitations for more complex dynamics. Hybrid models, such as SARIMA-SVR, combine multiple approaches to capture both linear and nonlinear relationships. Tri-model fusion is a more advanced technique that combines three types of models, SARIMA, SVM, and Random Forest, to maximize forecast robustness and accuracy. This project also emphasizes data preparation, particularly handling missing data, normalization, and trend analysis. This combination of advanced techniques and rigorous data preparation aims to provide more accurate forecasts and better understand the underlying dynamics of the data studied.

Keywords

Seasonal AutoRegressive Integrated Moving Average (SARIMA), Support Vector Machine (SVM), Random Forest, Predictive model, Hybrid model.

TABLE DES MATIERE

Remerciements	ii
Résumé	iii
Abstract	iv
CHAPITRE 1 : INTRODUCTION	1
Introduction :	1
1. Contexte Énergétique du Canada :	1
1.1. Exploration des Ressources Énergétiques dans les Provinces Canadiennes :.....	2
1.2. Tendances de Consommation et Impacts :	5
1.3. Enjeux et Perspectives:.....	5
2. Importance de la Prévision Énergétique :	5
2.1. Les Multiples Facettes de la Gestion Énergétique :	6
2.2. L'Impact des Prévisions sur la Politique Énergétique :	6
3. Présentation des Modèles ARIMA et ANN :	6
3.1. Fondements Théoriques des Modèles ARIMA :	7
3.2. Potentiel des Réseaux de Neurones Artificiels (ANN) :	7
3.3. Comparaison et Contraste des Deux Modèles :	7
4. L'Émergence des Données En temps réel:	8
4.1. Révolution des Données En temps réel :	8
4.2. Impact sur la Prévision Énergétique :	8
4.3. Potentiel des Modèles Hybrides avec Données En temps réel :	9
5. La Synergie des Modèles Hybrides avec l'Intégration des Données En temps réel:....	9
5.1. L'Intégration Innovante d'ARIMA et ANN :	9
5.2. Avantages de l'Approche Hybride :	10
5.3. Révolution dans la Prévision Énergétique :	10

5.4.	Un Pas Vers l'Avenir : Combinaison des Modèles ARIMA, ANN:.....	10
5.4.1.	Intégration des Données en temps réel dans les Modèles ARIMA et ANN :.....	11
5.4.2.	Amélioration de la Précision des Prévisions :.....	11
5.4.3.	Applications Pratiques et Scénarios d'Utilisation :	11
5.4.4.	Défis et Solutions dans l'Intégration des Données en temps réel :..	12
5.4.5.	Perspectives avec les Modèles Hybrides :	12
6.	Échanges Énergétiques Transfrontaliers :	13
6.1.	La Dynamique des Échanges Énergétiques avec les États-Unis :	13
6.2.	Implications Économiques et Politiques :.....	13
7.	Objectifs et Structure du Projet :	13
7.1.	Déclaration des Objectifs de Recherche :.....	14
7.2.	Aperçu des Chapitres Suivants :	14
8.	Hypothèses de l'Étude :	14
8.1.	Hypothèse sur l'Évolution des Modèles ARIMA et ANN :	15
8.2.	Impact des Données en temps réel sur la Précision des Prévisions :.....	15
8.3.	Influence des Politiques Énergétiques et Environnementales :	15
8.4.	Prévision Énergétique et Transition Vers les Énergies Renouvelables :	15
8.5.	Rôle des Avancées Technologiques dans la Modélisation Énergétique :	16
8.6.	Globalisation et Interdépendances Énergétiques :	16
8.7.	Défis et Opportunités de l'Électrification :	16
8.8.	Répercussions du Changement Climatique sur la Demande Énergétique :	16
8.9.	Interaction entre la Politique Énergétique et la Prévision :	17
	Conclusion :.....	18
	CHAPITRE 2 : REVUE DE LITTÉRATURE - HISTOIRE ET ÉVOLUTION DES MÉTHODES DE.....	
	PRÉVISION ÉNERGÉTIQUE	19

Introduction :	19
1. Les Racines de la Prédiction Énergétique :	20
2. Formulation des Premières Théories et Modèles :	26
2.1. Théories Fondamentales :	26
2.2. Développement des Théories Statistiques:	28
2.2.1. L'Ère Pionnière (Fin du 19ème - Début du 20ème Siècle) :	28
2.2.2. L'Innovation des Années 1970 - Les Modèles ARIMA :	30
2.2.3. Les années 1970 et 1980 - Modélisation Numérique et Algorithmique :	31
2.2.4. Les années 1980 au-delà - Les Modèles de Prédiction à l'Ère du Numérique, de l'Intelligence Artificielle et des Big Data :	32
Conclusion :	34
CHAPITRE 3 : ANALYSE DU PAYSAGE ÉNERGÉTIQUE CANADIEN : POLITIQUE, RELATION INTERNATIONALES.....	36
1. Panorama de l'énergie au Canada : Sources, exploitation et perspectives :	36
1.1. Politiques gouvernementales en matière d'énergie :	38
1.2. Transition énergétique : Stratégies pour une économie à faible émission de carbone :	40
2. Modèles de consommation d'énergie :	41
3. Infrastructure énergétique :	44
3.1. Réseau électrique : Interconnexions, fiabilité et défis :	44
3.2. Pipelines et transport du pétrole et du gaz : Principaux corridors, enjeux politiques et environnementaux :	45
3.3. Stockage d'énergie et technologies émergentes :	45
4. Relation Internationales et commerce énergétique :	46
4.1. Le Canada et les USA : Interdépendance énergétique et impact de l'ALENA :.	46
4.2. Exportations d'énergie : Marchés clés, défis logistiques, et perspectives :	47

4.3.	Coopération internationale : Participation du Canada aux forums.....	
	énergétiques mondiaux et initiatives multilatérales :.....	48
5.	Défis et Opportunités :.....	49
5.1.	Enjeux environnementaux :	49
5.2.	Innovations technologiques :.....	49
5.3.	Perspective :	50
	Conclusion :.....	50
CHAPITRE 4 : DE L'ANALYSE DES SÉRIES CHRONOLOGIQUES VERS LA PRÉVISION PAR.....		
	L'APPRENTISSAGE MACHINE.....	51
	Introduction :	51
1.	Données Énergétiques :	53
2.	Modes de Production d'Énergie Électrique:	53
3.	Analyse des Séries Chronologiques :	55
3.1.	Collecte et visualisation des données :	55
3.2.	Décomposition de la série :.....	59
3.2.1.	Estimation de la Tendance :	60
3.2.2.	Les Cycles :.....	60
3.2.3.	Estimation de la Saisonnalité :	61
3.2.1.	Estimation de Résidu:	61
3.3.	Introduction du SARIMA, un outil d'analyse et de prévision des séries.....	
	chronologiques :.....	61
3.4.	Présentation du travail fait sur MATLAB pour l'application du.....	
	modèle SARIMA :	64
	Conclusion :.....	72
CHAPITRE 5 : SYNERGIE DES TECHNIQUES : LES MODÈLES DE BASE, HYBRIDE ET.....		
	FUSION TRI-MODÈLE POUR L'ANALYSE PRÉDICTIVE	73
	Introduction :	73

1.	Explication des nominalisations:.....	74
2.1.	Modèles de base :	74
1.1.1.	Caractéristiques des Modèles de Base :	74
2.2.	Modèles Hybrides :	75
1.2.1.	Caractéristiques des Modèles Hybrides :	75
1.2.2.	Exemples de Modèles Hybrides :	75
2.3.	Fusion Tri-Modèle :.....	76
1.3.1.	Caractéristiques de la Fusion Tri-Modèle :	76
1.3.2.	Exemples de Fusion Tri-Modèle :	77
2.4.	Conclusion	78
2.	Présentation des données :.....	78
2.1.	Description des Données :.....	78
2.2.	Préparation des Données :	79
2.2.1.	Gestion des Données Manquantes :	79
2.2.2.	Normalisation des Données :	80
2.2.3.	Analyse des Tendances et Saisonnalité :	80
2.2.4.	Division des Données :	80
2.2.5.	Visualisation des Données :	80
2.2.6.	Réduction des Données dans les études théoriques :	
	Simplicité et Compréhension :	81
2.3.	Objectif de la Réduction des Données :	82
2.3.1.	Simplicité de Présentation :	82
2.3.2.	Faciliter la Compréhension :	82
2.3.3.	Validation Efficace des Modèles :	82
2.3.4.	Concentration sur les Événements Récents :	82
3.	Modèles Linéaires :	84
2.1.	Moyenne Mobile Intégrée Autorégressive -ARIMA- :	84

2.2.	Introduction de ARIMA Saisonnière -SARIMA- :.....	91
4.	Modèles non-Linéaires :.....	97
2.1.	Machine à Vecteurs de Support -SVM- :.....	97
4.1.1.	SVM :.....	99
4.1.2.	Histoire et Évolution de la SVM :.....	99
4.1.3.	Fondements Mathématiques de la SVM :.....	100
4.1.4.	Fonctions de Noyau :.....	101
4.1.5.	Entraînement du Modèle SVM :.....	101
4.1.6.	Calcul SVM :.....	102
4.1.7.	Calculer $X^T y$:.....	103
4.1.8.	Prévoir et Évaluer :.....	104
4.1.9.	Observations Clés :.....	105
4.1.10.	Interprétation des Résultats :.....	106
2.2.	Modèle SVM Amélioré :.....	107
4.2.1.	Équation du Noyau Polynomial:.....	108
4.2.2.	Définition des Hyperparamètres :.....	108
4.2.3.	Évaluation par Validation Croisée:.....	109
4.2.4.	Équation de Prédiction SVM:.....	109
4.2.5.	Équation d'Erreur:.....	110
4.2.6.	Interprétation des Résultats du SVM Amélioré par Noyau Polynomial :.....	111
4.3.	Forêt aléatoire (Random Forest) -RF- :.....	114
4.3.1.	Modélisation avec la Forêt Aléatoire :.....	115
4.3.2.	Exemple de Calcul Manuel :.....	116
4.3.3.	Données d'Exemple :.....	117
4.3.4.	Calcul des Prédictions Manuelles :.....	118
4.3.5.	Résultats sur le logiciel MATLAB :.....	121

5. Modèles Hybrides :	124
2.3. Modèle Hybride -SARIMA-SVR- (approche en Cascade) :	126
4.3.1. SARIMA :	127
4.3.2. SVR :	127
4.3.3. Fusion Tri-Modèle -SARIMA-SVM-RF-(Meta-Apprentissage):. .	133
4.3.4. Étape 1 : Prétraitement des Données :	133
4.3.5. Étape 2 : Modèle SARIMA :	135
4.3.6. Étape 3 : Modèle SVM :	136
4.3.7. Étape 4 : Modèle Foret aléatoires (Random Forest) :	137
4.3.8. ÉTAPE 5 : Prédiction Final :	138
4.3.9. ÉTAPE 6 : Dénormalisation des prédictions finales :	138
4.3.10. ÉTAPE FINAL : Interprétation Des Résultats Des Études	
Pratique Sur Le Logiciel MATLAB :	138
Conclusion:.....	152
Conclusion générale:.....	154
Bibliographie:	156

LISTE DES TABLEAUX

Tableau 1	Vue plus complète de l'évolution de la prévision énergétique.....	25
Tableau 2	Aperçu des développements mathématiques et statistiques qui ont joué un rôle clé dans l'évolution de la prévision énergétique	34
Tableau 3	Données real d'énergie électrique canadien - voir le fichier de..... programme pour l'ensemble des donnée-.....	79
Tableau 4	Données Réduites pour la compression du chapitre.....	83
Tableau 5	Les Coefficient de AR-I-MA.....	86
Tableau 6	Calcule de la Différenciation.....	86
Tableau 7	Coefficients du modèle SARIMA.....	94
Tableau 8	Données métrologiques et énergétiques de la province du..... Québec 2022-2023.....	98
Tableau 9	Donnée utiliser dans le model RF-voir le fichier de programme pour..... l'ensemble des donnée-.....	115
Tableau 10	Donnée d'exemple RF.....	117

LISTE DES FIGURES

Figure 1	Évolution de la prévision Énergétique.....	24
Figure 2	Andrey Kolmogorov 1903-1987.....	26
Figure 3	Norbert Wiener 1894-1964.....	26
Figure 4	Karl Pearson 1857-1936.....	29
Figure 5	Ronald Fisher 1890-196.....	29
Figure 6	George Box 1919-2013.....	30
Figure 7	Gwilym Jenkins 1932-1982.....	30
Figure 8	Dans la région du Saguenay-Lac-St-Jean au Québec [43].....	37
Figure 9	Comparaison des scénarios pour la production d'électricité..... (Pétajoule) [45].....	39
Figure 10	Demande d'électricité par secteur, scénario d'évolution..... des politiques [7].....	42
Figure 11	Production d'électricité par source, scénario d'évolution..... des politiques [7].....	43
Figure 12	Ajouts cumulatifs de capacités jusqu'en 2050, tous scénarios..... d'électricité net-zéro [7].....	43
Figure 13	Lignes de transport transfrontalières du Canada, 2022 [8].....	44
Figure 14	Répartition du commerce de l'énergie au Canada [53].....	46
Figure 15	Flux d'Énergie en Amérique du Nord [54].....	47
Figure 16	Valeur (Billion-USD) et volumes (Méga barils équivalents pétrole /j)..... du commerce énergétique entre les États-Unis et le Canada [53].....	48
Figure 17	Les provinces du Canada.....	51
Figure 18	a- Les ressources énergétiques Canadiennes. b- Répartition..... de la population Canadienne.....	52
Figure 19	Énergie électrique totale produite par chaque Zone.....	56
Figure 20	Énergie totale reçue par chaque Zone.....	57
Figure 21	Énergie totale livrée par chaque Zone.....	58

Figure 22 Énergie totale disponible pour une utilisation dans une limite géographique spécifique	59
Figure 23 Flux gramme du cycle itératif du Modèle ARIMA saisonnière	63
Figure 24 Prévision de la Production totale d'Énergie électrique par SARIMA - Zone 6 Québec	64
Figure 25 Prévision par SARIMA l'énergie Totale disponible - Zone 6	65
Figure 26 Prévision par SARIMA l'énergie totale reçus - Zone 6	66
Figure 27 Prévision par SARIMA l'Énergie totale délivrée - Zone 6	67
Figure 28 Prévision de l'énergie Hydraulique par le modèle SARIMA – Zone 6	68
Figure 29 Prévision de l'énergie Turbine à Vapeur classique par le modèle SARIMA Zone 6	69
Figure 30 Prévision de l'énergie Solaire par le modèle SARIMA – Zone 6	70
Figure 31 Prévision d'électricité produite à partir du combustible fossile par SARIMA – zone6	71
Figure 32 Données real d'énergie électrique canadien	81
Figure 33 Production Totale (MWh) - CANADA	84
Figure 34 Les prévisions de l'Énergie électrique par le Model ARIMA	89
Figure 35 Les prévisions de l'Énergie électrique par le Model SARIMA	92
Figure 36 Model SARIM entrainer sur l'ensemble de données de 2008 à 2023	95
Figure 37 Vladimir Vapnik 1936	99
Figure 38 Alexey Chervonenkis 1938-2014	99
Figure 39 Régression SVM 'Noyau Linéaire' pour la Prédiction de la Production d'Électricité Québec de 2022-2023	105
Figure 40 Les Erreurs Quadratiques	111
Figure 41 Absolues de Prédiction	112
Figure 42 Prédiction de la Production d'Électricité en Fonction de la Température Moyenne Mensuelle Québec 2008-2022	113
Figure 43 Les Hyperparamètres dans cette étude de SVM améliorer	114
Figure 44 Le processus de la Forêt Aléatoire	120

Figure 45 Prédiction de Consommation d'Électricité par RF	121
Figure 46 Tendance de l'Erreur OOB	123
Figure 47 Schéma du Modèle Hybride approche en Cascade.....	128
Figure 48 Modèle Hybride SARIMA-SVM	129
Figure 49 Comparaison des Observations Réelles, Prédiction SARIMA-SVM.....	
et Observations d'entraînement SVM.....	131
Figure 50 Schémas explicative du Processus d'étude	140
Figure 51 les prédictions du modèle SARIMA-SVM-RF	141
Figure 52 les prédictions et les résidus pour chaque modèle	143
Figure 53 Production Totale Réelle et Prédiction – Entraînement	147
Figure 54 Production Totale Réelle et Prédiction – Test	149
Figure 55 Comparaison des Observations Réelles et Prédiction.....	
SARIMA-SVM-RF - Test.....	151

CHAPITRE 1 : INTRODUCTION

Introduction :

En abordant ce premier chapitre, nous nous engageons dans une exploration approfondie de la prévision énergétique, en mettant un accent particulier sur la prévision de l'énergie électrique pour les provinces du Canada. Nous naviguons à la confluence des modèles statistiques traditionnels, tels que les modèles ARIMA (Autoregressive Integrated Moving Average), réputés pour leur capacité à interpréter les tendances historiques des séries temporelles, et les techniques avancées de l'intelligence artificielle incarnées par les réseaux de neurones artificiels (ANN). Dans le domaine de la prévision énergétique, l'union des modèles ARIMA et ANN représente une avancée significative, marquant un tournant dans notre capacité à comprendre et anticiper les dynamiques énergétiques. Cette approche hybride est encore renforcée par l'intégration des données En temps réel, apportant une révolution dans la collecte et l'analyse des données énergétiques, et offrant une perspective plus dynamique et immédiate sur la consommation et la production d'énergie. En outre, cette étude se penche sur les implications des échanges énergétiques transfrontaliers entre le Canada et les États-Unis, et comment ces interactions, influencées par des facteurs politiques, commerciaux et environnementaux, façonnent la prévision de l'énergie électrique dans les différentes provinces canadiennes. En posant les bases de cette recherche multidimensionnelle, nous visons à fournir des connaissances précis et avant-gardistes pour la prévision de l'énergie électrique, un élément crucial pour la planification énergétique dans le contexte complexe et évolutif du Canada.

1. Contexte Énergétique du Canada :

Plongeons dans l'univers complexe et fascinant de l'énergie au Canada, une mosaïque de ressources naturelles et d'innovations technologiques. Ce vaste pays, caractérisé par sa diversité géographique, offre un panorama énergétique unique, façonné par des

siècles d'exploitation, d'innovation et de défis. De l'hydroélectricité abondante du Québec aux riches réserves de pétrole de l'Alberta, chaque région dépeint une facette différente de la stratégie énergétique nationale, reflétant à la fois des opportunités économiques et des défis environnementaux [30].

1.1. Exploration des Ressources Énergétiques dans les Provinces

Canadiennes :

Le Canada, avec sa vaste étendue géographique et sa diversité naturelle, présente un tableau énergétique complexe et varié. Chaque province et territoire possède un ensemble unique de ressources, défis et opportunités qui façonnent son paysage énergétique [31]. De l'abondante hydroélectricité au Québec et en Colombie-Britannique à l'exploitation intensive des sables bitumineux en Alberta, en passant par la diversification énergétique en Ontario et les initiatives renouvelables dans les Maritimes, le Canada illustre la dynamique d'un pays en transition vers un avenir énergétique plus durable et moins dépendant des combustibles fossiles. Cette exploration détaillée des ressources énergétiques de chaque province canadienne met en lumière non seulement les pratiques actuelles mais aussi les stratégies futures visant à répondre aux besoins énergétiques tout en tenant compte des impératifs environnementaux et du développement durable [32].

▪ **Colombie-Britannique :**

En Colombie-Britannique, l'hydroélectricité domine le paysage énergétique, avec des barrages hydroélectriques importants fournissant la majorité de l'électricité de la province. En plus de l'hydroélectricité, la Colombie-Britannique exploite le gaz naturel et investit dans des énergies renouvelables telles que l'éolien, le solaire et la biomasse. Ces initiatives s'inscrivent dans une politique globale de réduction des émissions de gaz à effet de serre et de transition vers une économie à faible émission de carbone.

- **Alberta :**

L'Alberta est célèbre pour ses vastes réserves de sables bitumineux, une source majeure de pétrole brut. L'exploitation de ces sables bitumineux est au cœur de l'économie de la province, mais elle pose également d'importants défis environnementaux. En parallèle, l'Alberta développe des projets d'énergie éolienne et solaire, cherchant à diversifier son mix énergétique.

- **Saskatchewan :**

La Saskatchewan repose fortement sur les énergies fossiles, notamment le charbon et le gaz naturel, pour sa production d'électricité. Cependant, la province s'engage progressivement vers des sources d'énergie renouvelable, en particulier l'énergie éolienne. Elle est également un leader dans la recherche sur la capture et le stockage du carbone.

- **Manitoba :**

Le Manitoba tire la majorité de son électricité de l'hydroélectricité grâce à ses nombreux fleuves et lacs. La province est un exemple de l'utilisation efficace des ressources hydrauliques pour une énergie propre et renouvelable. En outre, le Manitoba explore activement d'autres sources d'énergie renouvelable comme l'éolien et le solaire.

- **Ontario :**

L'Ontario possède un mix énergétique diversifié, avec une forte dépendance à l'énergie nucléaire, qui représente une part significative de sa production d'électricité. L'hydroélectricité et le gaz naturel complètent ce mix, tandis que l'éolien et le solaire sont en croissance. La province a fermé ses centrales au charbon, marquant une étape importante dans sa transition énergétique.

- **Québec :**

Le Québec est un leader mondial en matière d'hydroélectricité, avec une capacité installée importante gérée par Hydro-Québec. Cette abondance d'énergie renouvelable

positionne le Québec comme un acteur clé dans les efforts de réduction des émissions de gaz à effet de serre au Canada. La province exporte également une partie significative de son électricité hydroélectrique.

- **Nouveau-Brunswick :**

Le Nouveau-Brunswick utilise un mix d'énergie nucléaire, d'hydroélectricité, et de combustibles fossiles (notamment le pétrole et le gaz naturel). La province travaille à augmenter sa part d'énergies renouvelables, notamment l'éolien et le solaire.

- **Nouvelle-Écosse :**

La Nouvelle-Écosse dépend fortement des combustibles fossiles, mais elle fournit des efforts pour diversifier ses sources d'énergie. La province investit dans l'énergie éolienne et d'autres formes d'énergies renouvelables, en plus d'explorer les possibilités de l'énergie marémotrice.

- **Île-du-Prince-Édouard :**

Cette petite province se distingue par son engagement envers les énergies renouvelables, en particulier l'énergie éolienne. L'Île-du-Prince-Édouard est un leader dans la production d'énergie éolienne par habitant au Canada.

- **Terre-Neuve-et-Labrador :**

Terre-Neuve-et-Labrador possède d'importantes ressources hydroélectriques, notamment avec le projet de Muskrat Falls. La province se concentre également sur l'exploitation pétrolière offshore, jouant un rôle important dans l'économie locale.

- **Territoires :**

Les trois territoires du Canada — Yukon, Territoires du Nord-Ouest et Nunavut — ont des défis uniques en matière d'énergie en raison de leur climat rigoureux et de leur éloignement. Ils dépendent largement des générateurs diesel, mais cherchent à intégrer davantage d'énergies renouvelables, notamment l'hydroélectrique, le solaire, et l'éolien, dans leurs réseaux énergétiques.

Chaque province et territoire canadien présente un profil énergétique distinct, reflétant non seulement ses ressources naturelles mais aussi ses priorités politiques et environnementales. Le Canada, dans son ensemble, s'efforce de naviguer vers un avenir énergétique plus durable, avec une emphase croissante sur les énergies renouvelables et la réduction des émissions de gaz à effet de serre.

1.2. Tendances de Consommation et Impacts :

Le paysage énergétique est également façonné par les tendances de consommation des Canadiens. La demande énergétique, influencée par des facteurs tels que les changements démographiques, l'industrialisation et les initiatives de conservation, varie considérablement d'une province à l'autre. Cette diversité soulève des questions importantes sur l'adaptation des stratégies énergétiques aux besoins spécifiques de chaque région, en tenant compte des impératifs de durabilité et d'efficacité [33].

1.3. Enjeux et Perspectives:

La richesse énergétique du Canada, avec ses implications économiques et environnementales, représente un terrain fertile pour la recherche et l'innovation. Les défis liés à la transition vers des énergies plus propres et la gestion efficace des ressources actuelles sont au cœur des discussions sur l'avenir énergétique du pays. Ce panorama complexe offre une opportunité unique d'explorer les différentes stratégies et technologies pouvant contribuer à un avenir énergétique durable et prospère pour le Canada [34].

2. Importance de la Prévision Énergétique :

Immergeons-nous maintenant dans l'importance cruciale de la prévision énergétique au Canada, un aspect fondamental pour orienter les politiques et la gestion des ressources naturelles. Imaginons les vastes implications d'une prévision inexacte – des villes plongées dans l'obscurité à cause de pannes de courant imprévues, des industries

paralysées par des erreurs de calcul de la demande, ou encore des opportunités de développement durable manquées.

2.1. Les Multiples Facettes de la Gestion Énergétique :

Considérez le défi colossal que représente la gestion des ressources énergétiques dans un pays aussi vaste et diversifié que le Canada. La prévision énergétique y joue un rôle prépondérant, agissant comme une boussole guidant les décisions stratégiques. Elle doit tenir compte de la variabilité de la demande, des fluctuations des prix sur les marchés mondiaux, et des enjeux environnementaux comme le changement climatique. La précision de ces prévisions est cruciale pour équilibrer la croissance économique, la sécurité énergétique, et la durabilité environnementale [35].

2.2. L'Impact des Prévisions sur la Politique Énergétique :

Plongez dans l'univers des décideurs politiques, où chaque choix stratégique en matière d'énergie est influencé par des prévisions. Une prévision fiable est un outil puissant pour les gouvernements, permettant de planifier les investissements dans les infrastructures, de promouvoir les énergies renouvelables, et de répondre aux engagements internationaux sur la réduction des émissions. Des prévisions inexactes, en revanche, peuvent conduire à des politiques mal orientées, affectant non seulement l'économie, mais aussi le quotidien des citoyens [35].

3. Présentation des Modèles ARIMA et ANN :

Dans cette partie fascinante, nous explorons les fondements et les applications des modèles ARIMA (Autoregressive Integrated Moving Average) et ANN (Artificial Neural Networks), deux piliers de la prévision énergétique moderne. Imaginez ces modèles comme des outils sophistiqués dans l'arsenal des prévisionnistes, chacun avec ses propres forces et limitations, travaillant ensemble pour peindre une image complète de l'avenir énergétique du Canada.

3.1.Fondements Théoriques des Modèles ARIMA :

Le modèle ARIMA, un vétéran dans le domaine de l'analyse des séries temporelles, excelle dans la prévision de données où les tendances historiques et les schémas saisonniers jouent un rôle crucial. Pensez à ARIMA comme à un historien énergétique, traçant les tendances passées pour prédire l'avenir. Ce modèle est particulièrement utile pour comprendre et prévoir les variations de consommation énergétique sur une base saisonnière ou annuelle, mais il atteint ses limites lorsqu'il s'agit de modéliser des dynamiques plus complexes ou non linéaires [36].

3.2.Potentiel des Réseaux de Neurones Artificiels (ANN) :

En contraste, les ANN sont les virtuoses de la modélisation des relations complexes et non linéaires. Ces réseaux, inspirés du fonctionnement du cerveau humain, apprennent des modèles de données en constante évolution, les rendant idéaux pour intégrer et analyser des données volumineuses et variées. Les ANN peuvent détecter des tendances subtiles et des anomalies dans les données énergétiques, offrant des perspectives précieuses pour les prévisions à court terme et pour réagir aux changements rapides dans les habitudes de consommation [37].

3.3.Comparaison et Contraste des Deux Modèles :

En juxtaposant ARIMA et ANN, nous obtenons une vision équilibrée de leurs capacités respectives. Tandis que ARIMA fournit une base solide grâce à son analyse des tendances historiques, les ANN apportent une flexibilité et une adaptabilité à des données plus volatiles et moins prédictibles. Cette combinaison peut s'avérer particulièrement puissante pour les prévisionnistes énergétiques, offrant une approche complète pour naviguer dans le paysage énergétique en constante évolution du Canada [38].

4. L'Émergence des Données En temps réel:

La révolution des données en temps réel, transforme radicalement le paysage de la prévision énergétique. Imaginez un monde où chaque appareil, chaque capteur, transmet des informations en continu, fournissant une image instantanée et détaillée de la consommation énergétique. Cette section explore comment ces technologies novatrices modifient non seulement notre capacité à collecter des données, mais aussi la manière dont nous les analysons et les utilisons pour prédire les besoins en énergie.

4.1. Révolution des Données En temps réel :

Les dispositifs En temps réel, des compteurs intelligents aux capteurs environnementaux, collectent des données essentielles sur la consommation d'énergie, les conditions climatiques, et d'autres facteurs influençant la demande énergétique. Ces données, par leur nature instantanée, offrent une opportunité sans précédent de comprendre en temps réel les dynamiques énergétiques. Plongez dans l'exploration des défis et des possibilités que ces technologies apportent à la prévision énergétique. Comment les données En temps réel peuvent-elles être utilisées pour rendre les prévisions plus précises et réactives [38]?

4.2. Impact sur la Prévision Énergétique :

La richesse des données en temps réel ouvre de nouvelles voies pour l'analyse énergétique. Ces informations permettent aux modèles prédictifs de s'adapter rapidement aux changements de consommation et de production. Réfléchissez à la manière dont l'intégration de ces données dans des modèles tels que ARIMA et ANN peut améliorer significativement la précision des prévisions. Des exemples concrets illustrent comment l'analyse en temps réel transforme la gestion de l'énergie, rendant les réponses plus agiles face aux fluctuations imprévues de la demande [39].

4.3.Potentiel des Modèles Hybrides avec Données En temps réel :

L'intégration des données En temps réel dans des modèles hybrides ARIMA-ANN représente une avancée majeure dans le domaine de la prévision énergétique. Volatilisant l'analyse historique robuste des modèles ARIMA avec la capacité des ANN à traiter des données complexes et volatiles, les prévisionnistes peuvent désormais accéder à des outils plus sophistiqués et précis. Cette section détaille comment la synergie de ces technologies offre une approche plus complète et adaptative, capable de répondre efficacement aux exigences d'un secteur énergétique en évolution [40].

5. La Synergie des Modèles Hybrides avec l'Intégration des Données En temps réel:

L'union des modèles ARIMA et ANN dans le domaine de la prévision énergétique représente une avancée remarquable, marquant un tournant dans notre capacité à comprendre et anticiper les dynamiques énergétiques. Cette section nous invite à explorer la synergie de ces modèles hybrides, où la rigueur statistique rencontre l'intelligence artificielle, offrant une approche révolutionnaire et adaptative de la prévision énergétique.

5.1.L'Intégration Innovante d'ARIMA et ANN :

Imaginez les modèles ARIMA, avec leur capacité éprouvée à analyser les tendances historiques, s'alliant aux ANN, ces puissants outils capables de décrypter des schémas complexes et de s'adapter à des données en constante évolution. Cette combinaison crée un modèle hybride qui capitalise sur les forces de chaque approche. Les ARIMA fournissent un cadre solide basé sur des données historiques, tandis que les ANN apportent une dimension de flexibilité, idéale pour intégrer les variations et les anomalies imprévues [40].

5.2. Avantages de l'Approche Hybride :

Cette fusion de techniques offre une précision sans précédent dans la prévision énergétique. Les modèles hybrides peuvent s'adapter rapidement aux changements, qu'ils soient saisonniers, climatiques ou dus à des événements imprévus. Ils permettent une analyse plus riche et nuancée, capable de prendre en compte une gamme plus large de variables et de scénarios. En outre, ces modèles peuvent être constamment mis à jour et améliorés au fur et à mesure que de nouvelles données sont disponibles, rendant les prévisions non seulement plus précises, mais aussi plus pertinentes.

5.3. Révolution dans la Prévision Énergétique :

Les implications de cette synergie sont vastes. Dans un monde où l'énergie est un facteur clé de l'économie, de l'environnement et de la société, la capacité à prédire de manière fiable la demande et l'offre d'énergie est cruciale. Les modèles hybrides ARIMA-ANN représentent une révolution dans la manière dont les décideurs, les entreprises et les consommateurs abordent la question de l'énergie. Cette section détaille comment, grâce à ces modèles, nous pouvons non seulement mieux comprendre le présent, mais aussi façonner un avenir énergétique plus durable et efficace [40].

5.4. Un Pas Vers l'Avenir : Combinaison des Modèles ARIMA, ANN:

L'évolution rapide de la technologie et la disponibilité croissante des données en temps réel transforment la manière dont nous approchons la prévision énergétique. Dans cette section enrichie, nous explorons comment l'intégration des données En temps réel renforce la synergie entre les modèles ARIMA et ANN, offrant ainsi une perspective plus dynamique et précise pour la prévision énergétique.

5.4.1. Intégration des Données en temps réel dans les Modèles ARIMA et ANN :

- **Exploitation des Données en Temps Réel** : Les données en temps réel, collectées à partir de capteurs et de compteurs intelligents, fournissent des informations en temps réel sur la consommation et la production d'énergie. L'intégration de ces données dans les modèles ARIMA et ANN permet une analyse plus réactive et adaptative des tendances énergétiques.
- **Détection des Anomalies et Réactivité** : Les données en temps réel aident à détecter rapidement les anomalies de consommation ou de production, permettant aux modèles prédictifs de s'adapter en temps réel aux changements imprévus.

5.4.2. Amélioration de la Précision des Prévisions :

- **Combinaison des Forces des Modèles** : Les modèles ARIMA excellent dans l'analyse des tendances historiques, tandis que les ANN sont performantes dans la modélisation de relations complexes et non linéaires. L'ajout des données en temps réel enrichit ces modèles, en offrant une granularité et une pertinence accrues pour les prédictions.
- **Prévisions Plus Dynamiques** : En intégrant les données en temps réel, les prévisions deviennent plus dynamiques, reflétant mieux les conditions actuelles et anticipant plus efficacement les besoins futurs en énergie.

5.4.3. Applications Pratiques et Scénarios d'Utilisation :

- **Gestion de la Demande et de l'Offre** : Utilisation des prévisions pour une meilleure gestion de la demande et de l'offre d'énergie, en particulier dans les contextes où les sources d'énergie sont diversifiées et intermittentes.

- **Optimisation des Réseaux Énergétiques** : Les données en temps réel permettent une meilleure compréhension et un meilleur contrôle des réseaux énergétiques, en particulier

Dans les contextes de réseaux intelligents, où la gestion en temps réel est cruciale pour l'efficacité et la durabilité.

5.4.4. Défis et Solutions dans l'Intégration des Données en temps réel :

- **Gestion des Volumes de Données** : Face à l'afflux massif de données en temps réel, des techniques avancées de traitement et d'analyse des données sont nécessaires pour gérer efficacement les informations et extraire des connaissances pertinentes.
- **Sécurité et Confidentialité** : Alors que les données en temps réel offrent de précieuses informations, elles posent également des questions de sécurité et de confidentialité. Des mesures robustes sont nécessaires pour protéger ces données tout en les exploitant pour des prévisions énergétiques.

5.4.5. Perspectives avec les Modèles Hybrides :

- **Adaptation aux Changements Rapides** : Les modèles hybrides intégrant des données en temps réel sont particulièrement adaptés pour répondre aux défis d'un secteur énergétique en rapide évolution, marqué par des innovations technologiques et des changements dans les politiques environnementales.
- **Rôle Croissant de l'IA et de l'Apprentissage Automatique** : L'intégration des données en temps réel ouvre la voie à une utilisation plus poussée de l'intelligence artificielle et de l'apprentissage automatique dans la prévision énergétique, offrant des perspectives prometteuses pour des prévisions encore plus précises et personnalisées.

6. Échanges Énergétiques Transfrontaliers :

Les échanges énergétiques transfrontaliers du Canada, notamment avec les États-Unis, sont une facette cruciale et complexe du paysage énergétique national. Cette section nous invite à explorer les dynamiques de ces échanges, leurs implications économiques et politiques, et leur impact sur la prévision énergétique.

6.1. La Dynamique des Échanges Énergétiques avec les États-Unis :

Imaginez les flux d'énergie traversant les frontières - un ballet complexe d'importations et d'exportations qui façonne la stratégie énergétique du Canada. Ces échanges sont influencés par des facteurs tels que les politiques gouvernementales, les demandes du marché et les considérations environnementales. Comment le Canada équilibre-t-il ses besoins internes avec les opportunités et les défis de l'exportation d'énergie, notamment vers son voisin du Sud [41]?

6.2. Implications Économiques et Politiques :

La prévision énergétique dans le contexte des échanges transfrontaliers nécessite une analyse détaillée et nuancée. Les modèles prédictifs doivent tenir compte de variables internationales, des fluctuations des marchés mondiaux, et des changements de politiques énergétiques chez les partenaires commerciaux. Cette complexité ajoute une couche supplémentaire de défi à la prévision énergétique, soulignant la nécessité de modèles hybrides sophistiqués et adaptatifs [42].

7. Objectifs et Structure du Projet :

Dans cette partie de notre introduction, nous définissons clairement les objectifs de notre étude sur la prévision énergétique au Canada et présentons un aperçu structuré de ce rapport. Cette section sert de guide pour naviguer à travers les complexités et les nuances des prochains chapitres, établissant un cadre pour notre exploration approfondie.

7.1. Déclaration des Objectifs de Recherche :

Notre voyage dans le monde de la prévision énergétique canadienne vise à accomplir plusieurs objectifs clés. Premièrement, nous cherchons à comprendre comment les modèles prédictifs tels que ARIMA et ANN peuvent être optimisés et combinés pour améliorer la précision des prévisions énergétiques. De plus, nous explorons l'impact des données en temps réel sur ces modèles et analysons les implications des échanges énergétiques transfrontaliers. Enfin, notre but est d'évaluer comment ces avancées peuvent influencer les politiques énergétiques et la gestion des ressources au Canada.

7.2. Aperçu des Chapitres Suivants :

Ce projet de recherche est structuré pour nous guider à travers les différentes dimensions de la prévision énergétique. Après cette introduction. Dans le troisième chapitre, nous abordons les échanges énergétiques transfrontaliers et leur impact sur la prévision énergétique. Le quatrième chapitre traite des données réelles et de leur intégration dans les modèles prédictifs. Le cinquième chapitre se penche sur les modèles ARIMA et ANN en détail, explorant leurs fondements théoriques et applications pratiques. Les chapitres suivants se concentrent sur l'analyse des données collectées, la méthodologie utilisée dans notre étude, et enfin, nous concluons avec une synthèse de nos découvertes et perspectives pour l'avenir.

8. Hypothèses de l'Étude :

Dans cette section cruciale, nous formulons les hypothèses qui guideront notre étude sur la prévision énergétique au Canada. Ces hypothèses, basées sur des observations préliminaires et des revues de littérature, établissent le cadre conceptuel pour notre recherche.

8.1.Hypothèse sur l'Évolution des Modèles ARIMA et ANN :

Nous postulons que les modèles ARIMA, bien que performants pour analyser des données avec des tendances et des cycles historiques, rencontrent des limites dans des contextes dynamiques. L'intégration des modèles ANN, reconnus pour leur capacité à traiter des données complexes et changeantes, devrait offrir une amélioration substantielle dans la précision des prévisions énergétiques. Cette synergie entre les modèles historiques et l'apprentissage machine moderne pourrait transformer la manière dont les prévisions énergétiques sont réalisées.

8.2.Impact des Données en temps réel sur la Précision des Prévisions :

Grâce à leur nature en temps réel et leur volume, pourraient révolutionner la prévision énergétique. Nous supposons que l'intégration de ces données dans les modèles ARIMA et ANN permettra de capter des tendances subtiles, des anomalies et des changements comportementaux, ce qui rendrait les prévisions plus précises et adaptatives.

8.3.Influence des Politiques Énergétiques et Environnementales :

Les politiques énergétiques et environnementales jouent un rôle déterminant dans la configuration du paysage énergétique. Notre hypothèse est que les changements dans les politiques, en particulier ceux orientés vers la durabilité et la réduction des émissions, auront un impact significatif sur les tendances de consommation et de production d'énergie. Les prévisions devront donc intégrer ces variables politiques pour rester pertinentes et exactes.

8.4.Prévision Énergétique et Transition Vers les Énergies Renouvelables :

Nous envisageons que la transition vers les énergies renouvelables et les technologies vertes au Canada présente à la fois des défis et des opportunités pour la prévision énergétique. L'intermittence des sources renouvelables, comme l'éolien et le solaire,

impose des défis uniques en termes de prévision de la production d'énergie. Les modèles prédictifs devront être suffisamment flexibles pour gérer cette variabilité.

8.5. Rôle des Avancées Technologiques dans la Modélisation Énergétique :

Les avancées technologiques, telles que les réseaux intelligents et le stockage d'énergie, transformeront la manière dont l'énergie est consommée et distribuée. Nous supposons que ces innovations auront un impact profond sur les modèles de prévision, nécessitant des adaptations pour intégrer ces nouvelles dynamiques.

8.6. Globalisation et Interdépendances Énergétiques :

Dans un contexte de globalisation croissante, les marchés énergétiques sont de plus en plus interdépendants. Notre hypothèse est que les tendances mondiales, y compris les fluctuations des prix du pétrole et du gaz, les crises énergétiques internationales, et les changements dans les politiques étrangères, influenceront de manière significative le marché énergétique canadien. Les modèles prédictifs doivent donc prendre en compte ces facteurs externes pour être efficaces.

8.7. Défis et Opportunités de l'Électrification :

L'électrification croissante des transports et des industries est une tendance majeure qui remodelera la demande énergétique. Nous postulons que cette transition vers l'électrification présentera des défis en termes de prévision de la demande, mais offrira également des opportunités pour une gestion plus durable de l'énergie. Les modèles devront s'adapter pour anticiper et planifier cette augmentation de la demande électrique.

8.8. Répercussions du Changement Climatique sur la Demande Énergétique :

Le changement climatique aura des effets considérables sur les modèles de consommation énergétique. Nous envisageons que les effets du réchauffement climatique, tels que les vagues de chaleur et les changements de températures,

influenceront directement la demande en chauffage et en climatisation. Les prévisions énergétiques doivent intégrer ces variables climatiques pour rester pertinentes.

8.9. Interaction entre la Politique Énergétique et la Préviation :

Enfin, nous supposons que les décisions politiques en matière d'énergie ne sont pas seulement influencées par les prévisions existantes, mais qu'elles les façonnent également en retour. Les politiques énergétiques, telles que les subventions pour les énergies renouvelables ou les taxes sur le carbone, peuvent modifier de manière significative les tendances de production et de consommation d'énergie, et par conséquent, influencer les futures prévisions.

Ces hypothèses forment un cadre robuste pour notre étude, permettant une analyse approfondie des multiples facettes de la prévision énergétique dans le contexte canadien. Elles soulignent la complexité et l'importance de considérer un large éventail de facteurs – technologiques, politiques, environnementaux et économiques – pour comprendre et anticiper l'avenir énergétique du Canada.

À l'issue de ce premier chapitre, nous sommes parvenus à une compréhension approfondie des dynamiques complexes de la prévision énergétique, un domaine où la rigueur scientifique se mêle à des applications pratiques innovantes. L'intégration des modèles ARIMA et ANN, traditionnellement considérés comme distincts dans leurs approches de la modélisation statistique et de l'apprentissage machine, a révélé un potentiel considérable pour la précision des prévisions énergétiques. Les modèles ARIMA, avec leur capacité à décoder les tendances et les cycles historiques, se marient harmonieusement aux réseaux de neurones artificiels qui, grâce à leur flexibilité et leur capacité à modéliser des données complexes, offrent une perspective plus dynamique et adaptative.

Conclusion :

L'incorporation des données en temps réel dans ces modèles hybrides marque une avancée majeure dans le domaine. Elle permet une collecte de données plus fine et une réactivité accrue aux variations de consommation et de production d'énergie, éléments cruciaux dans un contexte où l'efficacité énergétique et la réponse aux demandes fluctuantes sont essentielles. Cette approche novatrice répond aux défis actuels du secteur énergétique, caractérisé par une variabilité accrue due aux sources d'énergie renouvelables et aux exigences de durabilité.

Dans le contexte des échanges énergétiques transfrontaliers, en particulier entre le Canada et les États-Unis, nous avons analysé comment ces interactions influencent la prévision énergétique. La gestion de ces échanges nécessite une compréhension approfondie des politiques énergétiques, des tendances du marché et des implications économiques et environnementales. Les modèles prédictifs doivent donc intégrer ces facteurs pour fournir des analyses précises et utiles pour la planification stratégique et la prise de décision.

En conclusion, ce chapitre pose les fondations pour une exploration approfondie des techniques avancées de prévision énergétique. Il souligne la nécessité d'une approche interdisciplinaire et innovante, intégrant des données et des modèles variés pour relever les défis et saisir les opportunités du secteur énergétique en évolution. Cette étude préliminaire ouvre la voie à des recherches plus poussées, guidant ainsi le développement de stratégies énergétiques plus résilientes et durables pour l'avenir.

CHAPITRE 2 : REVUE DE LITTÉRATURE- HISTOIRE ET ÉVOLUTION DES MÉTHODES DE PRÉVISION ÉNERGÉTIQUE

Introduction :

Lorsque l'on contemple l'immensité complexe et fascinante du réseau énergétique mondial, on peut facilement être submergé par l'ampleur des défis qui se dressent devant nous. Au cœur de ces défis se trouve une quête fondamentale : la prévision énergétique. Ce domaine, qui est à la fois un art et une science, a parcouru un long chemin depuis ses humbles débuts. Aujourd'hui, alors que nous nous trouvons à l'aube d'une nouvelle ère marquée par des avancées technologiques fulgurantes et des impératifs environnementaux pressants, il devient impératif de comprendre l'évolution historique et les développements actuels de la prévision énergétique.

Ce chapitre est une plongée dans le monde de la prévision énergétique, explorant les méthodologies qui ont façonné ce domaine au fil des décennies. Nous remontons le temps pour examiner les racines de ces méthodes, comment elles ont émergé des besoins primordiaux de l'humanité pour la gestion des ressources, et comment elles ont évolué face aux défis sans cesse croissants. Dans cette quête, nous entrelaçons les fils du passé et du présent pour tisser une histoire riche et nuancée de la prévision énergétique.

À l'aube de la révolution industrielle, la prévision énergétique était une tâche relativement simple, régie principalement par les besoins immédiats et les estimations rudimentaires. Cependant, avec l'augmentation exponentielle de la demande énergétique et la complexification des réseaux de distribution, la nécessité d'une approche plus systématique et scientifique est devenue évidente. C'est dans ce contexte que sont nées les premières méthodologies de prévision, ancrées dans la pensée

statistique et économétrique. Ces méthodes, bien qu'élémentaires par les standards actuels, ont jeté les bases pour des avancées futures.

Au fil des ans, le domaine de la prévision énergétique a été façonné par des innovations technologiques et des crises mondiales, chacune apportant son lot de défis et d'opportunités. Les chocs pétroliers des années 1970, par exemple, ont été un catalyseur majeur, incitant à des recherches plus approfondies et à l'adoption de méthodes prédictives plus sophistiquées. Cela a conduit à l'essor de modèles statistiques complexes tels que les modèles ARIMA, qui ont permis une compréhension plus nuancée des tendances de consommation énergétique.

Plus récemment, l'avènement de l'intelligence artificielle a marqué le début d'une ère nouvelle et passionnante dans la prévision énergétique. Ces technologies offrent la promesse d'une précision sans précédent et la capacité de traiter des quantités massives de données en temps réel. L'intégration de l'IA dans les modèles prédictifs a ouvert des horizons inexplorés, permettant non seulement de prévoir la demande mais aussi de réagir de manière proactive aux changements dynamiques du marché énergétique.

Dans ce chapitre, nous explorons cette trajectoire fascinante, en mettant en lumière les jalons clés et les figures influentes qui ont contribué à façonner le domaine de la prévision énergétique. De l'analyse des tendances historiques aux modèles prédictifs basés sur l'IA, chaque étape de cette évolution révèle une couche supplémentaire de complexité et d'ingéniosité. En parcourant cette chronologie, nous cherchons non seulement à comprendre le passé mais aussi à entrevoir l'avenir de la prévision énergétique, un domaine qui continue d'évoluer à un rythme vertigineux face aux défis et aux opportunités du 21^e siècle.

1. Les Racines de la Prévision Énergétique :

La prévision énergétique, un domaine crucial dans la gestion des ressources énergétiques, a évolué à travers diverses phases historiques et scientifiques, s'appuyant

sur une combinaison de méthodes empiriques, d'observations qualitatives, et d'approches statistiques et mathématiques avancées. Les origines de la prévision énergétique remontent aux sociétés primitives, où la nécessité de gérer des ressources naturelles essentielles, telles que le bois pour le chauffage et la cuisson, a conduit à l'élaboration de méthodes prédictives rudimentaires. Ces méthodes initiales, basées sur l'expérience directe et l'observation des cycles naturels, bien que limitées en précision et en portée, représentaient les premiers pas vers une compréhension empirique de la gestion des ressources énergétiques. Les communautés utilisaient des observations qualitatives pour estimer leurs besoins en ressources critiques, telles que la quantité de bois consommée durant des périodes similaires dans le passé, sans recourir à des méthodologies scientifiques structurées [13].

Cependant, avec l'avènement de la révolution industrielle et l'émergence des premières crises énergétiques, notamment les pénuries de charbon, un besoin pressant de méthodes de prévision plus fiables et systématiques est apparu. Ces crises ont agi comme des catalyseurs pour l'intégration de la pensée scientifique, mathématique et statistique dans les processus de prévision énergétique. Des figures historiques comme William Stanley Jevons au Royaume-Uni, avec son œuvre "The Coal Question" (1865), ont joué un rôle pionnier dans cette évolution. Jevons a analysé l'utilisation du charbon et son impact sur l'économie britannique, posant ainsi les bases de la réflexion moderne sur la prévision de la demande énergétique. Son travail a marqué un tournant décisif dans la manière de comprendre et de quantifier la consommation et la production d'énergie, passant d'une approche basée principalement sur l'observation et l'expérience à une méthodologie plus rigoureuse et analytique [14].

L'émergence des premières théories statistiques a également joué un rôle crucial dans l'évolution de la prévision énergétique. Les contributions de statisticiens et mathématiciens tels que Karl Pearson et Ronald Fisher ont été fondamentales. Leur

travail a établi les bases des méthodes statistiques modernes, introduisant des concepts tels que la corrélation, la régression linéaire, et les premières formes d'analyse de séries temporelles.

Pearson, par exemple, a introduit le concept de corrélation, fondamental pour les modèles prédictifs. La formule de corrélation de Pearson (r) est définie comme :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Où :

- r : est le coefficient de corrélation de Pearson.
- Σ : représente la somme sur l'ensemble des données.
- x_i et y_i : sont les valeurs individuelles des ensembles de données X et Y.
- \bar{x} et \bar{y} : sont les moyennes des ensembles de données X et Y, respectivement.
- n : est le nombre d'observations dans les ensembles de données.

Ces outils ont permis une application plus précise et quantitative dans l'analyse des données énergétiques. L'application de ces méthodes statistiques et mathématiques a transformé la prévision énergétique d'une discipline basée sur des estimations et des observations qualitatives en une science rigoureuse, capable de produire des prédictions précises et fiables sur la consommation et la production d'énergie [15].

Par exemple, la régression linéaire, une méthode introduite par Francis Galton (un protégé de Pearson), a commencé à être utilisée pour modéliser la relation entre la consommation énergétique et des facteurs comme la croissance économique. La formule de base de la régression linéaire est :

$$Y = a + bX + \varepsilon$$

Où :

- Y : est la variable dépendante (par exemple, la consommation d'énergie)
- X : est la variable indépendante (par exemple, le PIB "Produit Intérieur Brut").
- a : est la pente de la ligne de régression,
- b : est l'ordonnée à l'origine.
- ε : représente l'erreur résiduelle, c'est-à-dire la différence entre la valeur observée et la valeur prédite par le modèle.

La relation entre la régression linéaire et la corrélation de Pearson est que le coefficient de corrélation (r) est lié à la pente a dans un modèle de régression simple.

Pour être plus précis, r est le produit de la pente a et du rapport des écarts-types des variables X et Y . La formule est donnée par :

$$r = a \times \frac{\sigma_x}{\sigma_y}$$

Où :

- σ_x et σ_y : Sont les écarts- types des variables X et Y , respectivement.

Dans le contexte contemporain, la prévision énergétique s'appuie sur des techniques avancées d'intelligence artificielle (IA) et d'apprentissage automatique (ML), incorporant des analyses complexes de données historiques, des modèles météorologiques, des facteurs économiques et démographiques, des avancées technologiques, et des réglementations gouvernementales. Ces approches permettent non seulement de prédire la demande et l'offre d'énergie avec une précision accrue mais ouvrent également la voie à une gestion plus efficace et durable des ressources énergétiques. Ainsi, la prévision énergétique, en tant que champ d'étude et de pratique, continue d'évoluer, intégrant de nouvelles technologies et méthodologies pour répondre aux défis énergétiques de plus en plus complexes du 21e siècle [16].

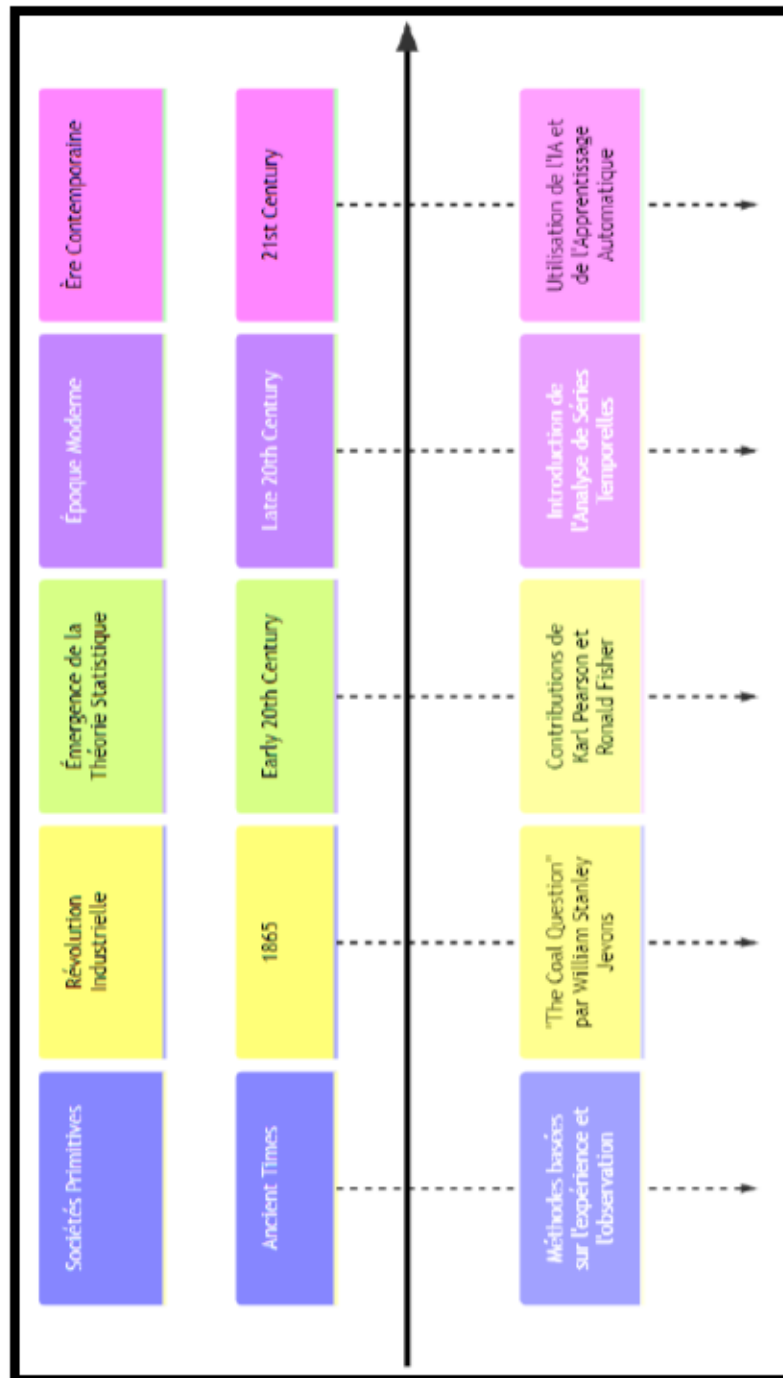


Figure 1 Évolution de la prévision Énergétique

Une vue plus complète de l'évolution de la prévision énergétique :

Tableau 1 Vue plus complète de l'évolution de la prévision énergétique

ANNEES	Événements et Innovations Clés dans la Prévision Énergétique	Relevance pour le Projet	Ma Contribution
1890s	Karl Pearson développe la corrélation de Pearson.	Fondations des modèles prédictifs.	Compréhension des principes statistiques fondamentaux.
1920s-1930s	Ronald Fisher étend les méthodes statistiques.	Amélioration des outils d'analyse.	Application de méthodes statistiques avancées.
1970s	George Box et Gwilym Jenkins introduisent les modèles ARIMA.	Base des méthodes de prévision traditionnelles.	Adaptation des modèles ARIMA aux données énergétiques actuelles.
1980s	Début de l'ère informatique dans la prévision.	Utilisation de l'informatique pour l'analyse des données.	Intégration de l'informatique dans l'analyse énergétique.
1990s	Montée des technologies numériques et de la modélisation.	Amélioration des capacités d'analyse des données.	Utilisation de modèles numériques pour la prévision.
2010s	Émergence des techniques d'apprentissage automatique.	Introduction des ANN pour l'analyse de données.	Exploitation des ANN pour des prédictions complexes.
2020	Développement de l'hybride ANN-ARIMA.	Cœur du projet pour une prévision plus précise.	Création de modèles hybrides ANN-ARIMA.
2020	Approche de durabilité dans la prévision énergétique.	Vision d'un avenir énergétique durable.	Contribution à une stratégie de durabilité énergétique.

Des points du Tableau 1 seront présentée dans les chapitres suivants avec une analyse détaillée.

2. Formulation des Premières Théories et Modèles :

2.1. Théories Fondamentales :

Des théoriciens comme « **Andrey Kolmogorov** » et « **Norbert Wiener** » ont jeté les bases des processus stochastiques, essentiels pour la modélisation des séries temporelles en prévision énergétique. Le travail de Wiener sur les séries temporelles et le bruit (Wiener Process) a été particulièrement influent.

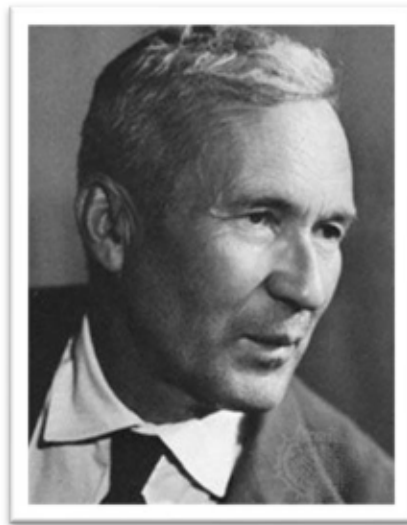


Figure 2 Andrey Kolmogorov 1903-1987



Figure 3 Norbert Wiener 1894-1964

Ils ont posé les bases des processus stochastiques. Leur contribution à la théorie des probabilités a été essentielle pour comprendre et modéliser le comportement aléatoire des séries temporelles, un aspect fondamental de la prévision énergétique. Nous explorerons comment leur travail a permis d'appliquer des méthodes mathématiques rigoureuses à l'analyse des données énergétiques.

« **Kolmogorov** » a établi les axiomes de base des probabilités, cruciaux pour les processus stochastiques. Son approche axiomatique a jeté les bases mathématiques des

probabilités, permettant une modélisation plus rigoureuse des phénomènes aléatoires [17], parmi les axiomes fondamentaux de la probabilité souvent attribués sont :

- **Axiome de non-négativité** : La probabilité d'un événement est toujours un nombre réel non négatif.

Pour tout événement A, la probabilité P(A) est supérieure ou égale à zéro.

$$\forall A ; P(A) \geq 0$$

- **Axiome de la normalisation** : La probabilité de l'ensemble des résultats possibles (l'univers) est égale à 1. C'est-à-dire que la somme des probabilités de tous les événements élémentaires doit être égale à 1.

Soit S l'univers des résultats possibles d'une expérience aléatoire, et $A_1, A_2, \dots, A_n, \dots$ les événements élémentaires associés à cet univers. Alors, l'axiome de la normalisation peut être formulé comme :

$$P(S) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots = 1$$

- **Axiome de l'addition des probabilités** : Pour deux événements mutuellement exclusifs (c'est-à-dire qui ne peuvent pas se produire simultanément), la probabilité de l'union de ces événements est égale à la somme de leurs probabilités individuelles.

Pour A et B deux événements mutuellement exclusifs :

$$P(A \cup B) = P(A) + P(B)$$

« Wiener » a développé le concept de mouvement brownien qui est un type de processus stochastique, ce qui signifie qu'il modélise l'évolution aléatoire d'une variable au fil du temps, les caractéristiques principales du mouvement brownien sont les suivantes :

1. **Continuité** : Le mouvement brownien est continu, ce qui signifie qu'il n'a pas de discontinuités dans son trajet.

2. **Variabilité** : Les trajectoires du mouvement brownien sont caractérisées par une variabilité constante. Cela est souvent représenté par des changements aléatoires et imprévisibles dans la position de la particule au fil du temps.
3. **Indépendance** : Les incréments du mouvement brownien sont indépendants les uns des autres, ce qui signifie que le comportement du processus à un moment donné est indépendant de son comportement antérieur.

Ce Processus est défini comme :

$$dW_t = \varepsilon_t \sqrt{dt}$$

Où :

- W_t : est le Processus de Wiener à l'instant t.
- ε_t : est une variable aléatoire qui suit une distribution normale centrée réduite (c'est-à-dire, une variable aléatoire normale avec une moyenne de 0 et un écart-type de 1).
- dt : est un petit incrément de temps.

Ce formule décrit l'incrément élémentaire du Processus de Wiener.

Le Processus de Wiener est largement utilisé dans la modélisation des mouvements aléatoires, en particulier dans le domaine de la finance et d'autres domaines où des phénomènes aléatoires sont observés. Il est également la base théorique pour le calcul stochastique [18].

2.2. Développement des Théories Statistiques:

2.2.1. L'Ère Pionnière (Fin du 19ème - Début du 20ème Siècle) :

« **Karl Pearson** », dans les années 1890, et « **Ronald Fisher** », au début du 20ème siècle, ont jeté les bases des méthodes statistiques modernes. La corrélation de Pearson,

introduite en 1896, a permis de mesurer quantitativement les relations entre les variables, un progrès fondamental pour l'analyse des données énergétiques.



Figure 4 Karl Pearson 1857-1936

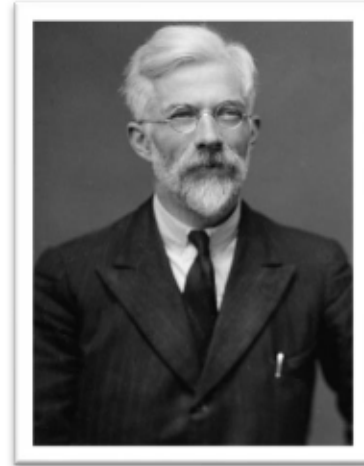


Figure 5 Ronald Fisher 1890-196

La corrélation de Pearson (r) entre deux variables x et y est calculée à l'aide de la formule suivante :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Où :

- n : est le nombre d'observations.
- x_i et y_i : sont les valeurs individuelles des deux variables.
- \bar{x} et \bar{y} : représentent les moyennes des variables x et y respectivement.

Cette formule mesure la force et la direction de la relation linéaire entre deux variables. Un coefficient de corrélation r proche de 1 indique une forte corrélation positive, -1 une forte corrélation négative, et 0 une absence de corrélation linéaire.

Cette avancée a été cruciale pour décrypter les complexités des données énergétiques, ouvrant la voie à des prévisions plus précises basées sur des analyses statistiques [15].

2.2.2. L'Innovation des Années 1970 - Les Modèles ARIMA :

Les modèles ARIMA (Auto Regressive Integrated Moving Average), développés par « George Box » et « Gwilym Jenkins » dans les années 1970, ont révolutionné la modélisation des séries temporelles en prévision énergétique.



Figure 6 George Box 1919-2013

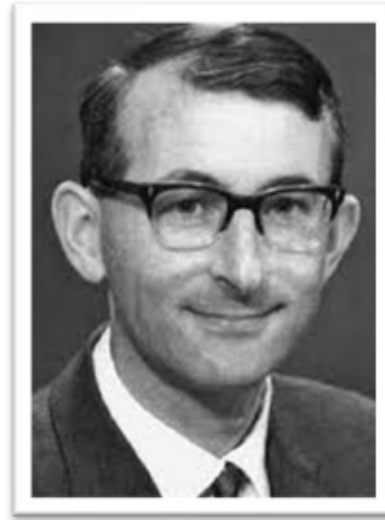


Figure 7 Gwilym Jenkins 1932-1982

La discussion portera sur la structure du modèle ARIMA, Leur approche a combiné les composants autorégressifs (AR) et de moyenne mobile (MA) y compris la différenciation intégrée (I) pour rendre les données stationnaires, et comment cela a amélioré la précision des prévisions énergétiques.

La formule générale d'un modèle ARIMA(p,d,q),

Avec :

- p (AR) : degré de l'ordre autorégressif (**AutoRegressive**) - il indique le nombre de termes autorégressifs dans le modèle. Ces termes capturent la relation linéaire entre la valeur actuelle de la série temporelle et ses valeurs passées.
- d (I) : degré de différenciation (**Integrated**) - il indique le nombre de différences non saisonnières nécessaires pour rendre la série temporelle

stationnaire. La stationnarité est importante pour l'application de modèles ARIMA.

- q (MA): degré de l'ordre de la moyenne mobile (Moving Average) - il indique le nombre de termes de moyenne mobile dans le modèle. Ces termes capturent la relation linéaire entre la valeur actuelle de la série temporelle et les erreurs passées.

Est :

$$Y'_t = c + \sum_{i=1}^p \phi_i Y'_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

Où :

- Y'_t : est la valeur transformée de la série temporelle à l'instant t après d différenciations.
- c : est une constante.
- ϕ_i : sont les coefficients autorégressifs pour les p termes autorégressifs.
- ε_t : est le terme d'erreur à l'instant t .
- θ_i : sont les coefficients de la moyenne mobile pour les q termes de moyenne mobile.

La détermination des valeurs optimales pour p , d et q implique généralement une analyse approfondie des données et l'utilisation de techniques telles que l'autocorrélation, l'autocorrélation partielle, et la vérification de la stationnarité de la série temporelle après les différenciations.

2.2.3. Les années 1970 et 1980 - Modélisation Numérique et Algorithmique :

L'introduction des méthodes comme le lissage exponentiel, particulièrement adapter aux prévisions à court terme, elle est utilisée pour calculer la moyenne pondérée d'une série temporelle (Y_t) en utilisant une constante de lissage (α) :

$$S_t = \alpha \times Y_t + (1 - \alpha) \times S_{t-1}$$

Où :

- S_t : est la valeur lissée à l'instant t.
- Y_t : est la valeur réelle de la série temporelle à l'instant t.
- α : est la constante de lissage, avec $0 \leq \alpha \leq 1$.
- S_{t-1} : est la valeur lissée à l'instant t-1.

Cette formule met à jour la valeur lissée S_t en combinant une fraction (α) de la valeur réelle (Y_t) avec une fraction complémentaire ($1 - \alpha$) de la valeur lissée précédente (S_{t-1}). La constante de lissage α détermine le poids relatif entre la valeur réelle et la valeur lissée précédente. Un α plus proche de 1 accorde plus de poids à la valeur réelle actuelle, tandis qu'un α plus proche de 0 donne plus de poids à la valeur lissée précédente [19].

2.2.4. Les années 1980 au-delà - Les Modèles de Prévision à l'Ère du Numérique, de l'Intelligence Artificielle et des Big Data :

Développements des Années 1990 et 2000 : L'évolution continue des technologies informatiques et numériques dans les années 1990 et 2000 a encore amélioré les capacités de prévision. L'intégration des systèmes d'information géographique (SIG) et des outils d'analyse spatiale a permis de contextualiser les données énergétiques dans des cadres géographiques et socio-économiques, offrant une perspective plus holistique de la demande et de l'offre énergétiques.

Les années 2010, les réseaux de neurones artificiels sont des modèles informatiques inspirés du fonctionnement du cerveau humain. Ils sont capables de reconnaître des modèles complexes à travers des couches de neurones et sont particulièrement utiles pour l'analyse de données non linéaires. Dans la formule, W représente les poids, b le biais, et f la fonction d'activation [20].

$$y = (W \times x + b)$$

Où :

- y : est la sortie du neurone (ou de la couche de neurones).
- W : représente les poids (matrice de poids), qui sont ajustés pendant l'apprentissage pour optimiser les performances du modèle.
- x : est le vecteur d'entrée.
- b : est le biais (vecteur de biais), un terme constant ajouté à la transformation linéaire.
- f : est la fonction d'activation, qui introduit une non-linéarité dans le modèle.

Les années 2020, l'hybride ANN-ARIMA combine la modélisation statistique des séries temporelles de l'ARIMA avec la puissance des réseaux de neurones artificiels. Cette combinaison permet de tirer parti des forces de chaque méthode, offrant une approche plus complète et précise pour la prévision, en particulier dans des domaines complexes comme la prévision énergétique.

Plus de détails sera présenter dans les chapitres suivants.

Nous avons parcouru un spectre étendu de l'évolution méthodologique, depuis les approches empiriques initiales jusqu'aux systèmes complexes d'aujourd'hui qui intègrent l'intelligence artificielle. Cette rétrospective met en évidence non seulement les avancées techniques mais aussi la croissance intellectuelle dans la compréhension des systèmes énergétiques [21].

Tableau 2 aperçu des développements mathématiques et statistiques qui ont joué un rôle clé dans l'évolution de la prévision énergétique

Période	Modèle/Formule	Description
Début 20ème Siècle	Corrélation de Pearson	$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$
Années 1970	Modèle ARIMA (p, d, q)	$Y'_t = c + \sum_{i=1}^p \phi_i Y'_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$
Années 1980- 1990	Lissage Exponentiel	$S_t = \alpha \times Y_t + (1 - \alpha) \times S_{t-1}$
Années 2010	Réseaux de Neurones (ANN)	$y = f(W \times x + b)$
Années 2020 et au-delà	Hybride ANN- ARIMA	Combinaison des techniques ARIMA et ANN pour des prévisions améliorées.

Conclusion :

Les premières méthodes de prévision, principalement basées sur des analyses statistiques simples et des observations qualitatives, ont posé les fondations pour une compréhension plus rigoureuse de la consommation énergétique. L'introduction des modèles ARIMA a marqué une évolution significative, permettant des prévisions plus précises grâce à la modélisation des séries temporelles. Cependant, ces modèles, bien qu'avancés pour leur époque, ont rencontré des limites dans leur capacité à gérer la complexité croissante des données énergétiques et les influences externes fluctuantes.

Cette revue de littérature révèle que la prévision énergétique est un domaine en constante évolution, où l'innovation technologique joue un rôle clé dans l'adressage des défis émergents. Les progrès dans ce domaine ne sont pas uniquement technologiques mais reflètent également une meilleure compréhension des systèmes

énergétiques et de leur interaction avec les facteurs économiques, environnementaux et sociaux.

Enfin, cette exploration historique et méthodologique souligne l'importance cruciale de la recherche continue et du développement dans la prévision énergétique. Alors que nous faisons face à des défis tels que la transition énergétique et le changement climatique, l'amélioration des méthodes de prévision est essentielle pour assurer une gestion efficace et durable des ressources énergétiques dans le futur.

CHAPITRE 3 : ANALYSE DU PAYSAGE ÉNERGÉTIQUE CANADIEN : POLITIQUE, RELATION INTERNATIONALES

1. Panorama de l'énergie au Canada : Sources, exploitation et perspectives :

Le Paysage Énergétique Canadien révèle une diversité et une richesse remarquables en termes de ressources énergétiques. Le Canada se distingue par sa capacité à produire de l'énergie à partir d'une variété de sources, notamment l'hydroélectricité, le pétrole, le gaz naturel, et un éventail croissant d'énergies renouvelables.

➤ **Hydro-électricité :**

Le Canada, considéré comme le troisième plus grand producteur d'hydroélectricité au monde, tire avantage de sa géographie propice pour générer de l'énergie à partir de l'eau en mouvement. La quantité d'énergie produite dépend du volume et de la vitesse de l'eau, ce qui rend les sites proches des chutes d'eau particulièrement précieux pour la production d'hydroélectricité. Les provinces de Québec, de Colombie-Britannique, de Terre-Neuve-et-Labrador et d'Ontario génèrent la majorité de l'énergie hydroélectrique du Canada, avec le Québec à lui seul produisant plus de 50 % de l'hydroélectricité du pays [43].



Figure 8 Dans la région du Saguenay-Lac-St-Jean au Québec [43]

Malgré l'exploitation importante de cette ressource, on estime que le Canada possède encore environ 160 000 mégawatts de potentiel hydroélectrique non exploité, promettant une expansion future significative dans ce domaine [44].

➤ **Pétrole et gaz naturel :**

Les réserves de pétrole et de gaz naturel du Canada, principalement situées en Alberta, constituent une part importante de son économie énergétique. L'exploitation des sables bitumineux d'Alberta, en particulier, soulève des défis environnementaux en raison des émissions de gaz à effet de serre et de l'impact sur les écosystèmes locaux. La production et l'exportation de gaz naturel sont également cruciales pour l'économie énergétique canadienne, avec une grande partie de cette production destinée aux marchés étrangers.

➤ **Énergies renouvelables :**

Le potentiel des énergies renouvelables, telles que l'énergie solaire, éolienne et la biomasse, est de plus en plus reconnu et exploité au Canada. Bien que leur contribution au mix énergétique global reste modeste en comparaison de l'hydroélectricité et des combustibles fossiles, ces sources d'énergie connaissent une croissance rapide. Cette expansion est soutenue par des avancées

technologiques et des politiques gouvernementales favorables, contribuant à l'objectif du Canada de promouvoir une transition énergétique durable.

Cette richesse énergétique joue un rôle crucial non seulement dans l'économie nationale mais aussi sur le marché énergétique mondial, où le Canada figure comme un acteur majeur en tant que producteur et exportateur d'énergie.

1.1. Politiques gouvernementales en matière d'énergie :

Le Canada a entrepris une transformation ambitieuse de son système énergétique, encadrée par des politiques gouvernementales claires et un cadre réglementaire axé sur les objectifs climatiques et la transition énergétique. Les politiques énergétiques et le cadre réglementaire du Canada se caractérisent par leur approche multidimensionnelle, visant à réduire les émissions de gaz à effet de serre tout en favorisant le développement durable et la croissance économique.

La série " **L'Avenir Énergétique du Canada 2023**" de la Régie de l'énergie du Canada souligne l'importance de la diversité énergétique du Canada et explore les scénarios possibles pour l'avenir, notamment l'objectif ambitieux de zéro émission nette d'ici 2050. Ce rapport illustre la complexité de la transition énergétique canadienne, qui doit tenir compte de l'équilibre entre les sources d'énergie conventionnelles et le développement accéléré des technologies renouvelables, dans le contexte d'une économie décentralisée et diverse [45].

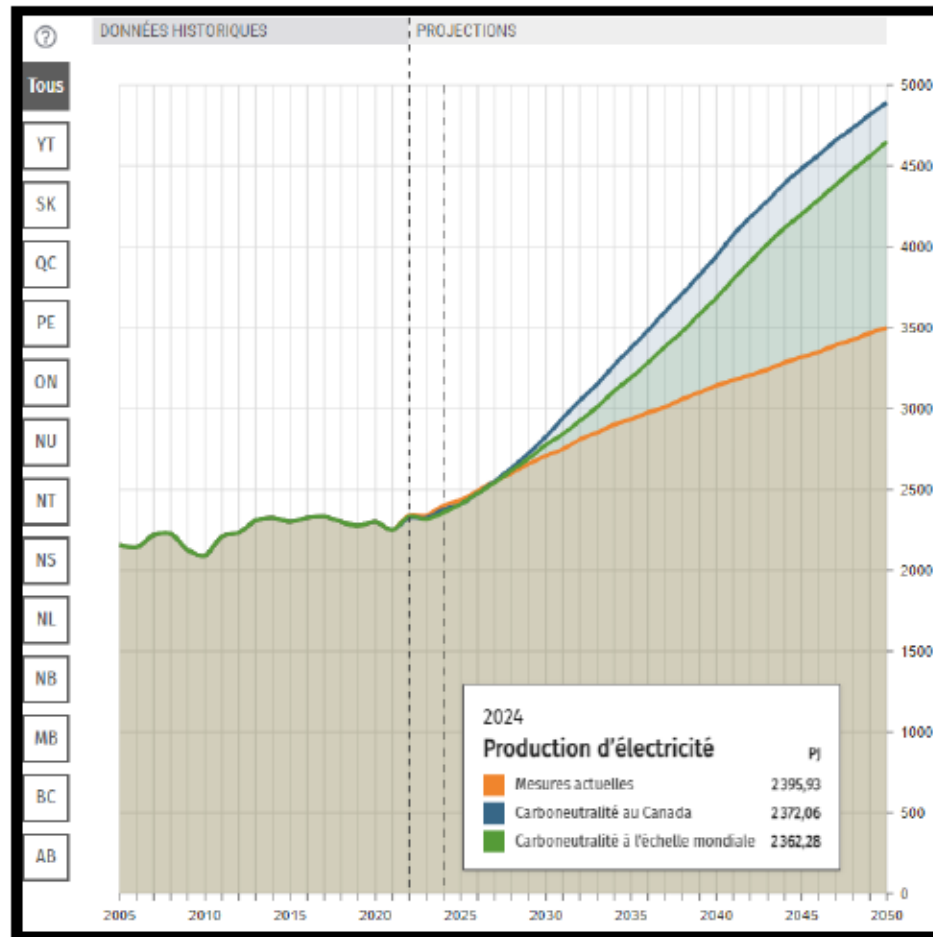


Figure 9 Comparaison des scénarios pour la production d'électricité (Pétajoule) [45].

Selon l'Agence Internationale de l'Énergie (IEA), le Canada a pris des engagements significatifs tant au niveau national qu'international pour transformer son système énergétique. L'accent est mis sur la réduction des émissions de gaz à effet de serre, avec des objectifs de réduction ambitieux fixés pour 2030 et la vision de parvenir à la neutralité carbone d'ici 2050. Ce cadre politique témoigne de la volonté du Canada de s'aligner sur les initiatives globales de lutte contre le changement climatique tout en tenant compte des défis uniques liés à son statut de grand producteur, consommateur et exportateur d'énergie [46]. Pour soutenir ces cibles, le Canada a adopté plusieurs mesures politiques, y compris un système ambitieux de tarification du carbone, des

normes sur les carburants propres, l'engagement d'éliminer progressivement l'utilisation du charbon non atténué d'ici 2030, l'extension des centrales nucléaires, des réglementations sur le méthane dans le secteur pétrolier et gazier, ainsi que des programmes d'efficacité énergétique et des mesures pour décarboniser le secteur des transports [52].

1.2. Transition énergétique : Stratégies pour une économie à faible émission de carbone :

La transition énergétique au Canada est également soutenue par des innovations et des investissements dans des technologies propres, telles que le captage et le stockage du carbone, l'hydrogène propre et les petits réacteurs nucléaires modulaires. Le soutien fédéral à la recherche, au développement et à la démonstration de ces technologies est essentiel pour accélérer les progrès vers la décarbonisation des secteurs de la production pétrolière et gazière, ainsi que des transports et de l'industrie [52].

L'aperçu fourni par le **Conseil Canadien de l'Énergie** révèle une autre dimension importante du secteur énergétique canadien : sa main-d'œuvre. Le secteur est marqué par une diversité croissante, avec une participation significative des femmes, des communautés autochtones et des immigrants. Cette diversité reflète les changements sociétaux au Canada et met en lumière l'importance d'une transition énergétique inclusive qui offre des opportunités équitables à tous les segments de la société canadienne [47].

En conclusion, le paysage énergétique canadien met en évidence la position unique du pays en tant que puissance énergétique mondiale, dotée d'une richesse en ressources naturelles et d'un engagement ferme envers les objectifs de durabilité et de réduction des émissions. Les défis liés à la transition vers un avenir énergétique plus durable sont nombreux, mais le Canada est bien positionné pour les relever, grâce à ses politiques innovantes, sa main-d'œuvre diversifiée et sa capacité d'adaptation.

2. Modèles de consommation d'énergie :

L'analyse des modèles de consommation d'énergie au Canada par secteur révèle des dynamiques complexes influencées par des facteurs tels que les avancées technologiques, les politiques énergétiques et les tendances démographiques et économiques. Ces facteurs façonnent la demande d'énergie dans les secteurs résidentiel, commercial, industriel et des transports, et ont un impact significatif sur les efforts de réduction des émissions de gaz à effet de serre.

Dans le secteur résidentiel, la consommation d'énergie est principalement motivée par le besoin de chauffage et de refroidissement, ce qui varie considérablement selon les types de logements et les conditions climatiques régionales. La méthodologie de désagrégation spatiale et temporelle de la consommation d'énergie utile dans ce secteur tient compte des types de bâtiments résidentiels, de l'année de construction, de la taille et du nombre de résidents, en ajustant la consommation annuelle de chauffage en fonction de la zone climatique spécifique à chaque région [6].

Le secteur industriel et commercial, quant à lui, est caractérisé par une consommation d'énergie déterminée par la combinaison de modèles Top-down et Bottom-up, basée sur la consommation finale d'énergie pour les sous-secteurs allemands du commerce et de l'industrie. Cette consommation est ensuite désagrégée spatialement au niveau NUTS3 '**Nomenclature des unités territoriales statistiques niveau 3**', Ce système permet la collecte, le développement et l'harmonisation des statistiques régionales, en fonction des chiffres d'emploi spécifiques aux sous-secteurs et des intensités énergétiques au sein de ces sous-secteurs, traduisant la consommation d'énergie finale en séries temporelles de consommation d'énergie utile selon des efficacités de conversion spécifiques à l'usage final [48].

Dans le secteur des transports, l'estimation de la consommation d'énergie utile repose sur les profils de conduite régionaux des différents types de véhicules et types de routes, avec une répartition spatiale des kilométrages des véhicules basée sur des statistiques

officielles de stock et de kilométrage de véhicules. Cette méthodologie permet de désagréger temporellement les valeurs annuelles en une résolution horaire, offrant ainsi une image détaillée de la consommation dans le secteur de la mobilité [48].

En outre, les scénarios d'électricité nette zéro au Canada suggèrent que les systèmes électriques continueront à être très distincts à travers le pays, même dans un avenir faible en carbone. Dans chaque scénario d'électricité nette zéro, les dix provinces répondent à leurs demandes d'électricité de manière diverse, avec des mélanges largement variés d'hydro, nucléaire, combustibles fossiles avec CCS (Capture et stockage du CO₂), éolien, solaire, hydrogène et biomasse avec CCS. La transmission entre provinces est un facteur clé qui permet au système électrique d'atteindre le net zéro, démontrant l'importance de la coopération interprovinciale dans la décarbonisation du secteur de l'énergie au Canada [49].

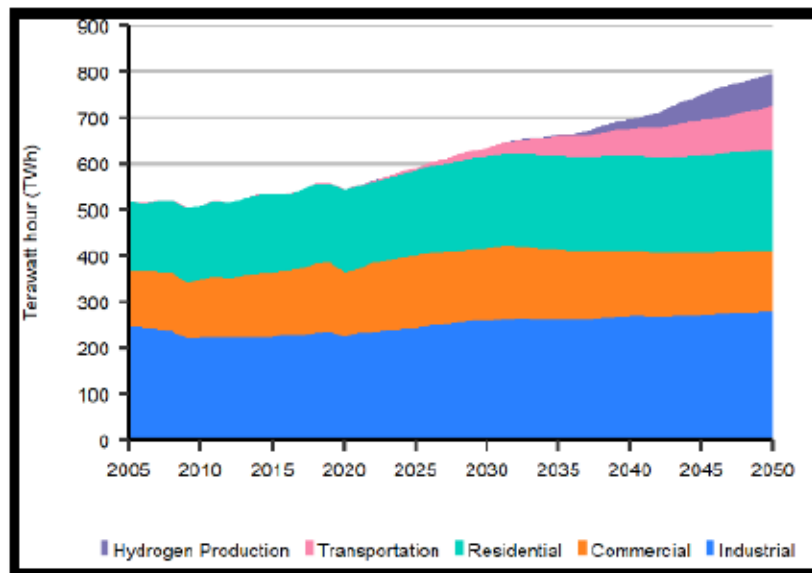


Figure 10 Demande d'électricité par secteur, scénario d'évolution des politiques [7].

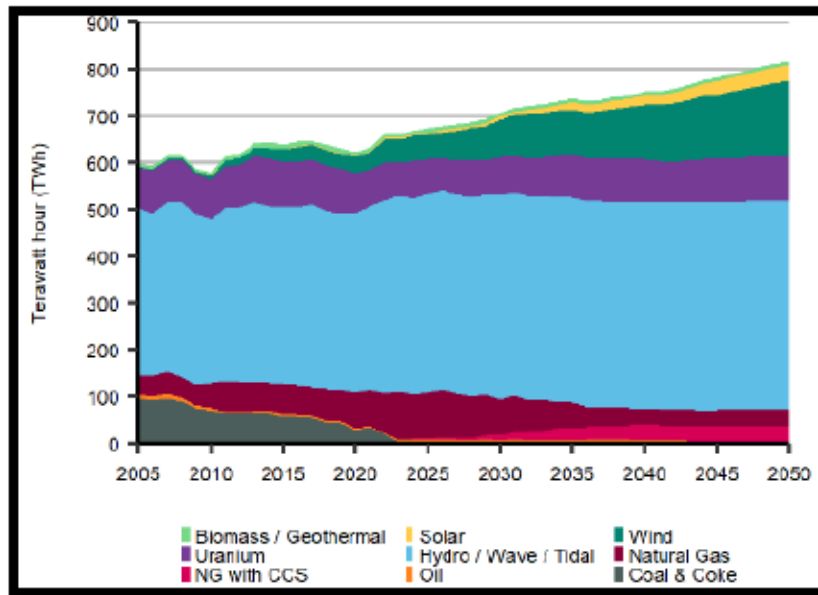


Figure 11 Production d'électricité par source, scénario d'évolution des politiques [7].

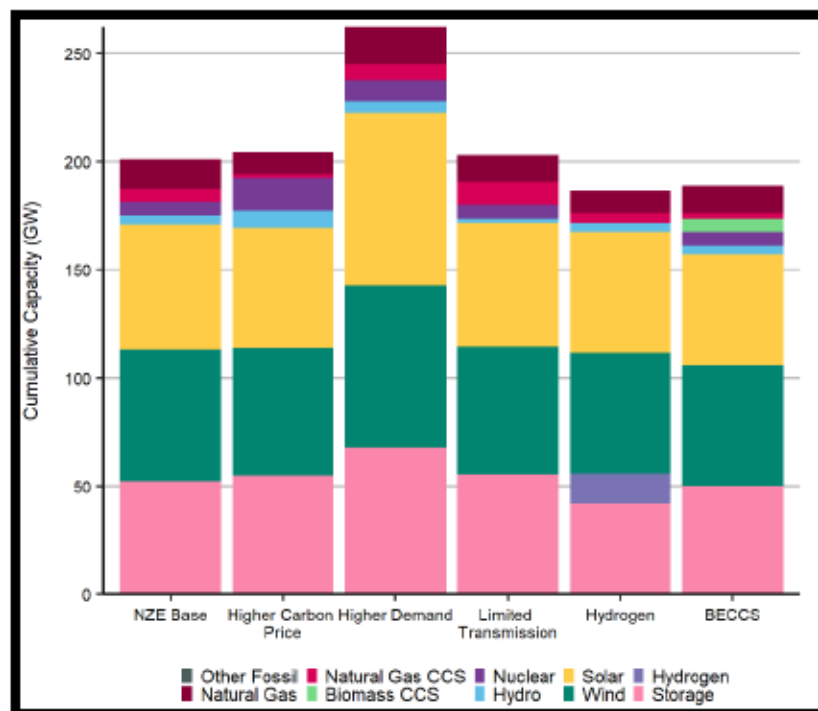


Figure 12 Ajouts cumulatifs de capacités jusqu'en 2050, tous scénarios d'électricité net-zéro [7].

3. Infrastructure énergétique :

L'infrastructure énergétique au Canada est un élément clé de son économie et de son engagement envers un avenir énergétique durable. Cette infrastructure se compose principalement du réseau électrique, des pipelines pour le transport du pétrole et du gaz, et des installations de stockage d'énergie.

3.1. Réseau électrique : Interconnexions, fiabilité et défis :

Le Canada est fortement intégré dans le marché nord-américain de l'électricité, avec 98 % des Canadiens connectés au système d'énergie en vrac du continent. Cette interconnectivité robuste, comprenant 37 lignes de transmission transfrontalières nord-sud, renforce la fiabilité et la résilience du réseau canadien face aux variations saisonnières ou imprévues de l'offre et de la demande. La fiabilité du réseau est assurée par les normes, pratiques et procédures de la 'Société Nord-Américaine de Fiabilité de l'Électricité' (NERC), adoptées par la plupart des provinces canadiennes [50].

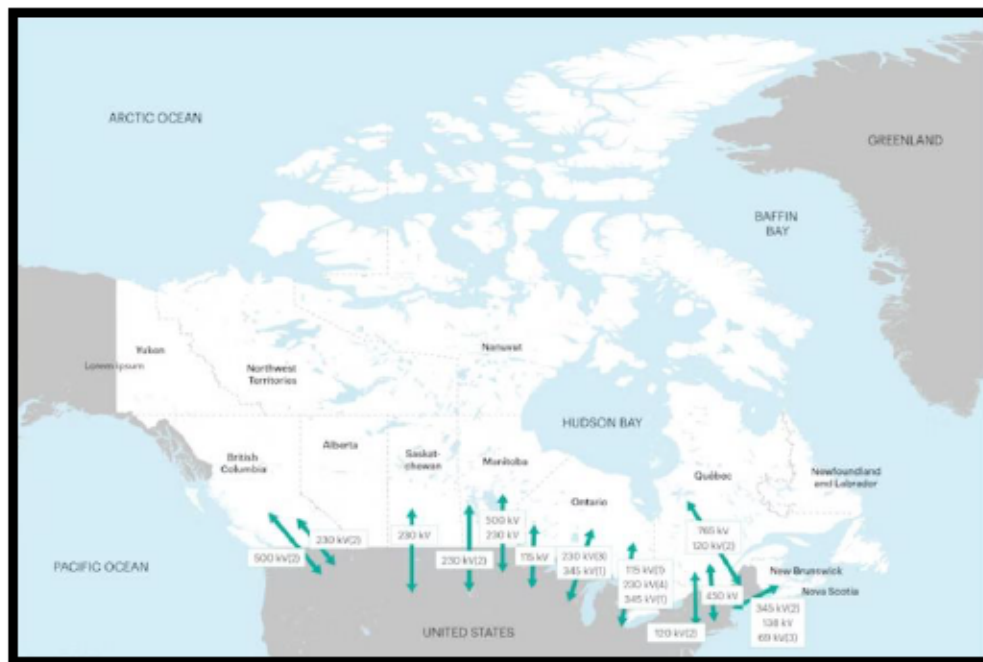


Figure 13 Lignes de transport transfrontalières du Canada, 2022 [8].

Le Canada bénéficie de l'un des systèmes électriques les plus propres au monde, avec une majorité de sa production en 2019 provenant de sources non émettrices, telles que l'hydroélectricité, d'autres renouvelables et le nucléaire. L'objectif à long terme de zéro émission nette d'ici 2050 nécessitera d'électrifier plusieurs secteurs dominés par les combustibles fossiles, tout en intégrant davantage de renouvelables variables [8].

3.2. Pipelines et transport du pétrole et du gaz : Principaux corridors, enjeux politiques et environnementaux :

Le transport du pétrole et du gaz naturel au Canada est principalement effectué par pipelines, qui traversent le pays d'est en ouest et du nord au sud. Ces pipelines sont essentiels pour acheminer les hydrocarbures des sites de production vers les marchés intérieurs et d'exportation. Toutefois, l'expansion des réseaux de pipelines soulève des défis environnementaux et politiques, notamment en ce qui concerne les préoccupations des communautés autochtones, la protection des écosystèmes sensibles et la contribution aux émissions de gaz à effet de serre.

3.3. Stockage d'énergie et technologies émergentes :

Le stockage de l'énergie est devenu une composante cruciale de l'infrastructure énergétique, permettant de gérer la variabilité des sources d'énergie renouvelables comme le solaire et l'éolien. Le Canada explore diverses solutions de stockage, y compris les batteries, le stockage par pompage et d'autres technologies émergentes. Ces technologies sont essentielles pour améliorer la fiabilité du réseau électrique et faciliter l'intégration des renouvelables dans le mix énergétique.

Les interconnexions avec les États-Unis fournissent à la fois au Canada et à ses voisins du sud de nombreux avantages, notamment une fiabilité électrique accrue, une sécurité, une abordabilité et une résilience améliorée, ainsi que des avantages économiques accrus. Grâce à son mix électrique comparativement propre, l'augmentation des exportations canadiennes pourrait aider les États-Unis, ainsi que les États et les villes individuels, à atteindre leurs objectifs d'énergie propre [51].

4. Relation Internationales et commerce énergétique :

Les relations internationales et le commerce énergétique du Canada, en particulier avec les États-Unis, jouent un rôle crucial dans son économie énergétique. L'interdépendance énergétique entre le Canada et les États-Unis est renforcée par l'Accord de libre-échange nord-américain (ALENA), désormais remplacé par l'Accord Canada-États-Unis-Mexique (ACEUM). Cette relation est marquée par une intégration profonde des marchés énergétiques, qui influence à la fois la sécurité énergétique et la dynamique économique des deux pays.

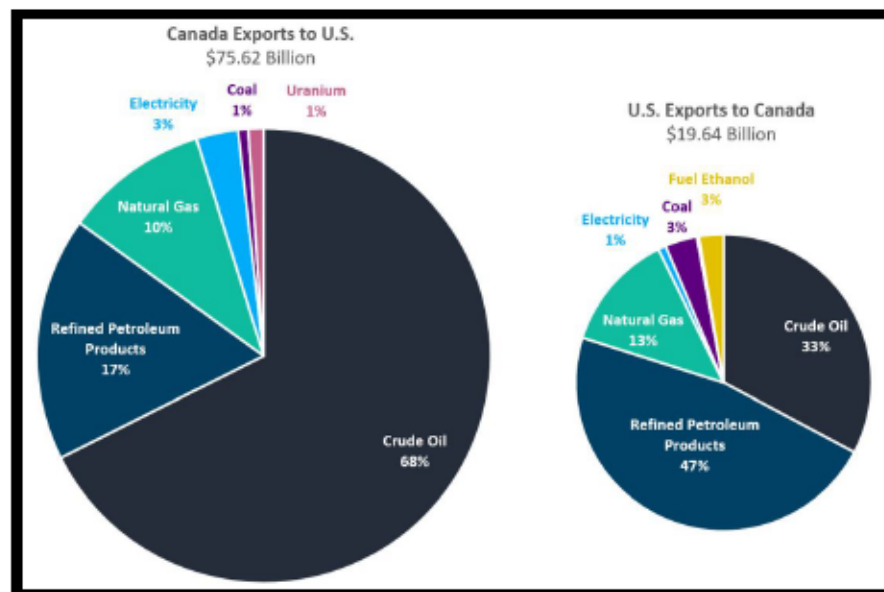


Figure 14 Répartition du commerce de l'énergie au Canada [53].

4.1. Le Canada et les USA : Interdépendance énergétique et impact de l'ALENA :

L'ALENA, et plus récemment l'ACEUM, a favorisé une augmentation significative des échanges commerciaux de l'énergie entre le Canada et les États-Unis, faisant d'eux des partenaires commerciaux essentiels l'un pour l'autre dans le secteur énergétique. Le Canada est un exportateur net d'énergie vers les États-Unis, avec des exportations

importantes en pétrole brut, produits pétroliers raffinés, gaz naturel et électricité. Alberta se distingue comme le centre de la production de pétrole et de gaz du Canada, avec un volume d'échange commercial énergétique significatif avec les États-Unis [53].

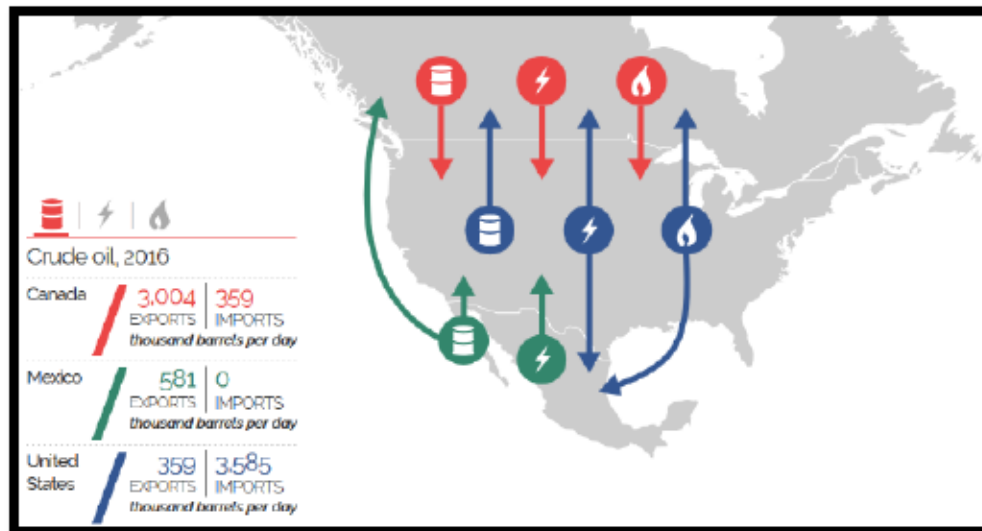


Figure 15 Flux d'Énergie en Amérique du Nord [54].

4.2. Exportations d'énergie : Marchés clés, défis logistiques, et perspectives :

Les exportations énergétiques du Canada vers les États-Unis comprennent principalement du pétrole brut, des produits pétroliers raffinés, du gaz naturel et de l'électricité. Ces échanges sont facilités par un vaste réseau de pipelines et d'infrastructures transfrontaliers, soulignant l'importance géographique de la proximité pour le commerce énergétique bilatéral. Les défis logistiques, tels que le développement et l'expansion des infrastructures de transport énergétique, ainsi que les préoccupations environnementales, jouent un rôle clé dans le façonnement de ce commerce [53].

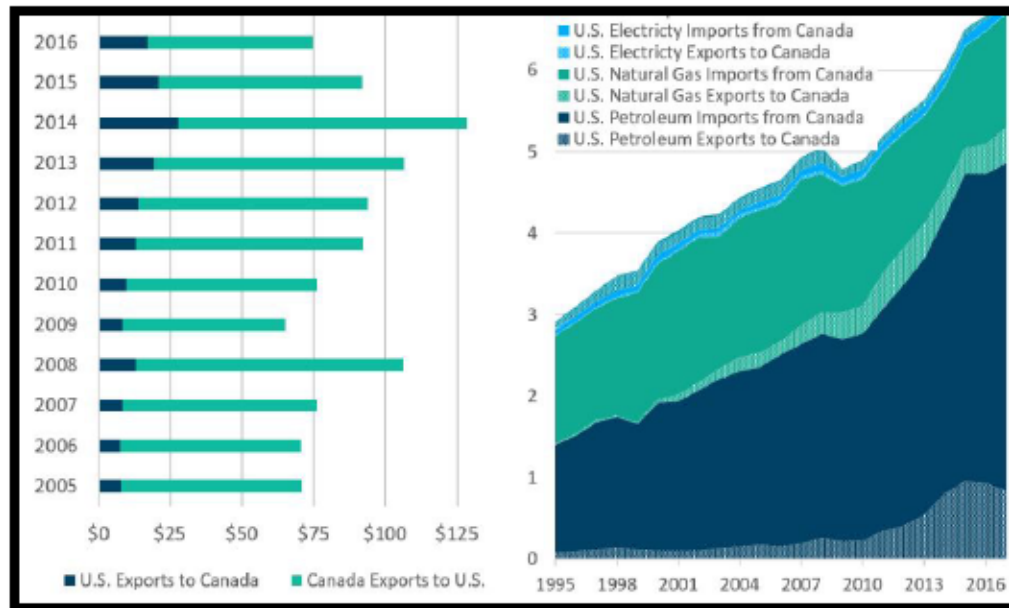


Figure 16 Valeur (Billion-USD) et volumes (Méga barils équivalents pétrole /j) du commerce énergétique entre les États-Unis et le Canada [53].

4.3. Coopération internationale : Participation du Canada aux forums énergétiques mondiaux et initiatives multilatérales :

Le Canada joue un rôle actif dans les forums énergétiques mondiaux et les initiatives multilatérales, s'engageant dans des efforts de collaboration pour aborder les défis énergétiques mondiaux, la sécurité énergétique et la transition vers des systèmes énergétiques durables. Son engagement dans des accords internationaux sur le climat, tels que l'Accord de Paris, souligne l'importance de la coopération internationale dans la réalisation de ses objectifs de réduction des émissions de gaz à effet de serre et de transition énergétique [54].

L'intégration économique et énergétique avec les États-Unis, renforcée par des décennies de commerce et de collaboration, ainsi que la participation active du Canada sur la scène internationale, sont essentielles pour comprendre le paysage énergétique et les politiques énergétiques du Canada. Ces éléments contribuent non seulement à la

sécurité énergétique du Canada mais également à sa capacité à influencer et à s'adapter aux dynamiques énergétiques mondiales.

5. Défis et Opportunités :

5.1. Enjeux environnementaux :

Les défis environnementaux associés à l'expansion des combustibles fossiles, en particulier les développements de gaz naturel liquéfié (GNL), attirent l'attention des militants et des écologistes concernés par les impacts écologiques de tels projets. Les préoccupations environnementales comprennent non seulement les émissions de gaz à effet de serre (GES) mais aussi la gestion de l'eau et la conservation de la biodiversité. Les initiatives visant à minimiser les émissions fugitives lors de la séparation du sable après la fracturation hydraulique et l'utilisation de capteurs de haute précision pour détecter les fuites de méthane des puits inactifs ou orphelins sont des exemples d'efforts en cours pour atténuer les impacts environnementaux dans le secteur énergétique canadien [55].

5.2. Innovations technologiques :

L'intégration des technologies avancées, comme l'intelligence artificielle (IA) et les modèles prédictifs tels que ANN-ARIMA, joue un rôle croissant dans l'amélioration de l'efficacité énergétique et la gestion des ressources. Ces technologies offrent des opportunités pour optimiser les opérations énergétiques, réduire les coûts et minimiser les impacts environnementaux. Par exemple, l'application de systèmes de détection rapide et de classification des événements peut aider à surmonter les défis liés à la transition des prototypes à un environnement de production, en tenant compte des objectifs divers des parties prenantes [56].

5.3.Perspective :

Le paysage énergétique canadien continue d'évoluer avec des avancées dans les énergies renouvelables et la mise en œuvre de pratiques plus durables. La recherche sur le stockage d'énergie comprimée dans des milieux géologiques souterrains, comme le montre l'expérience de la Chine et du Canada, indique des progrès significatifs qui pourraient résoudre certains problèmes environnementaux associés aux sources d'énergie traditionnelles. Ces avancées, combinées à une sensibilisation accrue aux impacts environnementaux et à l'engagement en faveur de la durabilité, ouvrent la voie à un avenir énergétique plus propre et plus durable pour le Canada [55-56].

En résumé, le secteur énergétique canadien fait face à de nombreux défis environnementaux qui nécessitent une attention continue et des efforts d'atténuation. Les innovations technologiques, en particulier dans les domaines de l'IA et des modèles prédictifs, offrent des opportunités substantielles pour aborder ces défis de manière efficace. À mesure que le Canada progresse vers une économie à faible émission de carbone, les initiatives visant à améliorer la durabilité et à réduire les impacts environnementaux seront cruciales pour le futur paysage énergétique du pays.

Conclusion :

Ce chapitre mettant en avant la diversité des ressources énergétiques, incluant le pétrole, le gaz naturel, les énergies renouvelables, et l'hydroélectricité.

Les politiques gouvernementales se concentrent sur la réduction des émissions de gaz à effet de serre tous en soutenant la croissance économique durable.

CHAPITRE 4 : DE L'ANALYSE DES SÉRIES CHRONOLOGIQUES VERS LA PRÉVISION PAR L'APPRENTISSAGE MACHINE

Introduction :

Ce chapitre a pour objectif d'analyser la production d'énergie électrique dans les provinces du Canada on se basant sur des séries chronologiques, l'analyse nous permettra également de prévoir l'évolution future de cette production dans le but de mieux gérer les ressources énergétiques et de réduire les impacts négatifs sur l'environnement [26].

Comme annoncé dans le chapitre 3, qui traite la diversité des ressources énergétiques, et étant donné que Canada est un pays vaste, nous allons procéder à une division en Zones Géographiques.



Figure 17 Les provinces du Canada

La division est faite par province :

- Zone 1 : Canada en ensemble
- Zone 2 : Terre-Neuve-et-Labrador
- Zone 3 : Île-du-Prince-Édouard
- Zone 4 : Nouvelle-Écosse
- Zone 5 : Nouveau-Brunswick
- Zone 6 : Québec
- Zone 7 : Ontario
- Zone 8 : Manitoba
- Zone 9 : Saskatchewan
- Zone 10 : Alberta
- Zone 11 : Colombie-Britannique
- Zone 12 : Yukon
- Zone 13 : Territoires du Nord-Ouest
- Zone 14 : Nunavut

Les sources d'énergie les plus abondantes ne se trouvent pas là où il y a une grande demande d'énergie, chaque province travaille à accroître l'utilisation des énergies renouvelables, y compris l'éolien, le solaire, géothermique et la biomasse, afin de diversifier les sources d'énergie et d'augmenter l'autonomie énergétique des territoires.

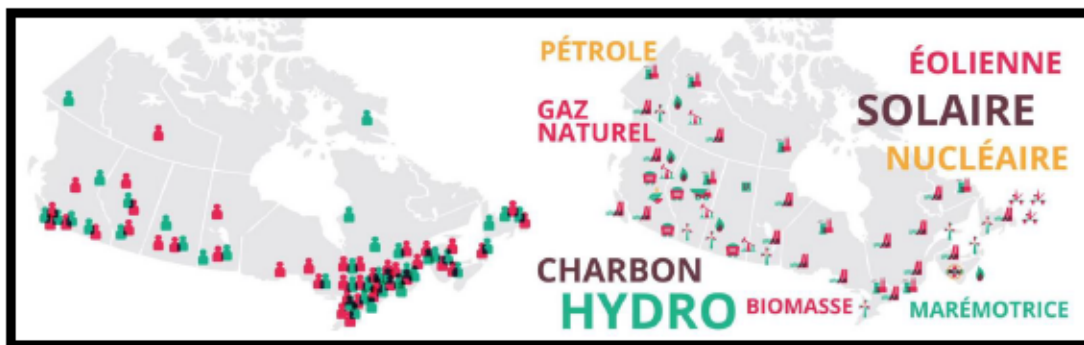


Figure 18 a- Les ressources énergétiques Canadiennes. b- Répartition de la population Canadienne.

Chaque Zone possède des particularités énergétiques basées sur ses ressources naturelles (comme les rivières pour l'énergie hydraulique ou les zones venteuses pour l'éolien) et ses infrastructures.

1. Données Énergétiques :

➤ Production totale d'énergie électrique:

La Production Totale d'Énergie dans chaque zone est la somme des différentes sources d'énergie disponible :

$$P_{Totale} = P_{Hydraulique} + P_{Vapeur_Classique} + P_{Nucleaire} + P_{Combustion} + P_{Autres}$$

➤ Électricité disponible pour l'utilisation :

L'Électricité Disponible est ce qu'il reste après déduction des pertes de transmission et de distribution :

$$E_{Disponible} = P_{Totale} - P_{Pertes}$$

➤ Energie totale reçue et livrée :

Les provinces échangent de l'énergie entre elles et avec d'autres pays. L'Énergie Nette est la différence entre l'énergie reçue et celle livrée :

$$E_{Net} = E_{Recue} - E_{Livree}$$

➤ Énergie électrique vendue et achetée des États-Unis :

Les échanges avec les États-Unis sont également modélisés pour certaines provinces, ou l'énergie est soit vendue, soit achetée :

$$E_{Net_US} = E_{Vendue_US} - E_{Achete_US}$$

2. Modes de Production d'Énergie Électrique:

Les modes de production d'énergie varient d'une province à l'autre en fonction des ressources naturelles disponibles.

Voici les principaux types de production d'énergie qui seront analysés :

➤ **Turbine Hydraulique :**

L'Énergie Hydraulique est produite par la force de l'eau. Le modèle de la production hydraulique est souvent fonction du temps et des précipitations.

$$P_{Hydraulique} = f(t)$$

➤ **Turbine à Vapeur Classique :**

La production par turbine à Vapeur utilise des combustibles fossiles (comme le charbon ou le gaz naturel) pour générer de l'électricité :

$$P_{Vapeur_Classique} = f(t, \text{Prix du charbon}, \text{Prix du gaz})$$

➤ **Turbine Nucléaire à Vapeur :**

L'Énergie Nucléaire est également plus stable et moins sujette aux variations saisonnières. Elle peut être modélisée ainsi :

$$P_{Nucleaire} = \text{Constante} + \epsilon_t$$

➤ **Turbine à Combustion Interne :**

Les Turbines à Combustion Interne, souvent utilisées dans des régions éloignées, produisent de l'électricité en fonction des besoins locaux :

$$P_{Combustion_Interne} = f(t, \text{Demandes locales})$$

➤ **Turbine à Combustion (Gaz) :**

Les Turbines à Combustion fonctionnant au gaz suivent un modèle basé sur les prix du gaz et la demande énergétique :

$$P_{Combustion_Gaz} = f(t, \text{Prix du gaz}, \text{Demande energetiques})$$

➤ **Production d'électricité à partir de combustibles fossiles :**

La production d'électricité à partir des Combustibles Fossiles regroupe plusieurs types de production (vapeur classique, combustion interne, etc.) :

$$P_{Combustibles} = P_{Vapeur_Classique} + P_{Combustion_Interne} + P_{Combustion_Gaz}$$

➤ **Hydrolienne:**

Les Hydroliennes utilisent les courants marins pour générer de l'énergie :

$$P_{Hydrolienne} = f(t, \text{Courant maritime})$$

➤ **Éolienne:**

L'Énergie Éolienne dépend de la vitesse du vent et modélisée en fonction du temps et des conditions météorologiques :

$$P_{Eolienne} = f(t, \text{Vitesse du vent})$$

➤ **Solaire:**

L'Énergie Solaire est fonction de l'ensoleillement, qui varie selon les saisons et la région géographique :

$$P_{Solaire} = f(t, \text{Ensoleillement})$$

➤ **Autres types de productions d'électricité :**

D'autres sources d'énergie incluent la biomasse, la géothermie, et d'autres technologies émergentes :

$$P_{Autres} = f(t, \text{Facteurs environnementaux et locaux})$$

3. Analyse des Séries Chronologiques :

L'analyse des Séries Chronologiques est une méthode statistique utilisée pour analyser des données collectées au fil du temps afin de déceler des tendances, des cycles, des motifs saisonniers, ou des relations à long terme.

3.1. Collecte et visualisation des données :

La série chronologique est notée $\{Y_t\}$, avec t représente le temps et Y_t la Valeur observée à l'instant t , les graphes de la série montre la relation entre la valeur observée à un moment spécifique [27].

Les figures suivantes représentent l'Énergie électrique produite (Figure 19), livré (Figure 21) et reçue (Figure 20) par chaque province du Canada entre la période de janvier 2008 et juin 2022 mensuellement.

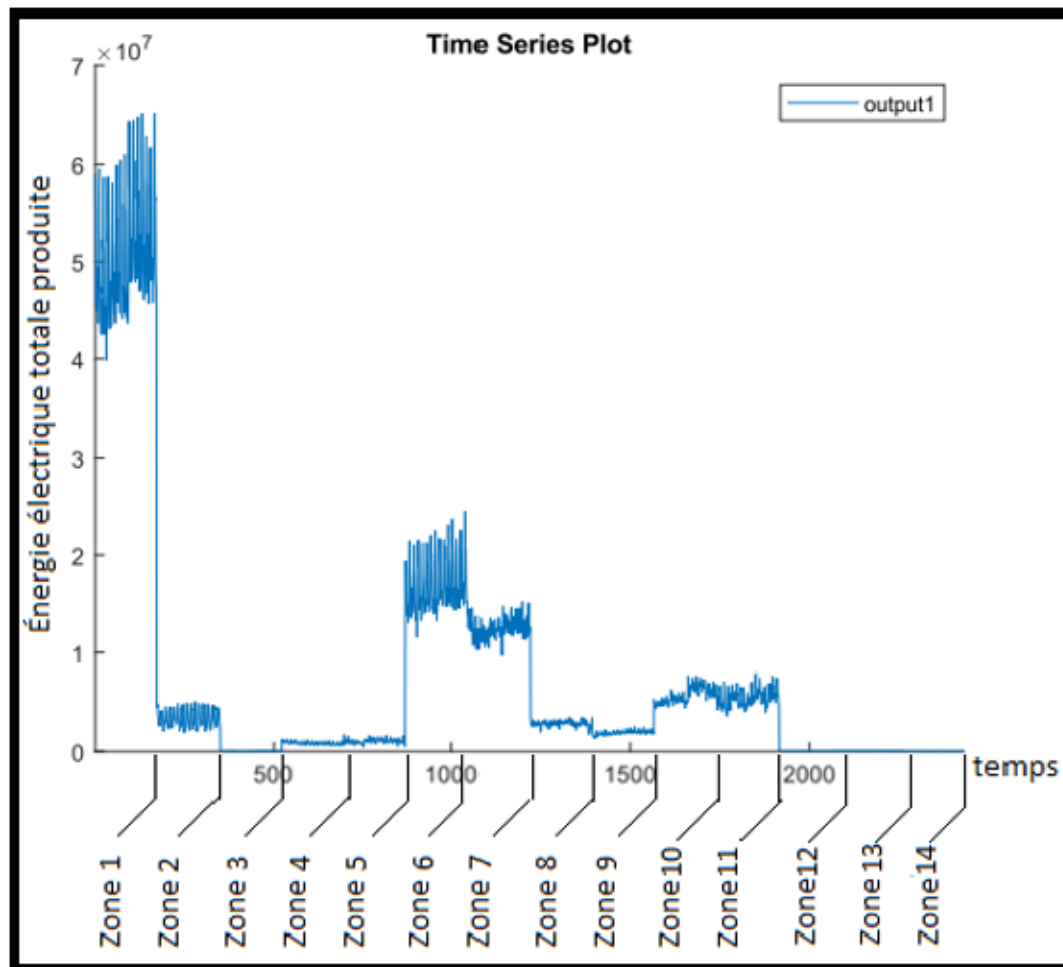


Figure 19 Énergie électrique totale produite par chaque Zone

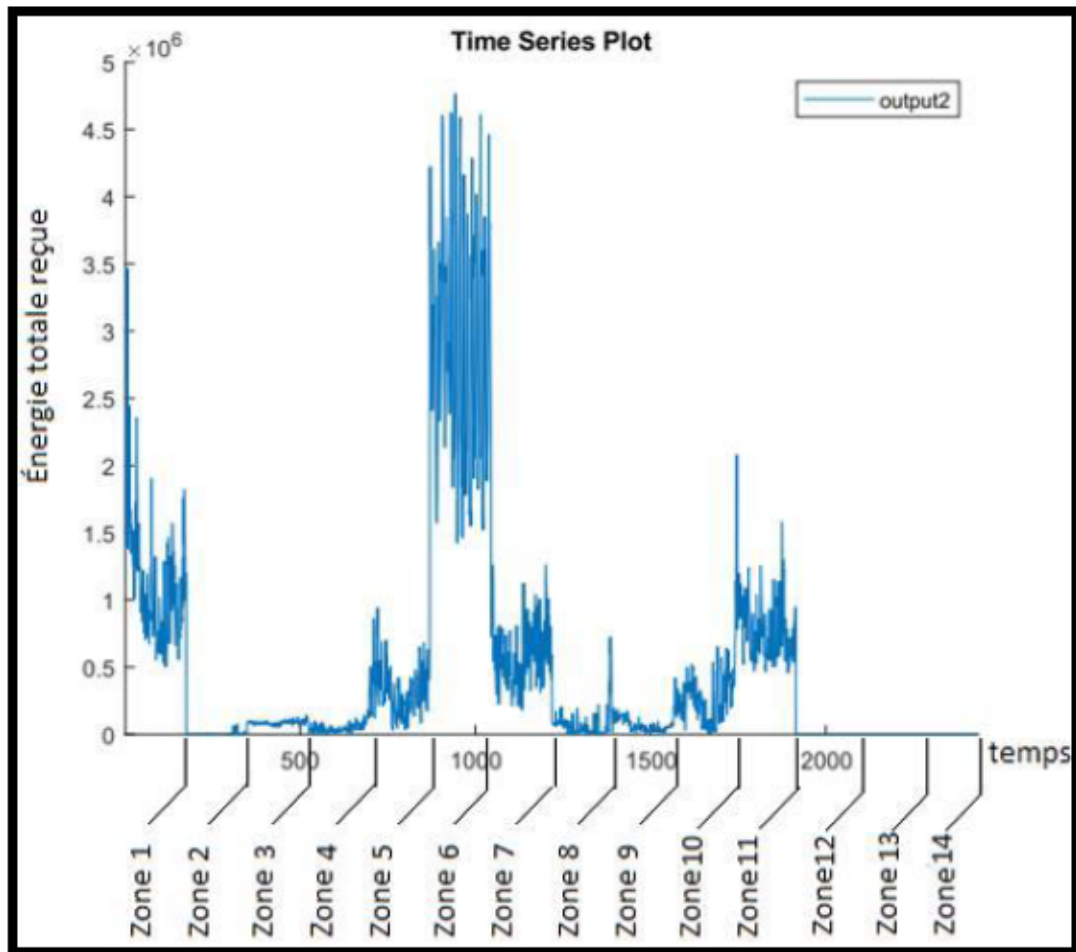


Figure 20 Énergie totale reçue par chaque Zone

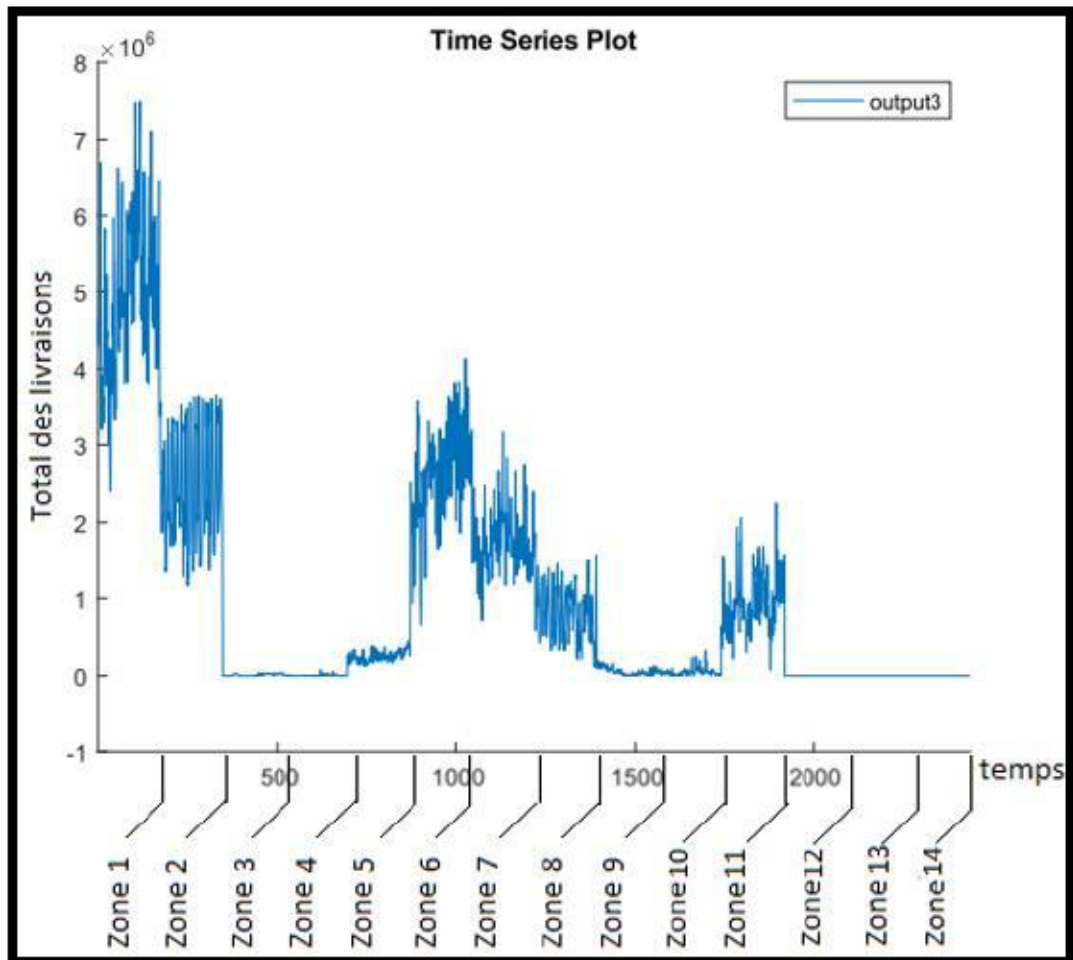


Figure 21 Énergie totale livrée par chaque Zone

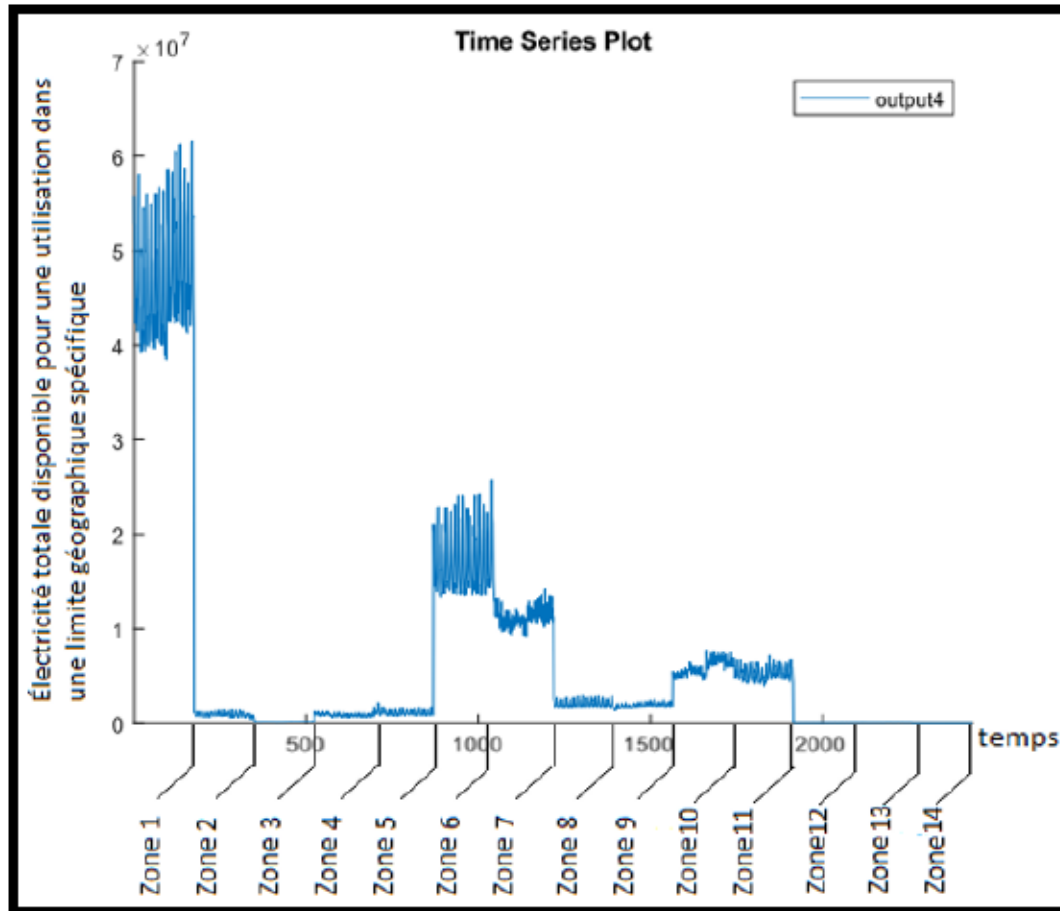


Figure 22 Énergie totale disponible pour une utilisation dans une limite géographique spécifique

3.2. Décomposition de la série :

La décomposition d'une série chronologique est souvent représentée par le modèle suivant :

$$Y_t = T_t + S_t + C_t + R_t$$

- Où :
- T_t : Représente la Tendence,
 - S_t : Représente la Saisonnalité,
 - C_t : Représente les Cycles,
 - R_t : Est le Résidu ou Bruit.

Il existe aussi un modèle Multiplicatif pour des série a forte Variance [28]:

$$Y_t = T_t \times S_t \times C_t \times R_t$$

3.2.1. Estimation de la Tendence :

La tendance peut être estimée à l'aide de différentes méthodes :

- **Moyenne Mobile** : La moyenne mobile d'ordre k est donnée par :

$$T_t = \frac{1}{k} \sum_{i=t-\frac{k-1}{2}}^{t+\frac{k-1}{2}} y_i$$

Où k est la fenêtre de lissage (par exemple 12 mois pour des données mensuelle) [28].

- **Régression linéaire** : Si la tendance est linéaire, on peut ajuster un modèle de la forme :

$$Y_t = \alpha + \beta t + \epsilon_t$$

Où α est l'ordonnée a l'origine, β la pente de la tendance, et ϵ_t l'erreur résiduelle.

3.2.2. Les Cycles :

Les Cycles C_t représentent des fluctuations à long terme. Contrairement à la saisonnalité, les cycles n'ont pas de période fixe. Les cycles peuvent être modélisés en utilisant des outils comme le Filtre de Hodrick-Prescott (HP Filtre), une technique couramment utilisée pour isoler les cycles.

Le filtre HP tente de minimiser l'expression suivante :

$$\min_{T_t} \sum_{t=1}^T \left((Y_t - T_t)^2 + \lambda \sum_{t=2}^{T-1} [(T_{t+1} - T_t) - (T_t - T_{t-1})]^2 \right)$$

Où λ contrôle la souplesse de la tendance T_t . Pour des données trimestrielles, $\lambda = 1600$ est généralement utilisé [29].

3.2.3. Estimation de la Saisonnalité :

La saisonnalité représente les motifs périodiques qui se répètent à intervalles réguliers. Il peut être calculé une fois que la tendance T_t a été estimée.

Si Y_t est une série mensuelle sur plusieurs années, on peut calculer l'indice saisonnier pour chaque mois m (par exemple : Janvier, Février, etc.) en prenant la moyenne des écarts pour chaque mois au fil des années.

Pour un modèle additif, on peut calculer les indices saisonniers comme :

$$S_m = \frac{1}{n} \sum_{t=1}^n (Y_{t+m} - T_{t+m})$$

Où n est le nombre de cycles dans la série [29].

3.2.1. Estimation de Résidu:

Le résidu ou bruit R_t est la composante aléatoire, c'est-à-dire la partie de la série qui ne peut pas être expliquée par les autres composantes. Une fois les composantes T_t , S_t et C_t extraites, le résidu est :

- Model Additive : $R_t = Y_t - (T_t + S_t + C_t)$
- Model Multiplicatif : $R_t = \frac{Y_t}{T_t \times S_t \times C_t}$

3.3. Introduction du SARIMA, un outil d'analyse et de prévision des séries chronologiques :

Lorsque la série présente des schémas saisonniers réguliers qui doivent être pris en compte en plus des relations temporelles classiques (auto-régression et moyenne

mobile). L'outil SARIMA (Moyenne mobile intégrée autorégressive saisonnière) est applicable pour modéliser et prévoir des séries chronologiques [19].

L'équation représentative simplifier de SARIMA (p, d, q)(P, D, Q) est :

$$(1 - \varphi_1 L)Y_t = c + (1 + \theta_1 L)\varepsilon_t$$

Avec L Variable aléatoire.

- p est le nombre d'ordres autorégressifs dans le modèle.
- La différence (d) indique l'ordre de différenciation
- La moyenne mobile (q) indique comment l'écart des valeurs passées par rapport à la moyenne de la série
- P, D, Q sont les ordres du modèle saisonnier.

Plus de détails sur le modèle SARIMA sera être traiter dans le chapitre suivant.

Remplies modéliser une série chronologique avec SARIMA, certaines conditions doivent être remplies :

- ✓ Présence de saisonnalité :
 - La série chronologique doit montrer des motifs saisonniers clairs.
 - Une analyse de la fonction d'autocorrélation ACF est essentiel pour observer des pics significatifs a des intervalles réguliers, indiquant une saisonnalité.
- ✓ Stationnarité de la série :
 - La série doit être stationnaire, c'est-à-dire que sa moyenne et sa variance ne doivent pas changer avec le temps.
 - Le test de Dickey-Fuller ADF est utiliser pour vérifier la stationnarité. Une p-value inférieure a 0.05 indique que la série est stationnaire.
 - Sinon une différenciation ordinaire ou saisonnière est appliquée pour éliminer la tendance ou la saisonnalité et re rendre la série stationnaire
- ✓ Différenciation saisonnière (D) :

- Est appliquée si les variations autour de la moyenne changent avec le temps pour stabiliser la série.
- ✓ Structure de l'autocorrélation :
- Il doit y avoir des motifs d'autocorrélation qui indiquent que les valeurs actuelles dépendent des valeurs passées.
 - La fonction d'autocorrélation partielle PACF est utilisé pour identifier les termes autorégressifs (p), et la ACF pour déterminer les termes de moyenne mobile (q).

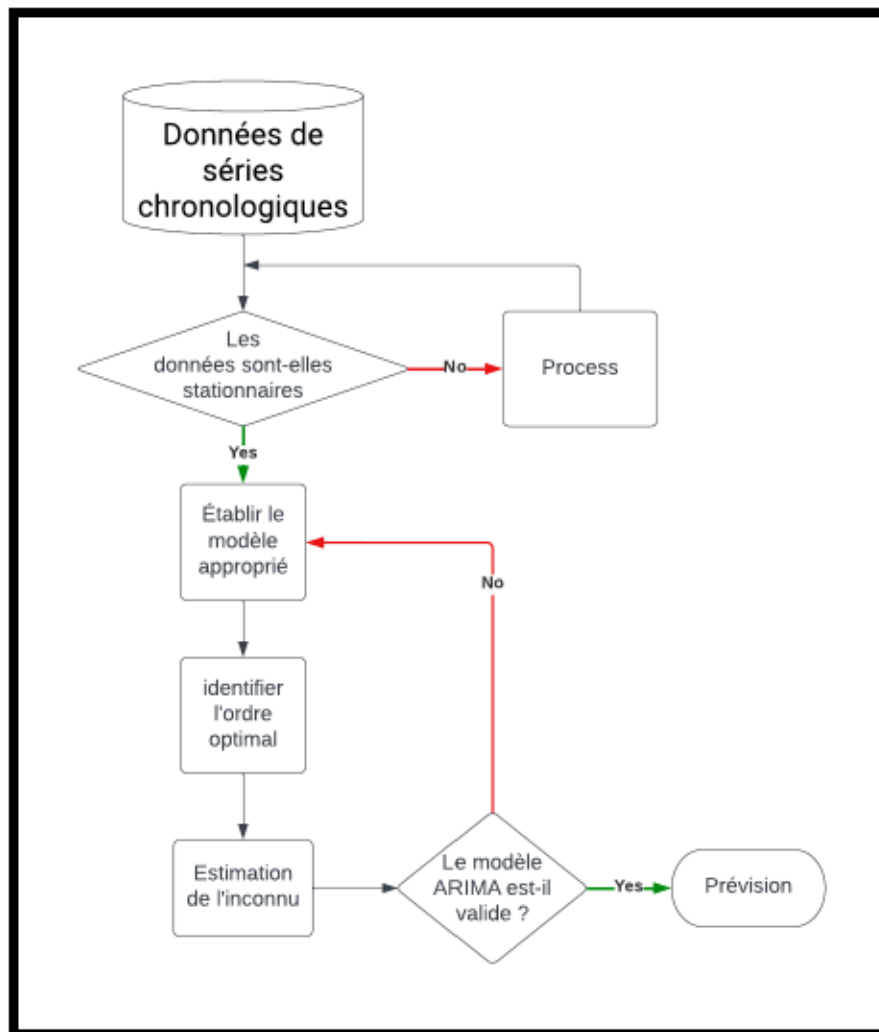


Figure 23 Flux gramme du cycle itératif du Modèle ARIMA saisonnière

3.4. Présentation du travail fait sur MATLAB pour l'application du modèle

SARIMA :

Dans le cadre de cette étude, nous avons utilisé MATLAB comme outil de modélisation et d'analyse pour traiter et prévoir l'évolutions des différentes sources d'Énergie dans la Zone 6 (Québec), Le projet visait principalement à analyser la disponibilité, la production, la livraison et la réception de l'énergie sur une période donnée, tout en prenant en compte les variations temporelles et saisonnières propres à cette zone.

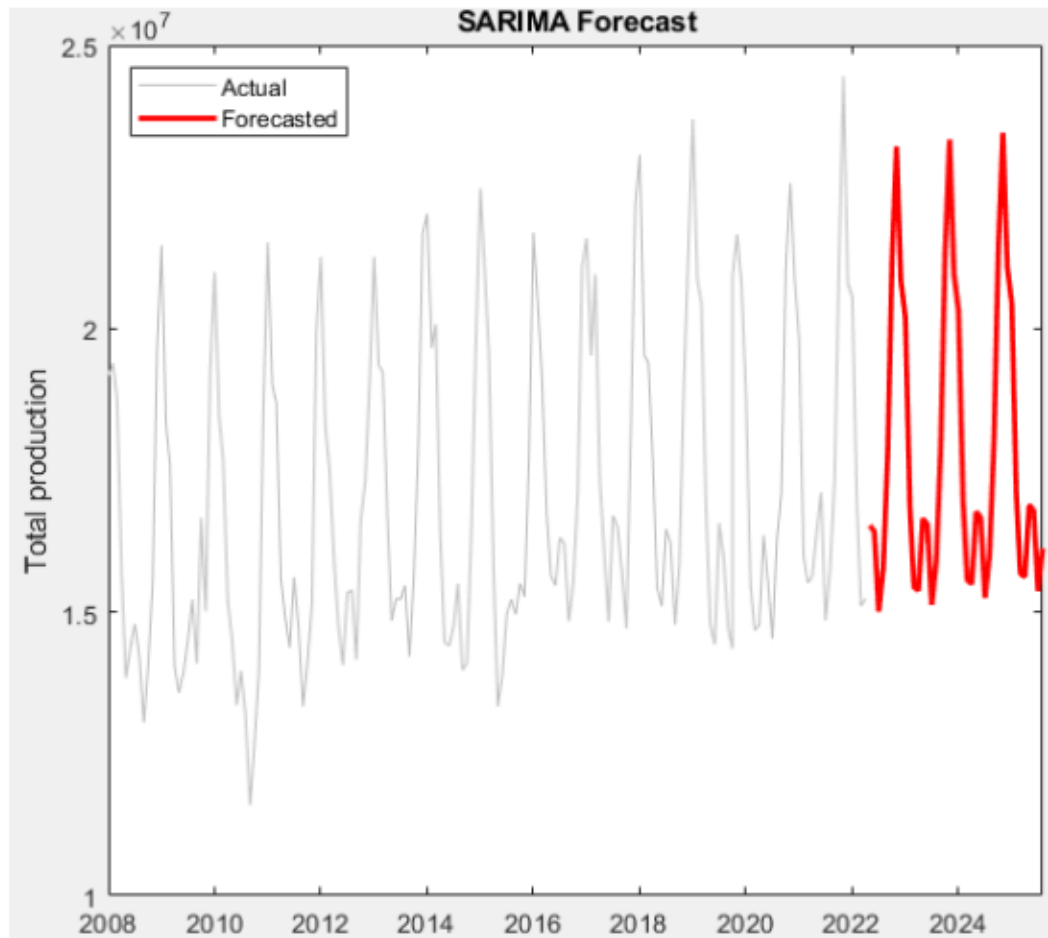


Figure 24 Prévission de la Production totale d'Énergie électrique par SARIMA - Zone 6 Québec

(Figure 24) Le modèle prédit des cycles similaires aux données passées, ce qui suggère que les fluctuations de production vont rester régulières avec des pics similaires à ceux observés historiquement.

SARIMA prévoit que la production totale d'énergie dans cette zone conservera la même structure saisonnière que celle observée dans les données historiques, avec des pics réguliers et des périodes de production plus faibles. Ce modèle semble indiquer une stabilité future dans les tendances de production énergétique.

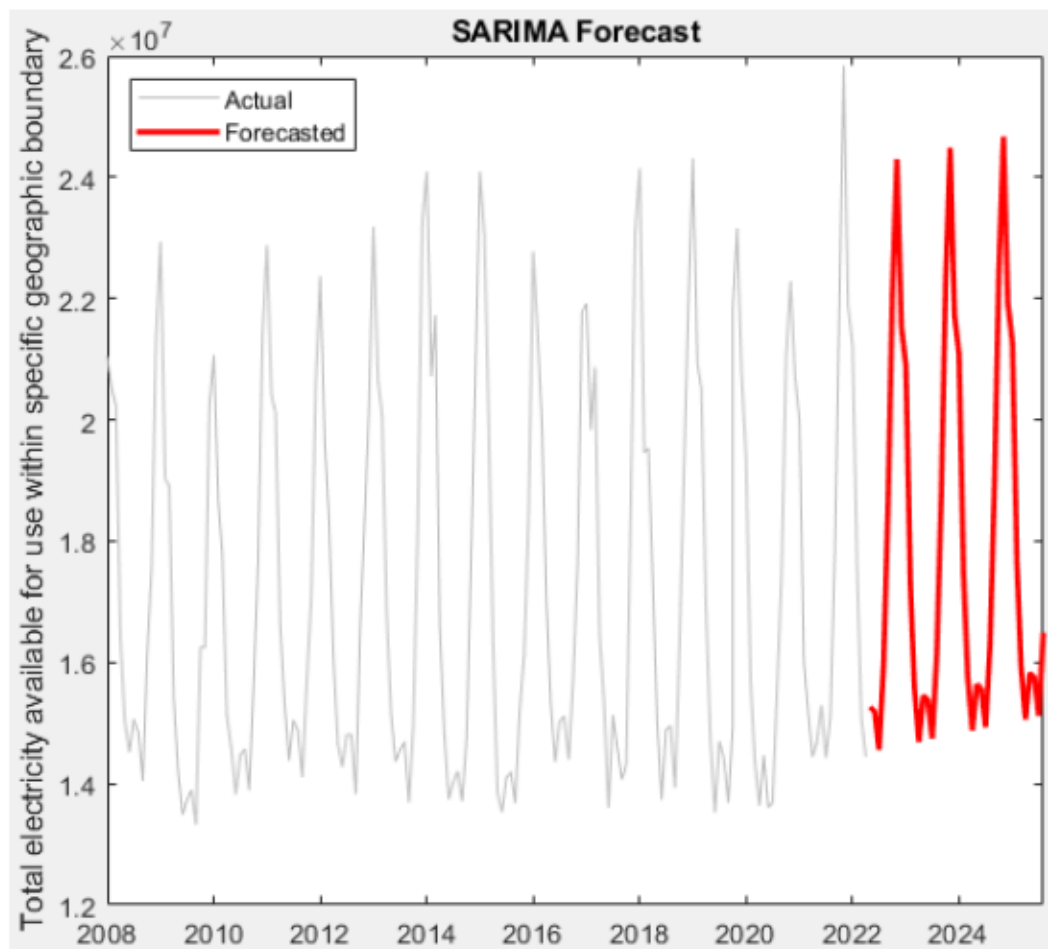


Figure 25 Prévision par SARIMA l'énergie Totale disponible - Zone 6

(Figure 25) Le modèle SARIMA a capturé cette saisonnalité et prévoit une répétition des mêmes motifs pour les années à venir, avec une forte similitude des pics et des

creux, correspondant probablement aux fluctuations saisonnières de la production ou de la demande en électricité.

Le modèle semble bien ajusté pour prévoir les variations annuelles de la disponibilité de l'électricité. SARIMA anticipe des cycles similaires pour les années futures, ce qui indique que les tendances saisonnières passées sont censées se maintenir

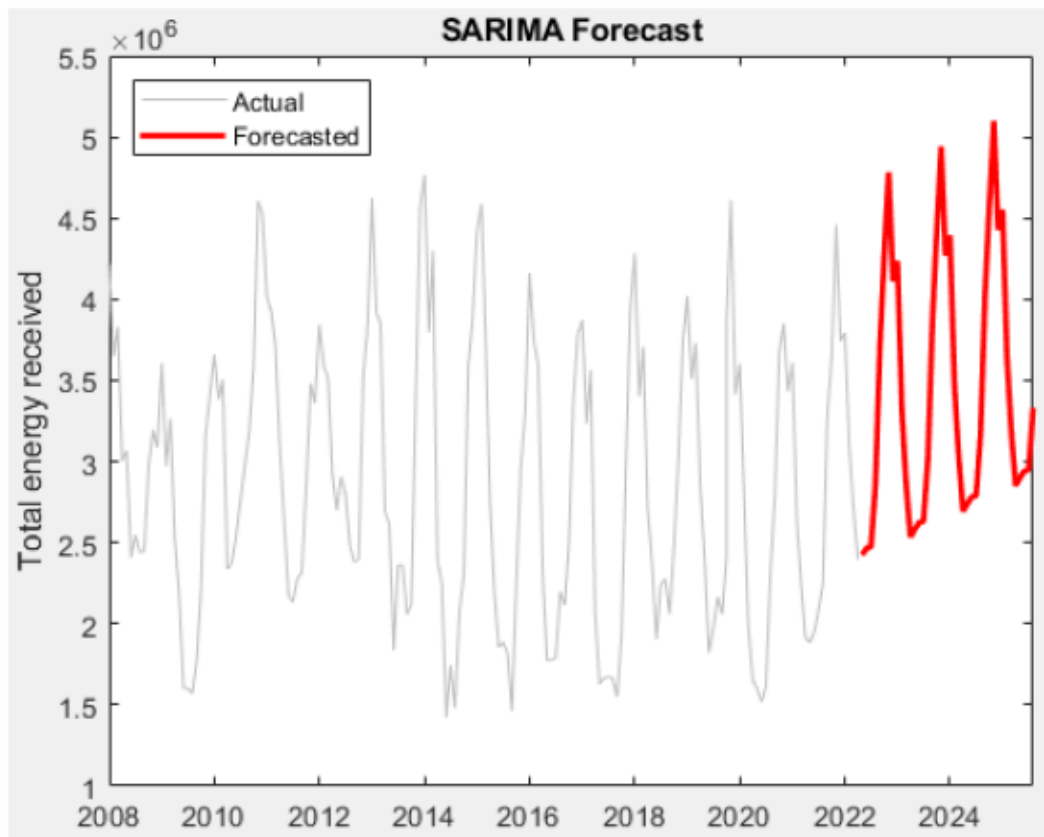


Figure 26 Prévision par SARIMA l'énergie totale reçus - Zone 6

(Figure 26) Le modèle SARIMA prévoit des cycles similaires dans le futur, avec des pics marqués et des creux similaires à ceux observés dans le passé.

Même si les données historiques sont plus volatiles, SARIMA a été capable de capturer la saisonnalité et d'anticiper des fluctuations similaires dans le futur. Cependant, la volatilité pourrait indiquer une incertitude plus élevée dans les prévisions, surtout si les

variations imprévues dans les données passées ont été influencées par des événements aléatoires ou externes.

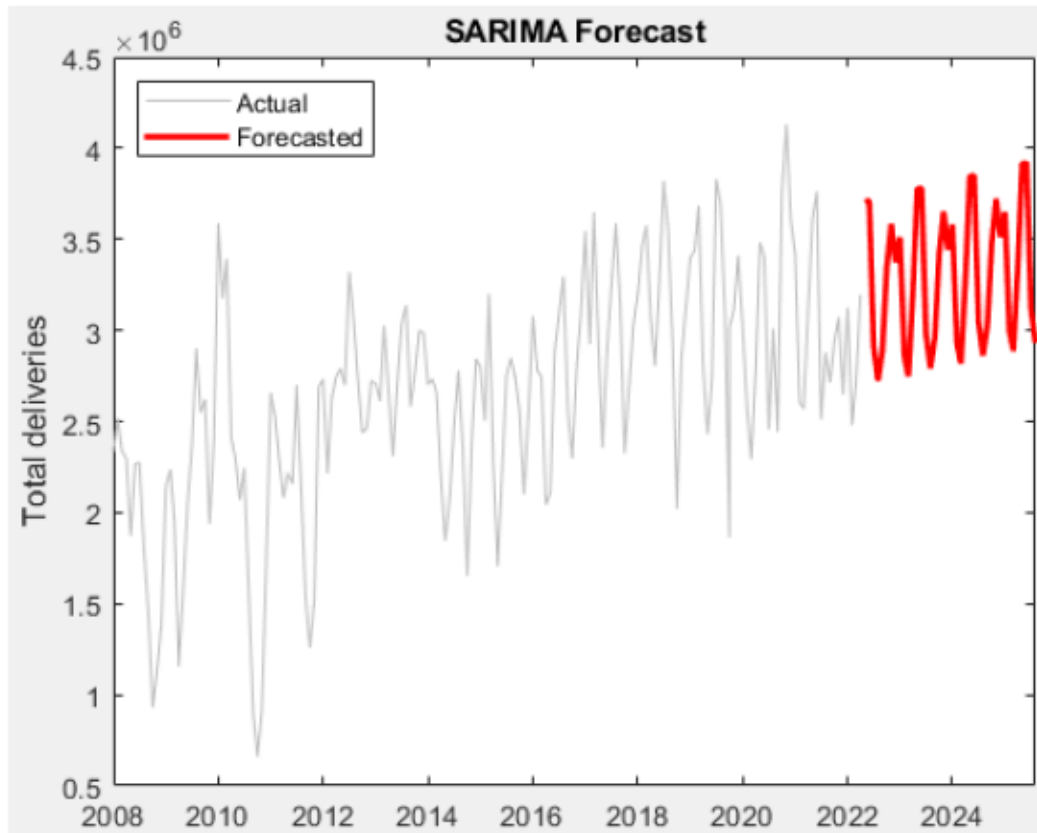


Figure 27 Prévision par SARIMA l'Énergie totale délivrée - Zone 6

Le modèle SARIMA capture à la fois la tendance légèrement croissante et les variations saisonnières. Les prévisions montrent une légère augmentation progressive des livraisons (Figure 27), mais avec des motifs saisonniers encore présents.

Le modèle prévoit que les livraisons totales d'énergie continueront d'augmenter légèrement au fil du temps, avec des fluctuations régulières en fonction des saisons. SARIMA est en mesure de modéliser ces cycles et d'anticiper une croissance modérée.

Dans cette section, nous allons passer en revue les **Différents Types de Production d'Électricité** disponible dans la même Zone 6 (Québec).

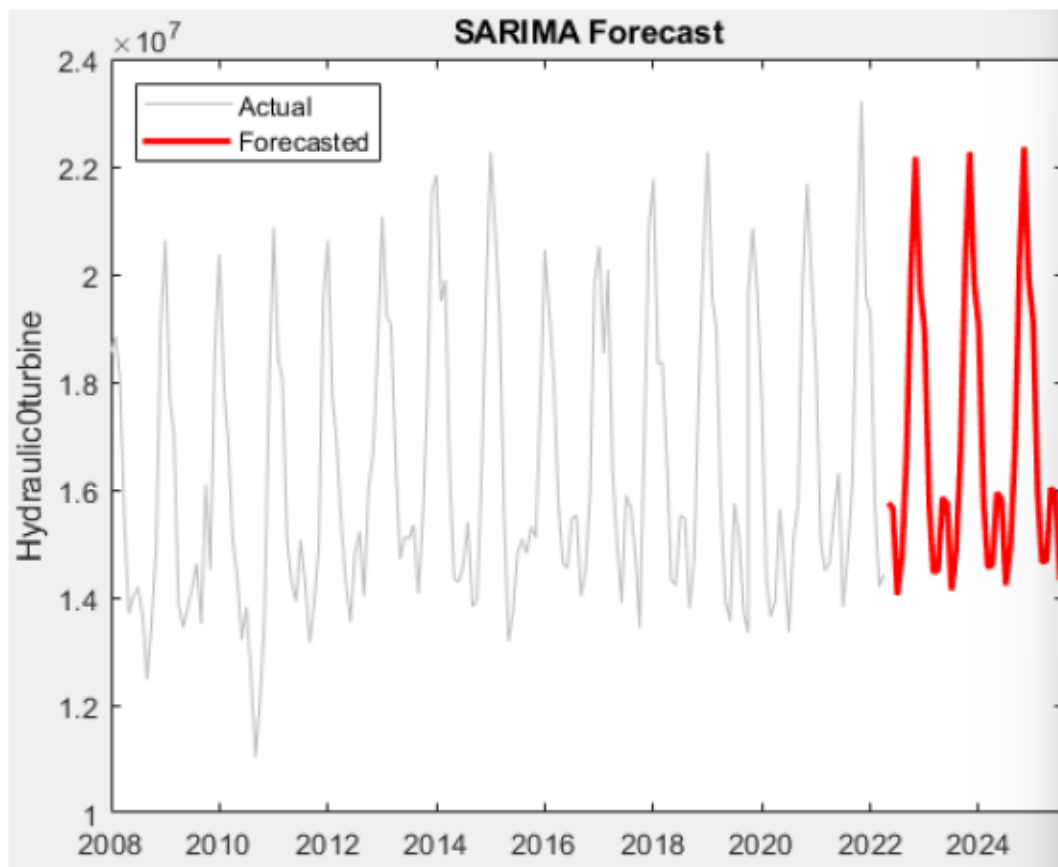


Figure 28 Prédiction de l'énergie Hydraulique par le modèle SARIMA – Zone 6

Le modèle SARIMA anticipe que cette saisonnalité va continuer, avec des fluctuations régulières dans la production hydroélectrique pour les années à venir.

L'énergie hydraulique (Figure 28) est stable et suit des cycles bien établis. La production semble bien modélisée par SARIMA, indiquant que la saisonnalité va rester une caractéristique clé de ce type de production.

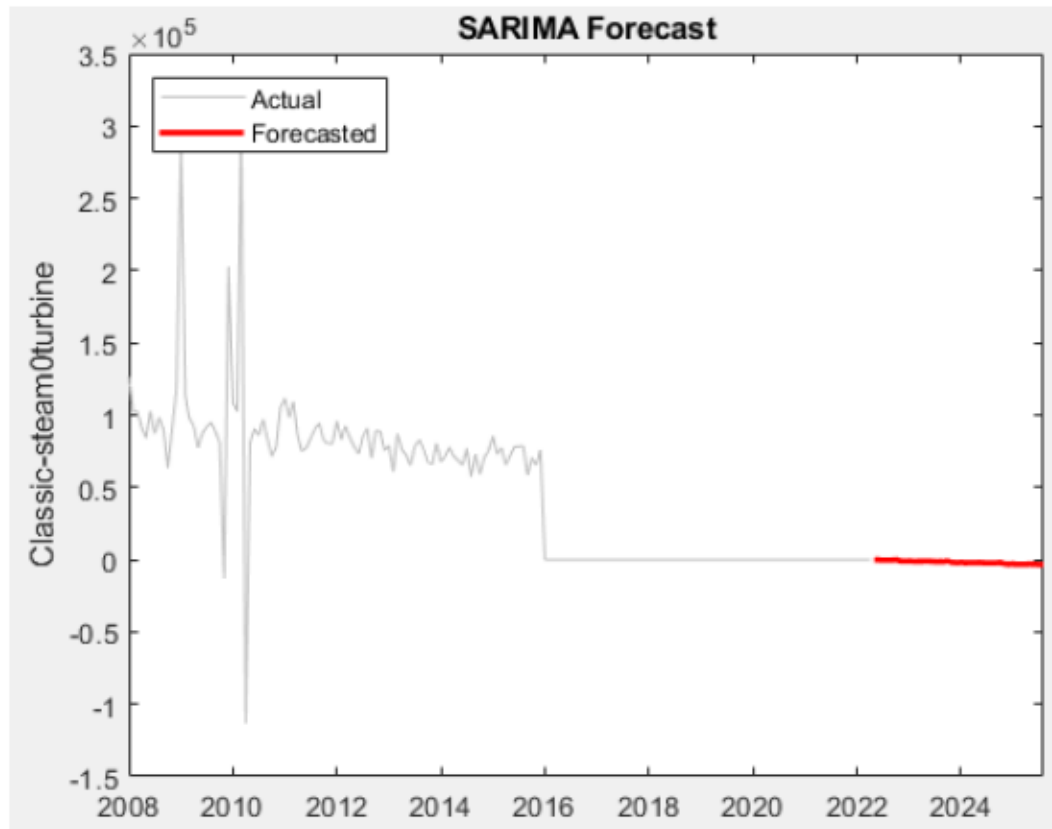


Figure 29 Prédiction de l'énergie Turbine à Vapeur classique par le modèle SARIMA Zone 6

Pour la turbine à vapeur classique (Figure 29) le modèle prévoit une production très faible, voire inexistante, pour les années à venir.

Cela reflète probablement une transition complète ou partielle vers des technologies de production d'électricité plus modernes et plus propres, marquant ainsi l'abandon progressif des turbines à vapeur classiques.

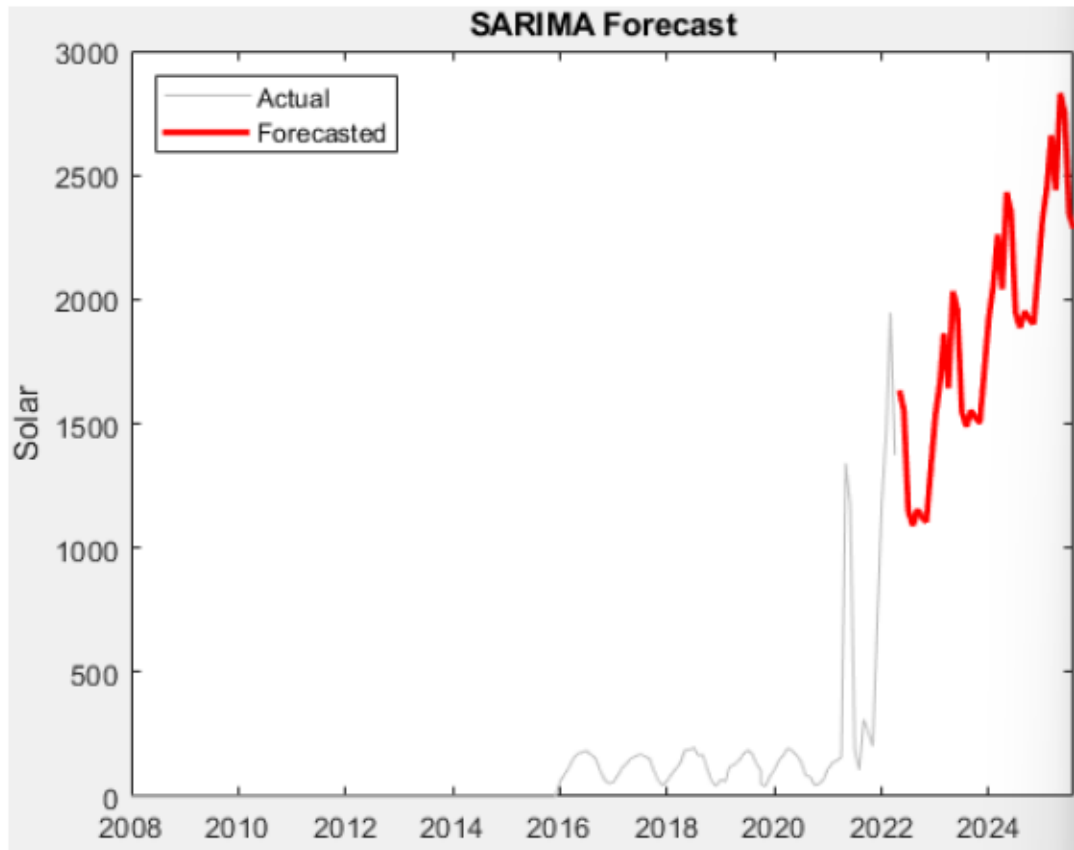


Figure 30 Prédiction de l'énergie Solaire par le modèle SARIMA – Zone 6

Le modèle prédit une croissance soutenue de la production solaire, avec une augmentation marquée à partir de 2020 et des prévisions suggérant une poursuite de cette tendance jusqu'en 2024.

La production d'énergie solaire est en forte croissance (Figure 30), probablement en raison des politiques favorisant les énergies renouvelables. SARIMA prévoit que cette croissance continue de manière significative dans les années à venir, ce qui est cohérent avec les tendances observées dans de nombreuses régions du monde.

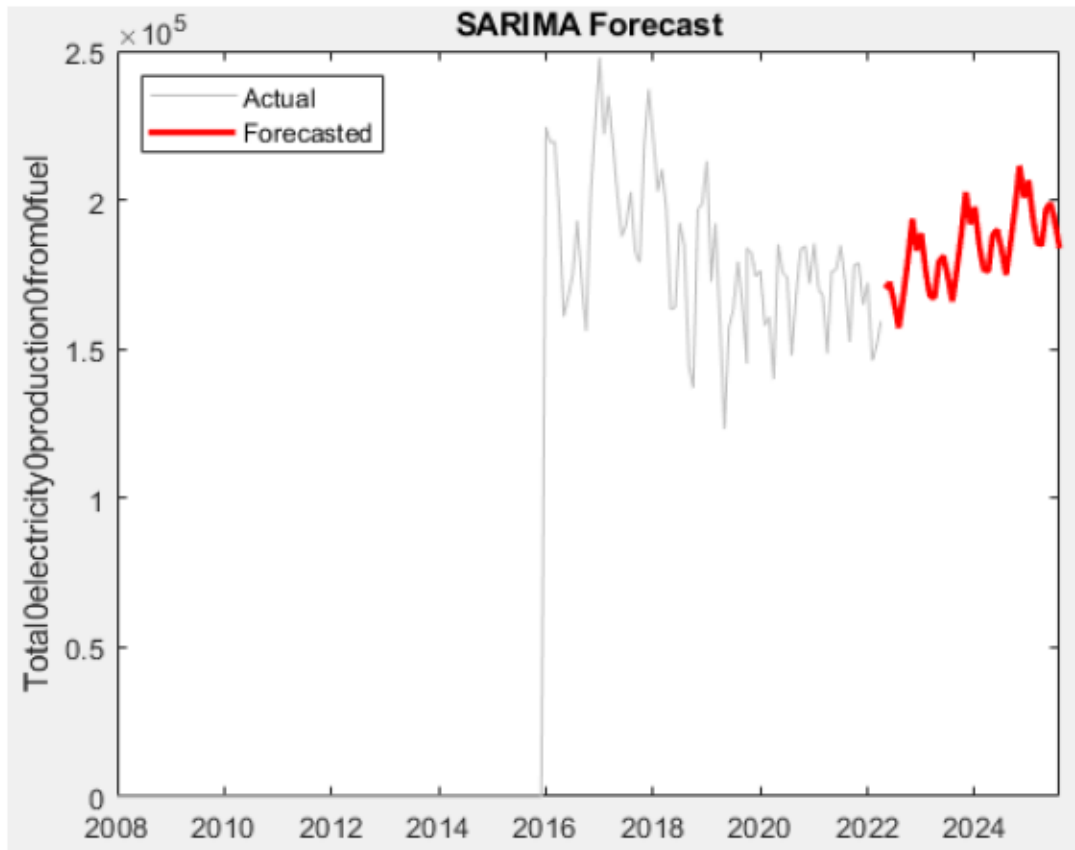


Figure 31 Préviation d'électricité produite à partir du combustible fossile par SARIMA – zone6

La figure 31 suggère que la production d'électricité à partir de combustibles fossiles a connu une interruption significative autour de 2016, suivie d'une légère reprise. Cela pourrait être dû à des politiques visant à réduire la dépendance aux combustibles fossiles ou à des investissements dans des énergies renouvelables. Cependant, la production à partir de combustibles fossiles reste importante, bien que la croissance semble modérée. SARIMA prévoit une stabilisation, ce qui pourrait indiquer que la dépendance à ces sources d'énergie traditionnelles restera présente mais contrôlée à moyen terme.

Conclusion :

Ce chapitre explore l'analyse des séries chronologiques de la production d'énergie électrique au Canada, en divisant le pays en zones géographiques par province. À l'aide de données historiques, le modèle SARIMA est appliqué pour prévoir la production future, en capturant efficacement les cycles, tendances et motifs saisonniers observés. Le modèle SARIMA a montré une capacité notable à représenter la saisonnalité et les tendances stables dans plusieurs types de production d'énergie, comme l'hydroélectricité et l'énergie solaire. Les simulations ont révélé une forte stabilité dans certaines zones, avec des prévisions cohérentes avec les cycles historiques. Cependant, bien que SARIMA réussisse à capturer les schémas réguliers et à anticiper les fluctuations saisonnières, il présente des limitations face aux changements soudains ou non linéaires dans les données. En particulier, SARIMA a du mal à s'adapter rapidement aux événements imprévus ou aux variations extrêmes, comme l'évolution rapide des politiques énergétiques ou les perturbations environnementales. Ainsi, bien que SARIMA soit efficace pour la prévision à court et moyen terme des séries chronologiques stables, d'autres modèles plus adaptés à la non-linéarité pourraient être nécessaires pour mieux gérer les changements imprévus ou dynamiques.

CHAPITRE 5 : SYNERGIE DES TECHNIQUES : LES MODÈLES DE BASE, HYBRIDE ET FUSION TRI-MODÈLE POUR L'ANALYSE PRÉDICTIVE

Introduction :

Ce chapitre se concentre sur le développement et l'évaluation de modèles de prévision de la consommation d'énergie mensuelle au Canada, en utilisant des données de janvier 2008 à décembre 2023. Une prévision précise de la consommation d'énergie est essentielle pour la planification des ressources, la gestion de l'offre et la formulation de politiques énergétiques efficaces.

Nous commencerons par explorer des modèles linéaires classiques comme l'ARIMA (Moyenne Mobile Intégrée Autorégressive) et le SARIMA (ARIMA saisonnier). Ensuite, nous intégrerons des techniques plus avancées, telles que les machines à vecteurs de support (SVM) et Arbres aléatoires (RF) et les réseaux de neurones artificiels (ANN), pour capturer les dynamiques complexes des données énergétiques.

L'étude progressera vers des combinaisons de modèles, cherchant à optimiser les prévisions en tirant parti des points forts de chaque méthode. Des modèles hybrides tels que SARIMA-SVM et autres seront évalués.

Ce voyage méthodique à travers divers modèles et combinaisons promet de révéler des connaissances précieuses et des solutions innovantes pour la prévision de la consommation d'énergie. En fin de compte, nous découvrirons comment empiler ces techniques pour atteindre une précision optimale, offrant ainsi des recommandations éclairées pour les décideurs et les gestionnaires de l'énergie.

Une exploration captivante et enrichissante des modèles de prévision énergétique, qui aboutira à des stratégies optimisées pour l'avenir. 'Les modèles seront intégrés progressivement selon les besoins, au fur et à mesure, tout au long du chapitre'.

1. Explication des nominalisations:

2.1.Modèles de base :

Dans le domaine de l'analyse de données et de la prévision, il est un élément clé de comprendre les modèles fondamentaux, souvent appelés **Modèles de base**. Ces modèles servent de référence et de point de départ pour de nombreuses analyses, offrant une base solide pour le développement et l'évaluation de modèles plus complexes. Cette partie explore le concept des modèles de base, leurs caractéristiques principales, et leur importance dans divers contextes d'application [1].

1.1.1. Caractéristiques des Modèles de Base :

Les Modèles de base se distinguent par leur simplicité et leur accessibilité. Ils sont généralement faciles à comprendre et à mettre en œuvre, ce qui les rend adaptés même pour les utilisateurs qui n'ont pas une formation approfondie en mathématiques ou en programmation. Cette simplicité n'altère en rien leur fiabilité; en effet, ces modèles ont démontré leur utilité à travers de nombreuses études et applications pratiques, fournissant des résultats cohérents et fiables [2-3].

Une autre caractéristique clé des Modèles de base est leur utilisation commune dans le domaine pour lequel ils ont été développés. Que ce soit pour la prévision de séries temporelles, la classification ou la régression, ces modèles sont largement reconnus et utilisés. Leur popularité découle de leur capacité à offrir des solutions satisfaisantes dans une variété de situations. De plus, ils servent souvent de base de comparaison pour d'autres modèles plus avancés. Les performances des nouveaux modèles sont

fréquemment évaluées par rapport à celles des modèles de base afin de mesurer leur efficacité relative [4].

2.2. Modèles Hybrides :

Les **Modèles hybrides** combinent les forces de plusieurs **Modèles de base**. Pour améliorer la précision et la robustesse des prévisions. Cette approche est particulièrement utile dans des domaines complexes où un seul modèle ne peut capturer toutes les caractéristiques des données. Cette partie du chapitre explore les concepts de base des Modèles hybrides, différentes approches de leur combinaison, notamment les **Approches en Cascade** et le **Méta-Apprentissage**, et leur importance dans l'analyse de données [5].

1.2.1. Caractéristiques des Modèles Hybrides :

Les Modèles hybrides se distinguent par leur capacité à combiner les avantages de plusieurs Modèles de base. Ils exploitent la complémentarité de différents algorithmes pour améliorer les performances globales. Par exemple, un modèle hybride peut utiliser un modèle linéaire pour capturer les tendances globales et un modèle non linéaire pour capturer les variations locales. Cette combinaison permet de mieux représenter les données et d'obtenir des prévisions plus précises [6].

Une autre caractéristique clé des modèles hybrides est leur flexibilité. Ils peuvent être adaptés à différents types de données et de problèmes, en ajustant les composants individuels pour répondre aux besoins spécifiques de l'application. De plus, les modèles hybrides peuvent réduire les erreurs et les biais associés à l'utilisation d'un seul modèle, en fournissant des prévisions plus robustes et fiables [7-4].

1.2.2. Exemples de Modèles Hybrides :

- **Approche en Cascade :**

L'approche en cascade consiste à utiliser les sorties d'un modèle comme entrées pour un autre modèle. Par exemple, un modèle SARIMA peut être utilisé pour capturer les tendances saisonnières des données, et ses résidus peuvent ensuite être utilisés comme entrées pour un modèle SVM ou RF. Cette approche permet de décomposer le problème en étapes successives, chaque modèle traitant une partie spécifique des données.

- **Méta-apprentissage :**

Le méta-apprentissage, ou apprentissage d'ensemble, combine les prédictions de plusieurs modèles pour obtenir une meilleure performance globale. Par exemple, un ensemble de modèles SARIMA, SVM, et RF peut être utilisé pour produire des prévisions individuelles, et une méthode de vote ou de pondération peut être appliquée pour combiner ces prévisions en une seule.

2.3.Fusion Tri-Modèle :

La **Fusion Tri-Modèle** combine les forces de trois modèles distincts pour améliorer la précision et la robustesse des prévisions. Cette approche est particulièrement utile dans des domaines complexes où même deux modèles combinés ne peuvent pas capturer toutes les caractéristiques des données. Ce chapitre explore les concepts de base de la fusion tri-modèle, différentes approches pour leur intégration, et leur importance dans l'analyse de données.

1.3.1. **Caractéristiques de la Fusion Tri-Modèle :**

Les modèles tri-modèle se distinguent par leur capacité à intégrer les avantages de trois modèles individuels. Ils exploitent la complémentarité de plusieurs algorithmes pour

améliorer les performances globales. Par exemple, un modèle peut utiliser SARIMA pour capturer les tendances saisonnières, SVM pour capturer les non-linéarités, et RF pour améliorer la généralisation des prévisions. Cette combinaison permet de mieux représenter les données et d'obtenir des prévisions plus précises [8].

Une autre caractéristique clé des modèles tri-modèle est leur complexité. Ils nécessitent une compréhension approfondie de chaque composant individuel et de leur interaction. De plus, ces modèles peuvent réduire les erreurs et les biais associés à l'utilisation de deux modèles, en fournissant des prévisions encore plus robustes et fiables [9].

1.3.2. Exemples de Fusion Tri-Modèle :

- Approche en Cascade :

L'approche en cascade dans une fusion tri-modèle consiste à utiliser les sorties séquentielles de chaque modèle. Par exemple, les résidus du modèle SARIMA peuvent être utilisés comme entrées pour un modèle SVM, et les résidus de ce modèle SVM peuvent ensuite être traités par un modèle RF. Cette approche permet de décomposer le problème en plusieurs étapes successives, chaque modèle traitant une partie spécifique des données.

- Méta-apprentissage :

Le méta-apprentissage dans une fusion tri-modèle combine les prédictions de trois modèles pour obtenir une meilleure performance globale. Par exemple, un ensemble de modèles SARIMA, SVM, et RF peut être utilisé pour produire des prévisions individuelles, et une méthode de vote, de pondération ou de régression peut être appliquée pour combiner ces prévisions en une seule.

2.4. Conclusion

La compréhension et la maîtrise des modèles de base, des modèles hybrides, et de la fusion tri-modèle sont essentielles pour exceller dans le domaine de l'analyse de données et de la prévision. Les modèles de base, tels que SARIMA, SVM et RF, offrent une simplicité, une accessibilité et une fiabilité qui en font des outils indispensables pour toute analyse de données, servant de référence et de point de comparaison. Les modèles hybrides, en combinant les forces de plusieurs modèles de base, permettent de capturer davantage de nuances dans les données, améliorant ainsi la précision et la robustesse des prévisions. La fusion tri-modèle va encore plus loin en intégrant trois modèles distincts, exploitant pleinement leurs complémentarités pour offrir des prévisions plus complètes et fiables, notamment dans des contextes complexes. Ces approches combinées permettent de surmonter les limitations des modèles individuels et d'offrir des solutions robustes et innovantes pour prendre des décisions informées et précises dans divers domaines d'application. En résumé, la maîtrise de ces différents types de modèles est fondamentale pour développer des solutions analytiques puissantes et efficaces.

2. Présentation des données :

Dans cette partie, nous décrivons les données de consommation d'énergie utilisées dans notre étude, couvrant la période de janvier 2008 à décembre 2023. La qualité et la préparation de ces données sont très importants pour la précision de nos prévisions.

Les données mensuelles de consommation d'énergie ont été obtenues auprès de sources fiables telles que Statistique Canada et d'autres bases de données énergétiques gouvernementales [13]. Ces sources fournissent des données détaillées et vérifiées sur la production, la réception, la livraison et la disponibilité de l'énergie au Canada.

2.1. Description des Données :

Les données utilisées dans cette étude (Tableau 3) comprennent les éléments suivants :

Tableau 3 Données real d'énergie électrique canadien - voir le fichier de programme pour l'ensemble des donnée-

Géographie	Période de Référence	Production Totale (MWh)	Énergie Totale Reçue (MWh)	Livraisons Totales (MWh)	Électricité Totale Disponible (MWh)
Canada	Jan-2008	59082501	1391169	4627072	55846598
	Feb-2008	55885424	1411545	4382286	52914683

	Dec-2023	63929085	1531422	6220295	56177368

Les variables incluses sont:

- **Production Totale (MWh)** : La quantité totale d'énergie produite en Mégawatt Heure.
- **Énergie Totale Reçue (MWh)** : L'énergie totale reçue d'autres sources.
- **Livraisons Totales (MWh)** : L'énergie totale livrée à d'autres consommateurs.
- **Électricité Totale Disponible (MWh)** : L'énergie totale disponible pour l'utilisation.

2.2.Préparation des Données :

Avant d'appliquer les modèles de prévision, il est essentiel de préparer et de nettoyer les données pour garantir leur qualité et leur pertinence. Les étapes de préparation des données incluent :

2.2.1. Gestion des Données Manquantes :

Identification et traitement des valeurs manquantes par interpolation ou remplacement par des moyennes.

2.2.2. Normalisation des Données :

Transformation des données pour qu'elles suivent une distribution normale, si nécessaire, pour améliorer la performance des modèles.

2.2.3. Analyse des Tendances et Saisonnalité :

Détection et extraction des tendances à long terme et des composants saisonniers dans les séries temporelles.

2.2.4. Division des Données :

Séparation des données en ensembles d'entraînement et de test pour valider les modèles de prévision. Typiquement, les données des dernières années (par exemple, 2021-2023) sont utilisées pour tester les modèles, tandis que les données antérieures servent à l'entraînement.

2.2.5. Visualisation des Données :

Pour mieux comprendre les caractéristiques des données, des visualisations graphiques seront utilisées. Des graphiques de séries temporelles, des histogrammes et des diagrammes de dispersion aideront à illustrer les tendances, la saisonnalité et les relations entre les variables.

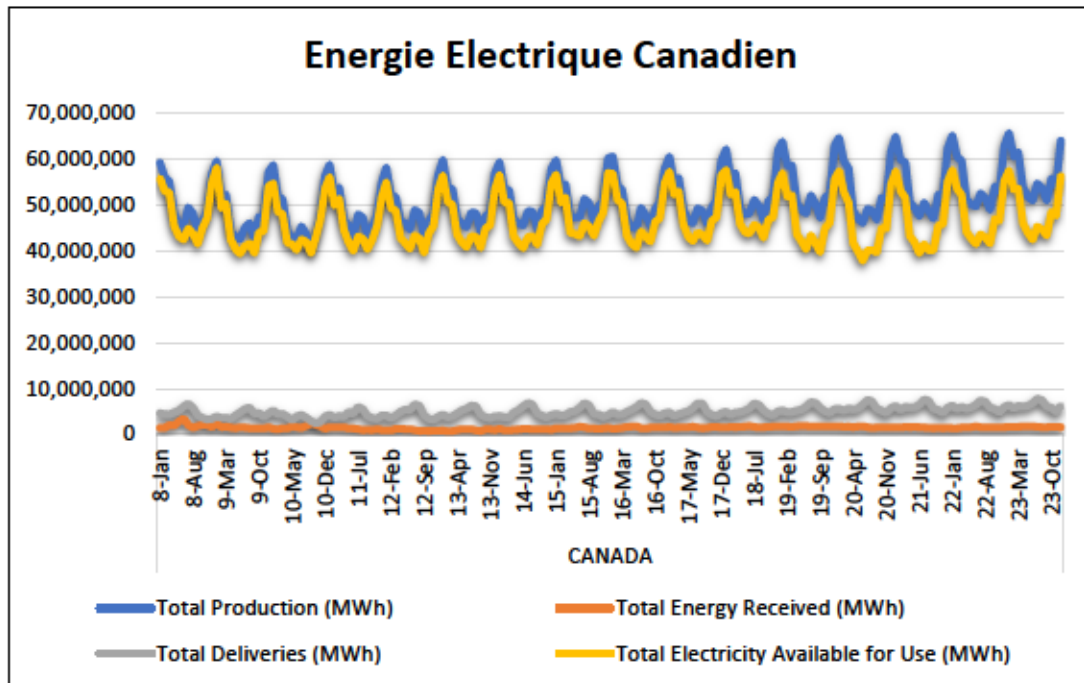


Figure 32 Données real d'énergie électrique canadien

Cette préparation minutieuse des données (Figure 32) constitue la base sur laquelle nos modèles de prévision seront construits et évalués. La prochaine section détaillera l'application et les résultats des différents modèles explorés dans cette étude.

2.2.6. Réduction des Données dans les études théoriques : Simplicité et Compréhension :

Nous avons réduit notre ensemble de données pour ne considérer que la production totale d'énergie au Canada de janvier 2022 à décembre 2023. Cette réduction a été effectuée afin de garantir la simplicité et de faciliter la compréhension du chapitre. Voici les détails et les raisons de cette approche :

2.3. Objectif de la Réduction des Données :

2.3.1. Simplicité de Présentation :

En se concentrant sur les données des deux dernières années, nous pouvons présenter les méthodes et les modèles de manière plus claire et concise. Cela permet de suivre plus facilement le déroulement de l'analyse sans être submergés par un trop grand volume de données. Une approche simplifiée aide à mieux illustrer les concepts clés et les processus méthodologiques.

2.3.2. Faciliter la Compréhension :

Les données récentes nous permettent de mettre en lumière les tendances actuelles et les changements récents dans la production d'énergie, offrant ainsi une perspective pertinente et actuelle. En se concentrant sur une période plus courte, nous pouvons analyser les impacts des événements récents et leur pertinence sur la production d'énergie.

2.3.3. Validation Efficace des Modèles :

Utiliser un ensemble de données plus restreint facilite les tests initiaux et la validation des modèles. Cela nous permet de vérifier l'efficacité des modèles de manière plus rapide et de les affiner avant de les appliquer à un ensemble de données plus large. Une validation préliminaire sur un échantillon réduit garantit une base solide pour des analyses ultérieures plus complexes.

2.3.4. Concentration sur les Événements Récents :

Les deux dernières années ont été marquées par des événements significatifs qui ont pu impacter la production d'énergie. En se concentrant sur cette période, nous pouvons mieux comprendre et modéliser ces impacts récents. Cela inclut les changements dans les politiques énergétiques, les innovations technologiques et les variations économiques.

Données Utilisées :

Voici les données réduites utilisées pour cette analyse (Tableau 4) :

Tableau 4 Données Réduites pour la compression du chapitre

Mois	Production Totale (MWh)
Jan-22	65051614
Fev-22	56206968
Mar-22	56654463
Avr-22	50183519
Mai-22	47824561
Jun-22	49631628
Jul-22	52455904
Auo-22	51435573
Sep-22	48846766
Oct-22	53980186
Nov-22	52748077
Dec-22	62611491
Jan-23	65621826
Fev-23	60434606
Mar-23	61433409
Avr-23	53689639
Mai-23	51863662
Jun-23	50867888
Jul-23	54656370
Auo-23	53559647
Sep-23	50876085
Oct-23	55777763
Nov-23	54117213
Dec-23	63929085

Avantages de cette Réduction :

- ✓ *Analyse Ciblée* : La réduction des données permet une analyse plus ciblée et compréhensible, facilitant ainsi la démonstration des modèles et leur application progressive.
- ✓ *Présentation Fluide* : Cela assure une fluidité dans la présentation du chapitre et une meilleure absorption des concepts.
- ✓ *Étapes Futuristes* : Une fois les modèles validés sur cet ensemble réduit, nous étendrons l'analyse à l'ensemble des données disponibles pour des prévisions encore plus robustes et précises.

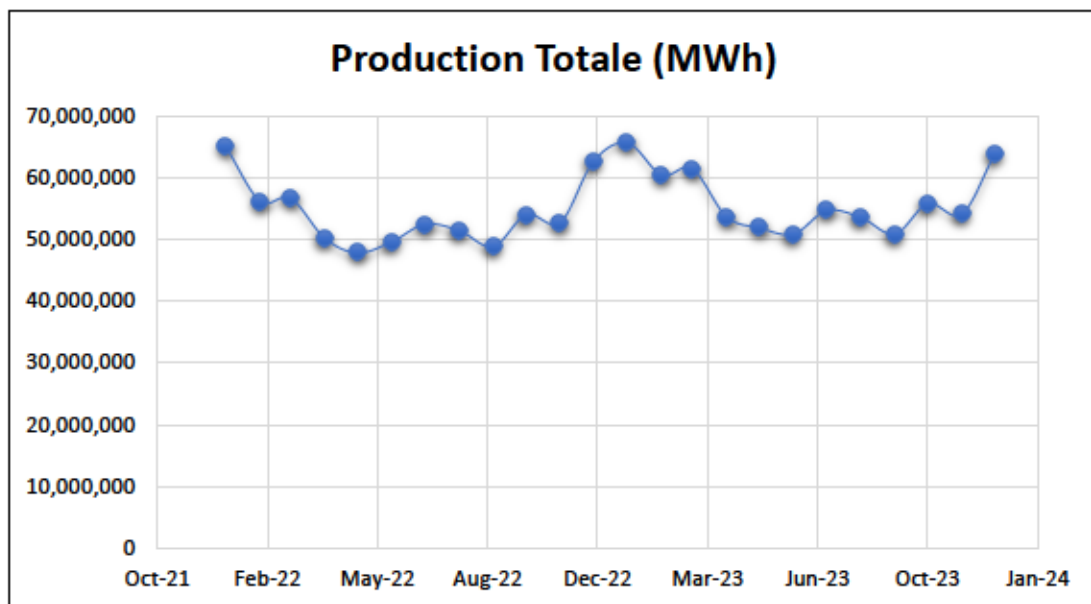


Figure 33 Production Totale (MWh) - CANADA

3. Modèles Linéaires :

2.1. Moyenne Mobile Intégrée Autorégressive -ARIMA- :

ARIMA, qui signifie Moyenne Mobile Intégrée Autorégressive, est une méthode statistique largement utilisée pour analyser et prévoir les données de séries temporelles.

Cette technique est particulièrement utile dans des domaines tels que la finance, l'économie et la consommation d'énergie, où les données sont collectées de manière séquentielle au fil du temps.

$$Y'_t = c + \sum_{i=1}^p \phi_i Y'_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

Composants Clés d'ARIMA (p, d, q):

Explication des Coefficients

- **Coefficients AR:**
 - ϕ_1 et ϕ_2 représentent les coefficients auto-régressifs. Ils montrent l'influence des valeurs précédentes (deux périodes en arrière) sur la valeur actuelle.
- **Coefficients MA:**
 - θ_1 et θ_2 représentent les coefficients de moyenne mobile. Ils montrent l'influence des erreurs passées (deux périodes en arrière) sur la valeur actuelle.

Ces valeurs sont dérivées des données historiques en utilisant la méthode d'estimation des moindres carrés ou une autre méthode d'estimation statistique, qui est implémentée dans la fonction *'fit'* de la bibliothèque *'statsmodels'* dans le logiciel de programmation Python.

Voici un tableau résumant les coefficients et les paramètres du modèle ARIMA (2,1,2) obtenus à partir des résultats du modèle ajusté :

Tableau 5 Les Coefficient de AR-I-MA

Composant	Coefficient	Valeur
AutoRégressif (AR)	ϕ_1	0.0955
	ϕ_2	0.3474
Moyenne Mobile (MA)	θ_1	-0.2526
	θ_2	-0.3426
Variance de l'erreur	σ^2	2.308e ¹³

- Composant Intégré (I) : Ce composant implique la différenciation des données pour les rendre stationnaires. Une série temporelle stationnaire a une moyenne et une variance constantes au fil du temps, ce qui est une hypothèse clé pour de nombreuses méthodes de prévision. La différenciation élimine les tendances et la saisonnalité, simplifiant ainsi le modèle. Le nombre d'opérations de différenciation requises est indiqué par le paramètre d .

Dans notre modèle, le paramètre d est 1, ce qui signifie que nous appliquons une différenciation d'ordre 1.

Formule de la différenciation d'ordre 1:

$$Y'_t = Y_t - Y_{t-1}$$

Tableau 6 Calcule de la Différenciation

Mois	Production Totale (MWh)	Différenciation (Y'_t)
Feb-22	56206968	56206968 - 65051614 = -8844646
Mar-22	56654463	56654463 - 56206968 = 447495
Apr-22	50183519	50183519 - 56654463 = -6470944
May-22	47824561	47824561 - 50183519 = -2358958

- **Composant AutoRégressif (AR) :** Le volet AR d'ARIMA examine la relation entre une observation et plusieurs observations décalées (c'est-à-dire des valeurs passées), il modélise essentiellement la dépendance de la valeur actuelle aux valeurs précédentes, la force de cette relation est capturée par le paramètre p .

Dans notre modèle ARIMA (2, 1, 2), le paramètre p est 2, ce qui signifie que nous utilisons les deux observations précédentes pour prédire la valeur actuelle.

Formule de l'AR (2):

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

Utilisons les coefficients fournis, nous pouvons calculer les valeurs prédites pour Mai 2022 :

$$Y_{Mai-22} = \phi_1 Y_{Avr-22} + \phi_2 Y_{Mar-22}$$

$$Y_{Mai-22} = 0.0955 \times (-6470944) + 0.3474 \times 447495$$

$$Y_{Mai-22} = -462768.149$$

- **Composant Moyenne Mobile (MA)** Le volet MA modélise la relation entre une observation et une erreur résiduelle d'un modèle de moyenne mobile appliqué aux observations décalées. Ce composant aide à lisser les fluctuations à court terme. Le paramètre q représente le nombre d'erreurs de prévision décalées dans l'équation de prédiction.

Le paramètre q est 2, ce qui signifie que nous utilisons les deux erreurs passées pour prédire la valeur actuelle.

Formule de MA (2):

$$\varepsilon_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t$$

μ : Constant

Calculons les résidus pour Mars 2022 ε_{Mar-22} et ε_{Avr-22} à partir des résultats :

$$\varepsilon_{Mar-22} = 9803572.35$$

$$\varepsilon_{Avr-22} = -6938536.82$$

Nous pouvons alors calculer les valeurs des résidus pour Mai 2022 :

$$\varepsilon_{Mai-22} = \theta_1 \varepsilon_{Avr-22} + \theta_2 \varepsilon_{Mar-22}$$

$$\varepsilon_{Mai-22} = \theta_1 \varepsilon_{Avr-22} + \theta_2 \varepsilon_{Mar-22}$$

$$\varepsilon_{Mai-22} = (-0.2526 \times -6938536.82) + (-0.3426 \times 9803572.35)$$

$$\varepsilon_{Mai-22} = -1601174.574$$

La combinaison des composant ARIMA donne la prévision future des données :

$$Y'_{Mai-22} = X_{Avr-22} + Y_{Mai-22} + \varepsilon_{Mai-22}$$

$$Y'_{Mai-22} = 50183519 - 4622768.149 - 1601174.574$$

$$Y'_{Mai-22} = 48159576.277MWh$$

En utilisant les coefficients fournis et les résidus calculés, nous avons calculé manuellement la prévision pour Mai 2022 en utilisant un modèle ARIMA (2,1,2). Les calculs montrent comment les composants AR et MA sont combinés pour produire la prévision finale.

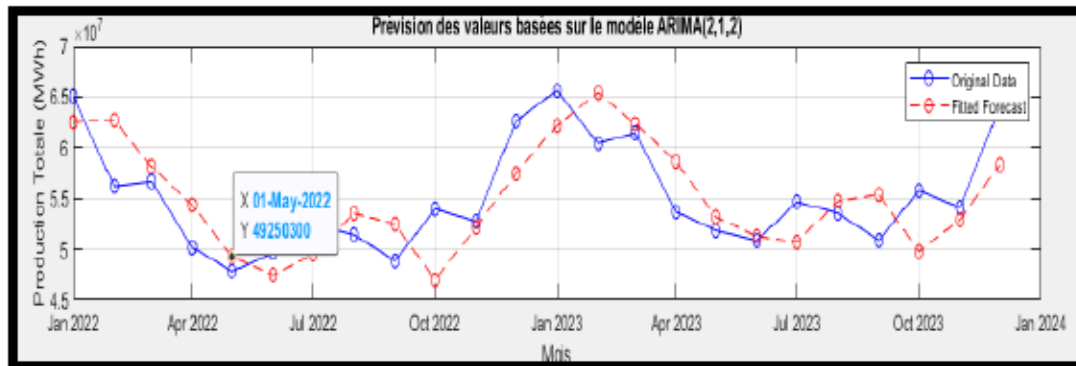


Figure 34 Les prévisions de l'Énergie électrique par le Model ARIMA

Les prévisions de la production totale d'énergie électrique utilisant le modèle ARIMA (2,1,2) sur MATLAB couvrent la période de janvier 2022 à décembre 2023. La figure 34 montre les données originales (ligne bleue) et les valeurs prévues (ligne rouge avec des cercles). En se concentrant sur les prévisions spécifiques, pour le mois de mai 2022, le modèle ARIMA prévoit une production de 49,250,300 MWh. En comparaison, les calculs manuels pour le même mois donnent une valeur de 48,159,576.277 MWh, révélant une différence de 1,090,723.723 MWh, soit environ 2.21%. Cette variation montre que les prévisions du modèle ARIMA sont raisonnablement précises, bien qu'il y ait de légères divergences probablement dues aux différences méthodologiques ou aux approximations inhérentes aux algorithmes.

Durant l'ensemble de la période, on observe que les prévisions suivent de près les tendances des données originales, avec des fluctuations saisonnières et des variations mensuelles capturées efficacement par le modèle. Les périodes de hausse et de baisse de la production sont bien reproduites, ce qui confirme la robustesse du modèle ARIMA (2,1,2) pour capturer les dynamiques sous-jacentes de la production d'énergie. Cette précision est très utile pour la planification et la gestion de l'énergie, permettant aux décideurs de disposer d'estimations fiables pour anticiper la demande et ajuster la production en conséquence. Malgré les petites différences observées, telles que celle

de mai 2022, le modèle montre une forte capacité à fournir des prévisions cohérentes et utiles pour l'ensemble de la période considérée.

Fonctionnement d'ARIMA :

- Identification : La première étape de la modélisation ARIMA consiste à identifier les valeurs appropriées pour p , d , et q . Cela se fait souvent à l'aide de graphiques des données et de tests statistiques pour déterminer la présence de tendances et de saisonnalité.
- Estimation Une fois les paramètres identifiés, le modèle ARIMA est estimé à l'aide des données historiques. Cela implique de faire correspondre le modèle aux données pour déterminer les coefficients des composants AR et MA.
- Vérification Diagnostique Après avoir ajusté le modèle, il est essentiel de vérifier les résidus (les différences entre les valeurs observées et les prévisions du modèle) pour s'assurer qu'ils ressemblent à du bruit blanc (erreurs aléatoires). Cette étape aide à valider l'exactitude du modèle.
- Prévion Avec un modèle ARIMA validé, des prévisions pour les périodes futures peuvent être générées. Le modèle utilise les motifs identifiés dans les données historiques pour faire des prédictions sur les valeurs futures.

Avantages et Limitations :

- Avantages : ARIMA est polyvalent et peut modéliser une large gamme de données de séries temporelles. Il est également bien adapté pour des prévisions à court et moyen terme.
- Limitations : Bien que le modèle ARIMA soit puissant, il présente certaines limitations lorsqu'il s'agit de séries temporelles avec des composantes saisonnières marquées. Les modèles ARIMA gèrent bien les tendances et les relations linéaires à court terme, mais ils peuvent ne pas capturer efficacement les motifs saisonniers qui se répètent à intervalles réguliers.

Par exemple, des séries temporelles comme les données de consommation d'énergie mensuelles peuvent présenter des fluctuations saisonnières dues aux changements climatiques et aux variations d'utilisation au cours de l'année.

2.2.Introduction de ARIMA Saisonnière -SARIMA- :

Pour pallier ces limitations, nous introduisons le modèle SARIMA, qui signifie ARIMA saisonnière.

SARIMA est une extension du modèle ARIMA qui intègre des termes supplémentaires pour capturer la saisonnalité. Un modèle SARIMA est noté ARIMA (p, d, q) (P, D, Q) [s], où :

- p, d, q sont les ordres du modèle ARIMA non saisonnier.
- P, D, Q sont les ordres du modèle saisonnier.
- s est la période de la saisonnalité (par exemple, 12 pour des données mensuelles avec une saisonnalité annuelle).

Les composantes saisonnières (P, D, Q) permettent de modéliser les motifs répétitifs qui se produisent à chaque période saisonnière.

Le modèle SARIMA (2,1,2) (1,1,1) [12] peut être écrit comme :

$$(1 - \phi_1 L - \phi_2 L^2)(1 - \phi_{12} L^{12})(1 - L)(1 - L^{12})Y_t = c + (1 + \theta_1 L + \theta_2 L^2)(1 + \theta_{12} L^{12})\varepsilon_t$$

Avec :

- L : est l'opérateur de décalage (ou backshift), défini par $LY_t = Y_{t-1}$
- $\phi_{12} = 0.7$: Coefficient du terme auto-régressif saisonnier (SAR).
- $\theta_{12} = 0.4$: Coefficient du terme de moyenne mobile saisonnier (SMA).

Calcul des Prévisions pour Mai 2022 :

Pour calculer la prévision pour mai 2022 (Y'_{Mai-22}), nous avons besoin des données antérieures :

$$Y_{Avr-22} = 50183519 ; Y_{Mars-22} = 56654463 ; Y_{Mai-21} = 47824561$$

Pour simplifier, nous allons négliger les erreurs antérieures ε :

$$Y'_{Mai-22} = \phi_1 Y_{Avr-22} + \phi_2 Y_{Mars-22} + \phi_{12} Y_{Mai-21} - \theta_1 \varepsilon_{Avr-22} - \theta_2 \varepsilon_{Mar-22} + \theta_{12} Y_{Mai-21}$$

Substituons les valeurs et les coefficients :

$$Y'_{Mai-22} = 0.0955 \times Y_{Avr-22} + 0.3474 \times Y_{Mars-22} + 0.7 \times Y_{Mai-21} - (-0.2526) \times \varepsilon_{Avr-22} - (-0.3426) \times \varepsilon_{Mar-22} + 0.4 \times Y_{Mai-21}$$

Additionnons ces valeurs :

$$Y'_{Mai-22} = 47995400 \text{ MWh}$$

Ce calcul manuel simplifié montre comment appliquer un modèle SARIMA pour prédire des valeurs futures en utilisant des données historiques. En pratique, des outils logiciels comme MATLAB, R ou Python sont utilisés pour gérer les données de manière plus efficace et précise, et pour inclure les erreurs passées dans les calculs.

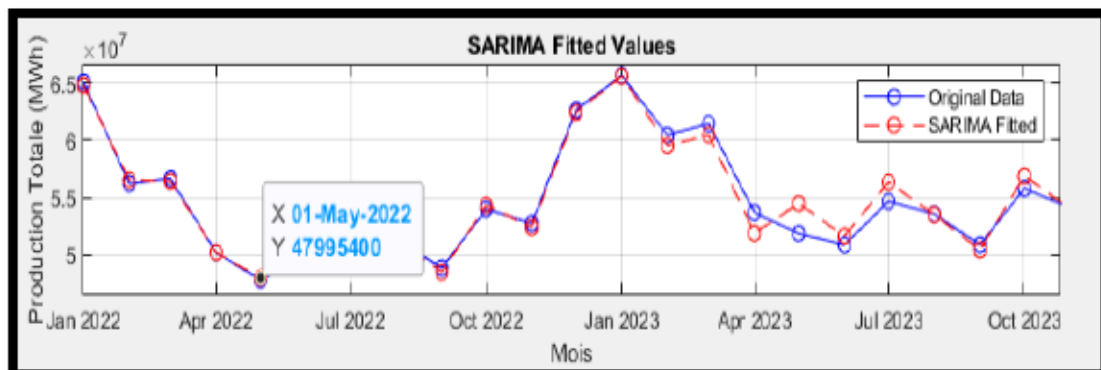


Figure 35 Les prévisions de l'Énergie électrique par le Model SARIMA

Le modèle SARIMA a également été utilisé pour prévoir la production totale d'énergie électrique sur la même période, de janvier 2022 à décembre 2023. La figure 35 montre les données originales en bleu et les valeurs ajustées par le modèle SARIMA en rouge. En mai 2022, le modèle SARIMA prévoit une production de 47,995,400 MWh.

Comparaison avec les Résultats de ARIMA et Calculs Manuels

➤ Pour le mois de mai 2022:

- Valeur Prédite par ARIMA: 49,250,300 MWh
- Valeur Prédite par SARIMA: 47,995,400 MWh
- Valeur Calculée Manuellement: 48,159,576.277 MWh

➤ Analyse de la Différence

- Différence entre ARIMA et SARIMA:

$$49,250,300 - 47,995,400 = 1,254,900 \text{ MWh}$$

- Différence entre SARIMA et Calculs Manuels:

$$48,159,576.277 - 47,995,400 = 164,176.277 \text{ MWh}$$

➤ Interprétation des Résultats sur la Période

Le modèle SARIMA semble également bien capturer les tendances saisonnières et les fluctuations mensuelles de la production d'énergie électrique, tout comme le modèle ARIMA. En comparant les deux modèles, on remarque que les prévisions de SARIMA pour mai 2022 sont légèrement plus proches des résultats calculés manuellement, avec une différence de seulement 164,176.277 MWh, soit environ 0.34%, comparé à la différence de 2.21% observée avec le modèle ARIMA.

Les résultats montrent que le modèle SARIMA est légèrement plus précis par rapport aux calculs manuels que le modèle ARIMA. Sur l'ensemble de la période de janvier 2022 à décembre 2023, le modèle SARIMA fournit des prévisions cohérentes qui suivent de près les données originales, avec une capacité robuste à capturer les dynamiques saisonnières et les variations mensuelles. Cette précision améliorée, bien que subtile, peut offrir des avantages supplémentaires pour la planification énergétique,

en réduisant encore les marges d'erreur et en fournissant des estimations plus fiables pour la gestion de la production d'énergie électrique.

Maintenant, nous allons entraîner le modèle SARIMA sur l'ensemble des données disponibles, couvrant la période de 2008 à 2023. En utilisant les valeurs de production totale mensuelle d'énergie sur cette période, nous ajusterons les coefficients du modèle pour optimiser sa précision prédictive. Cette étape essentielle consiste à estimer les paramètres du modèle autorégressif, de moyenne mobile, ainsi que les composants saisonniers. L'objectif est d'obtenir un modèle robuste capable de capturer les tendances, les cycles et les fluctuations saisonnières des données de consommation d'énergie. Une fois le modèle entraîné (Figure 36), nous pourrions générer des prévisions précises pour les mois futurs, en commençant par une validation sur les données historiques pour évaluer la performance du modèle.

Tableau 7 Coefficients du modèle SARIMA

Paramètre	Valeur	Erreur	Statistique t	Valeur-P
Constante	$-1.2737e^4$	$1.5563e^5$	-0.0818	0.9348
AR{1}	0.0632	0.4497	0.1405	0.8883
AR{2}	-0.4470	0.2328	-19197	0.0549
SAR{1}	-0.5122	0.4959	-1.0329	0.3017
MA{1}	-0.4583	0.8965	-0.5112	0.6092
MA{2}	0.2235	0.4135	0.5405	0.5889
SMA{1}	0.3806	0.9293	0.4096	0.6821
Variance	$3.213e^{12}$	0.0109	$2.9485e^{14}$	0

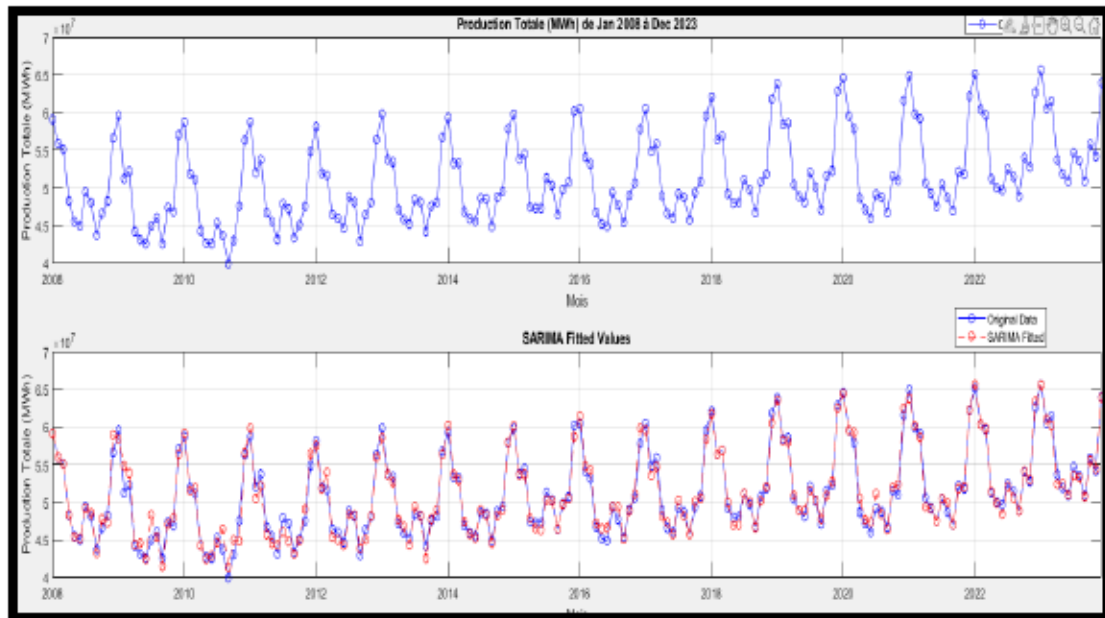


Figure 36 Model SARIM entrainer sur l'ensemble de données de 2008 à 2023

Conclusion sur les Modèles ARIMA et SARIMA

Les modèles linéaires ARIMA et SARIMA sont des outils puissants pour la prévision de séries temporelles, particulièrement utiles dans des contextes où les données présentent des comportements linéaires et, dans le cas de SARIMA, des variations saisonnières.

➤ Simplicité et Efficacité:

- ARIMA : Idéal pour des séries temporelles sans composantes saisonnières explicites. Il combine l'autorégression (AR), l'intégration (I) pour rendre la série stationnaire, et la moyenne mobile (MA) pour modéliser les résidus.
- SARIMA : Une extension d'ARIMA qui inclut des termes saisonniers, permettant de capturer des patterns répétitifs à intervalles réguliers. SARIMA est plus complexe mais offre une meilleure précision pour les séries temporelles présentant des fluctuations saisonnières.

- Flexibilité:
 - Les deux modèles sont très flexibles et peuvent être adaptés à une variété de séries temporelles grâce à la possibilité de combiner différentes composantes AR, I, MA, et, pour SARIMA, les termes saisonniers.
- Applications Pratiques:
 - ARIMA : Souvent utilisé pour la prévision des ventes, des cours boursiers, ou des indicateurs économiques, là où les patterns saisonniers ne sont pas prédominants.
 - SARIMA : Utilisé dans des contextes tels que la prévision de la demande énergétique, les données climatiques, ou toute série temporelle présentant des variations saisonnières claires.
- Limitations:
 - Les modèles ARIMA et SARIMA supposent une relation linéaire dans les données, ce qui peut ne pas capturer des comportements non linéaires ou des interactions complexes.
 - La sélection des paramètres (p, d, q) pour ARIMA et (P, D, Q, s) pour SARIMA peut être complexe et nécessite une analyse minutieuse des données.
- Complémentarité avec d'Autres Modèles:
 - Pour des séries temporelles avec des composantes non linéaires ou lorsque les modèles ARIMA/SARIMA ne sont pas suffisants, ils peuvent être combinés avec d'autres méthodes comme les réseaux de neurones artificiels (ANN), les machines à vecteurs de support (SVM), ou les forêts aléatoires (RF) pour améliorer la précision des prévisions.

En résumé, les modèles ARIMA et SARIMA restent des outils essentiels pour la prévision de séries temporelles, offrant une simplicité et une robustesse pour les

données linéaires et saisonnières. Cependant, ils doivent parfois être complétés par des modèles non linéaires pour capturer toutes les nuances des données réelles.

4. Modèles non-Linéaires :

2.1.Machine à Vecteurs de Support -SVM- :

La prévision de la production d'électricité est une tâche nécessaire dans le secteur de l'énergie. Des prédictions précises permettent une gestion efficace des ressources, une planification optimale et l'assurance de l'équilibre entre l'offre et la demande.

Parmi les nombreux modèles d'apprentissage automatique disponibles, la Machine à Vecteurs de Support (SVM) se distingue par sa robustesse et sa polyvalence. Cette partie propose une exploration approfondie de l'application des SVM pour prédire la production d'électricité en se basant sur les données météorologiques. Nous commencerons par détailler le jeu de données utilisé, suivrons avec une explication détaillée de la méthodologie SVM et de ses équations, effectuerons des calculs manuels, implémenterons le modèle en utilisant MATLAB et interpréterons les résultats de la courbe de prédiction SVM.

Données Utilisées dans la SVM :

Le jeu de données comprend les températures moyennes mensuelles et la production totale d'électricité pour les années 2022 à 2023.

Voici un résumé des données :

**Tableau 8 Données météorologiques et énergétiques de la province du Québec
2022-2023**

Mois	Température Moyenne (°C)	Production Totale (GWh)
Janvier 2022	-10.5	17 200
Février 2022	-8.7	16 745
Mars 2022	-3.2	16 560
Avril 2022	4	15 505
Mai 2022	12.1	14 030
Juin 2022	18.7	13 505
Juillet 2022	21.3	13 030
Août 2022	20.4	13 115
Septembre 2022	14.7	14 290
Octobre 2022	7.8	15 755
Novembre 2022	0.4	16 400
Décembre 2022	-7.5	16 960
Janvier 2023	-9.8	17 330
Février 2023	-8.4	16 875
Mars 2023	-2.8	16 690
Avril 2023	4.5	15 635
Mai 2023	12.4	14 145
Juin 2023	19	13 620
Juillet 2023	21.5	13 145
Août 2023	20.6	13 230
Septembre 2023	15	14 405
Octobre 2023	7.9	15 870
Novembre 2023	0.5	16 525
Décembre 2023	-7.2	17 090

4.1.1. SVM :

La machine à vecteurs de support (SVM) est un puissant algorithme d'apprentissage supervisé qui peut être utilisé pour des tâches de classification et de régression. La SVM est efficace dans les espaces de haute dimension et est polyvalente grâce aux différentes fonctions de noyau qu'elle prend en charge.

4.1.2. Histoire et Évolution de la SVM :

Le concept des machines à vecteurs de support a été introduit par « **Vladimir Vapnik** » et « **Alexey Chervonenkis** » dans les années 1960 et 1970.



Figure 37 Vladimir Vapnik 1936

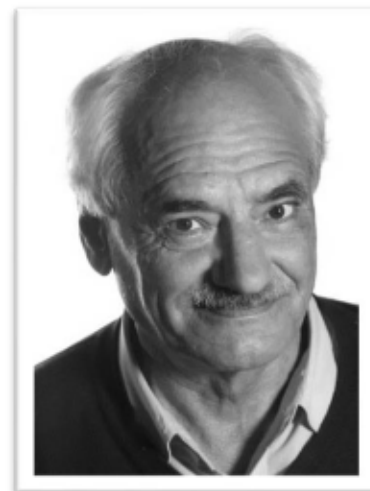


Figure 38 Alexey Chervonenkis
1938-2014

Cependant, ce n'est qu'à partir des années 1990 que les SVM ont gagné en reconnaissance dans la communauté de l'apprentissage automatique, grâce aux travaux de **Vapnik** et de ses collègues chez AT&T Bell Labs. L'introduction du "*truc du noyau*" a permis aux SVM de gérer les tâches de classification et de régression non linéaires, en faisant un choix populaire pour diverses applications, y compris la classification de texte, la reconnaissance d'images et la bio-informatique.

4.1.3. Fondements Mathématiques de la SVM :

Au cœur de la SVM se trouve la recherche de l'hyperplan optimal qui sépare les données en différentes classes (pour la classification) ou ajuste les données avec une erreur minimale (pour la régression). L'objectif est de maximiser la marge, qui est la distance entre l'hyperplan et les points de données les plus proches de chaque classe. Cela se fait en résolvant un problème d'optimisation sous contraintes [22].

Pour la régression, la SVM est souvent appelée régression à vecteurs de support (SVR). La SVR vise à trouver une fonction qui dévie des valeurs observées réelles d'une valeur ne dépassant pas ϵ (fonction de perte insensible à l'épsilon) et en même temps qui est aussi plate que possible. L'idée de base est de mapper les données d'entrée dans un espace de caractéristiques de haute dimension en utilisant une fonction de noyau, puis de réaliser une régression linéaire dans cet espace de haute dimension [23].

Le problème de régression SVM peut être défini comme suit :

$$\min_{w, b, \xi_i, \xi_i^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

Sous réserve de :

$$\begin{cases} y_i - (w \cdot x_i + b) \leq \epsilon + \xi_i \\ (w \cdot x_i + b) - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Avec :

- w : est le vecteur de poids,
- b : est le terme de biais,
- ξ_i et ξ_i^* : sont des variables de relâchement,

- C : est le paramètre de régularisation,
- ϵ : définit la marge de tolérance.

4.1.4. Fonctions de Noyau :

La flexibilité de la SVM provient de l'utilisation des fonctions de noyau, qui lui permettent de gérer les relations non linéaires en mappant implicitement les données d'entrée dans un espace de caractéristiques de plus haute dimension. Voici quelques fonctions de noyau courantes :

- **Noyau Linéaire** : $K(x_i, x_j) = x_i \cdot x_j$
- **Noyau Polynômial** : $K(x_i, x_j) = (x_i \cdot x_j + 1)^d$
- **Noyau à Base Radiale (RBF)** : $K(x_i, x_j) = e^{(-\gamma \|x_i - x_j\|^2)}$
- **Noyau Sigmoidé** : $K(x_i, x_j) = \tanh(\alpha x_i \cdot x_j + c)$

Ces noyaux permettent à la SVM de modéliser des relations complexes entre les caractéristiques d'entrée et la variable cible.

4.1.5. Entraînement du Modèle SVM :

Nous utiliserons une version simplifiée de la régression SVM pour cet exemple. En pratique, on utilise une bibliothèque comme 'scikit-learn' ou MATLAB pour effectuer ces calculs, mais dans cette partie nous allons décrire les étapes manuellement.

On commence par le calcul du vecteur de poids w et le biais b pour le noyau linéaire, on utilise les équations suivantes :

$$w = (X^T X)^{-1} X^T y$$

$$b = \bar{y} - w \cdot \bar{x}$$

4.1.6. Calcul SVM :

1. Calculer $X^T X$:

$$X_{\text{Température Moyenne (}^\circ\text{C)}} = \begin{bmatrix} -10.5 \\ -8.7 \\ -3.2 \\ 4.0 \\ 12.1 \\ 18.7 \\ 21.3 \\ 20.4 \\ 14.7 \\ 7.8 \\ 0.4 \\ -7.5 \\ -9.8 \\ -8.4 \\ -2.8 \\ 4.5 \\ 12.4 \\ 19.0 \\ 21.5 \\ 20.6 \\ 15.0 \\ 7.9 \\ 0.5 \\ -7.2 \end{bmatrix}$$

$$X^T X = \sum_{i=1}^{24} x_i^2 = (-10.5)^2 + (-8.7)^2 + \dots + (-7.2)^2 = 3252.95$$

4.1.7. Calculer $X^T y$:

$$y_{\text{Production Totale (GWh)}} = \begin{bmatrix} 17200 \\ 16745 \\ 16560 \\ 15505 \\ 14030 \\ 13505 \\ 13030 \\ 13115 \\ 14290 \\ 15755 \\ 16400 \\ 16960 \\ 17330 \\ 16875 \\ 16690 \\ 15635 \\ 14145 \\ 13620 \\ 13145 \\ 13230 \\ 14405 \\ 15870 \\ 16525 \\ 17090 \end{bmatrix}$$

$$\begin{aligned} X^T y &= \sum_{i=1}^{24} x_i y_i = (-10.5 \times 17200) + (-8.7 \times 16745) + \dots + (-7.2 \times 17090) \\ &= -319581.9 \end{aligned}$$

2. Calculer w :

$$w = \frac{X^T y}{X^T X} = \frac{-319581.9}{3252.95} \approx -98.24$$

3. Calculer b :

$$\bar{x} = \frac{\sum_{i=1}^{24} x_i}{24} \approx 6.55$$

$$\bar{y} = \frac{\sum_{i=1}^{24} y_i}{24} \approx 15367.92$$

$$b = \bar{y} - w \cdot \bar{x} = 15367.92 - (-98.24 \times 6.55) \approx 16013.08$$

Ainsi, notre modèle de régression SVM linéaire est :

$$y = -98.24x + 16013.08$$

4.1.8. Prévoir et Évaluer :

En utilisant le modèle, nous pouvons prédire la production d'électricité pour une température moyenne donnée. Par exemple, pour une température moyenne de 10°C :

$$y = -98.24x + 16013.08 = -98.24 \times 10 + 16013.08 \approx 15030.68 \text{ GW}$$

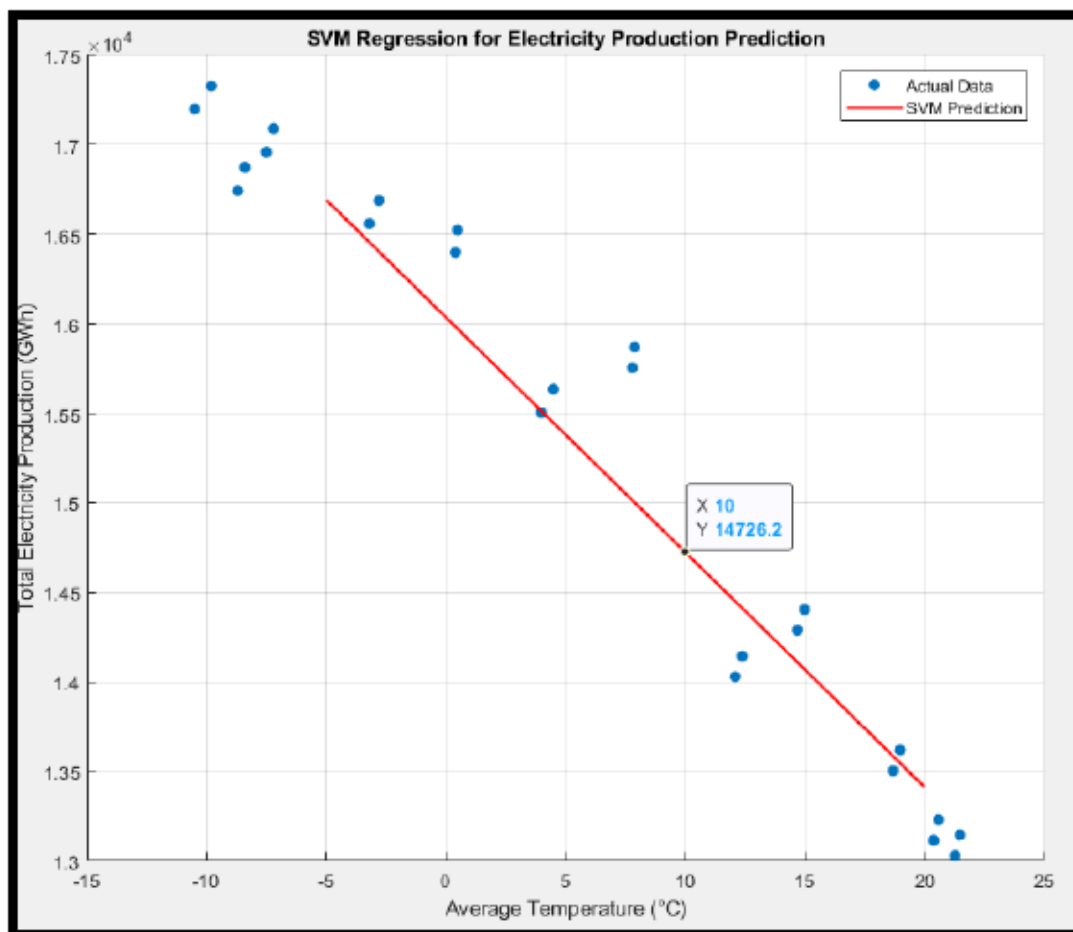


Figure 39 Régression SVM 'Noyau Linéaire' pour la Prédiction de la Production d'Électricité Québec de 2022-2023

La figure 39 montre les résultats de la régression par machine à vecteurs de support (SVM) pour la prédiction de la production totale d'électricité en fonction de la température moyenne. Les points bleus représentent les données réelles, tandis que la ligne rouge représente les prédictions faites par le modèle SVM.

4.1.9. Observations Clés :

- Données Réelles (Points Bleus):

- Les données montrent une relation inverse entre la température moyenne et la production totale d'électricité.
- Lorsque la température moyenne augmente, la production totale d'électricité tend à diminuer.
- Cette tendance est particulièrement évidente avec des températures plus élevées, où les valeurs de production sont significativement plus basses.

➤ Prédications du Modèle SVM (Ligne Rouge)

- La ligne de régression SVM suit une pente descendante, indiquant également une relation inverse entre la température et la production d'électricité.
- Le modèle SVM semble bien capturer la tendance générale des données, avec une ligne de régression qui passe proche de nombreux points de données réels.
- La valeur prédite par le SVM à une température moyenne de 10°C est d'environ 14,726.2 GWh.

4.1.10. Interprétation des Résultats :

➤ Relation Inverse:

- Le modèle SVM confirme une relation inverse forte entre la température moyenne et la production totale d'électricité. Cela peut être dû à plusieurs facteurs, tels que la diminution de la demande de chauffage en été ou des contraintes sur les systèmes de production d'énergie en raison de températures élevées.

➤ Précision du Modèle:

- Le modèle SVM semble être assez précis dans ses prédictions, comme le montre la proximité de la ligne rouge avec les points de données réels. Cependant, il y a quelques écarts notables, indiquant que le modèle pourrait bénéficier d'un ajustement supplémentaire ou de l'inclusion de variables supplémentaires pour améliorer sa précision.

➤ Utilisation Pratique:

- Ces prédictions peuvent être extrêmement utiles pour la gestion et la planification de la production d'électricité. En comprenant comment les variations de température affectent la production, les gestionnaires peuvent mieux anticiper les besoins en énergie et optimiser les ressources disponibles.

➤ Conclusion

La régression par machine à vecteurs de support (SVM) offre une représentation efficace de la relation entre la température moyenne et la production totale d'électricité. Les résultats montrent une forte relation inverse, avec une diminution de la production d'électricité à mesure que la température augmente. Bien que le modèle soit globalement précis, il pourrait encore être amélioré pour capturer certaines variations non expliquées dans les données réelles. Ces insights sont essentiels pour la planification stratégique et l'optimisation des ressources dans le secteur énergétique.

2.2.Modèle SVM Amélioré :

Pour améliorer la précision du modèle SVM en utilisant un noyau polynomial, Les noyaux polynomiaux permettent de capturer des relations non linéaires dans les

données, ce qui peut améliorer la performance du modèle pour des données qui ne sont pas linéairement séparables.

4.2.1. Équation du Noyau Polynomial:

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d$$

Ou bien plus généralement

$$K(x_i, x_j) = (\gamma \langle x_i, x_j \rangle + r)^d$$

Avec :

- γ : est un coefficient d'échelle
- $\langle x_i, x_j \rangle$: est le produit scalaire entre x_i et x_j
- r et d : sont les coefficients de décalage et le degré du polynôme successivement.

La recherche en grille (Grid Search) a été ajustée pour inclure différents ordres de polynômes (degrés). La recherche en grille permet de tester plusieurs combinaisons d'hyperparamètres pour trouver ceux qui offrent la meilleure performance.

4.2.2. Définition des Hyperparamètres :

$$\{C, Degree, \epsilon\}$$

Où ;

- C : est le paramètre de régularisation.
- $Degree$: est l'ordre du polynôme.
- ϵ : est un paramètre insensible à l'erreur dans la fonction de perte.

4.2.3. Évaluation par Validation Croisée:

La validation croisée 'cross-validation' est utilisée pour évaluer la performance des différentes configurations de noyaux polynomiaux. Elle permet de s'assurer que les résultats sont robustes et qu'ils ne sont pas dus à un surajustement des données d'entraînement.

L'équation par Validation Croisée est donnée par :

$$CV_{error} = \frac{1}{k} \sum_{i=1}^k Error_i$$

Avec :

- CV : Validation-Croisée (Cross-Validation)
- k : est le nombre de sous-échantillons
- $Error_i$: est l'erreur sur le sous-échantillon i

Des graphiques sont inclus pour visualiser les prédictions et les erreurs. Cela permet de mieux comprendre comment le modèle se comporte et d'identifier les éventuelles zones de faiblesse.

4.2.4. Équation de Prédiction SVM:

$$\hat{y} = \sum_{i=1}^n \alpha_i K(x_i, x) + b$$

Où :

- \hat{y} : est la valeur prédite.
- α_i : sont les coefficients du modèle.
- $K(x_i, x)$: est la fonction de noyau.
- b : est le biais.

4.2.5. Équation d'Erreur:

$$Erreur = y - \hat{y}$$

Où :

- y : est la valeur réelle.
- \hat{y} : est la valeur prédite.

Ces modifications permettent d'améliorer la précision du modèle SVM en utilisant un noyau polynomial. La recherche en grille et la validation croisée permettent de sélectionner les meilleurs hyperparamètres, tandis que la visualisation des prédictions et des erreurs aide à mieux comprendre les performances du modèle. En mettant en œuvre ces techniques, on obtient des prédictions plus précises et une meilleure compréhension de la relation entre les variables d'entrée et la production d'électricité.

4.2.6. Interprétation des Résultats du SVM Amélioré par Noyau Polynomial :

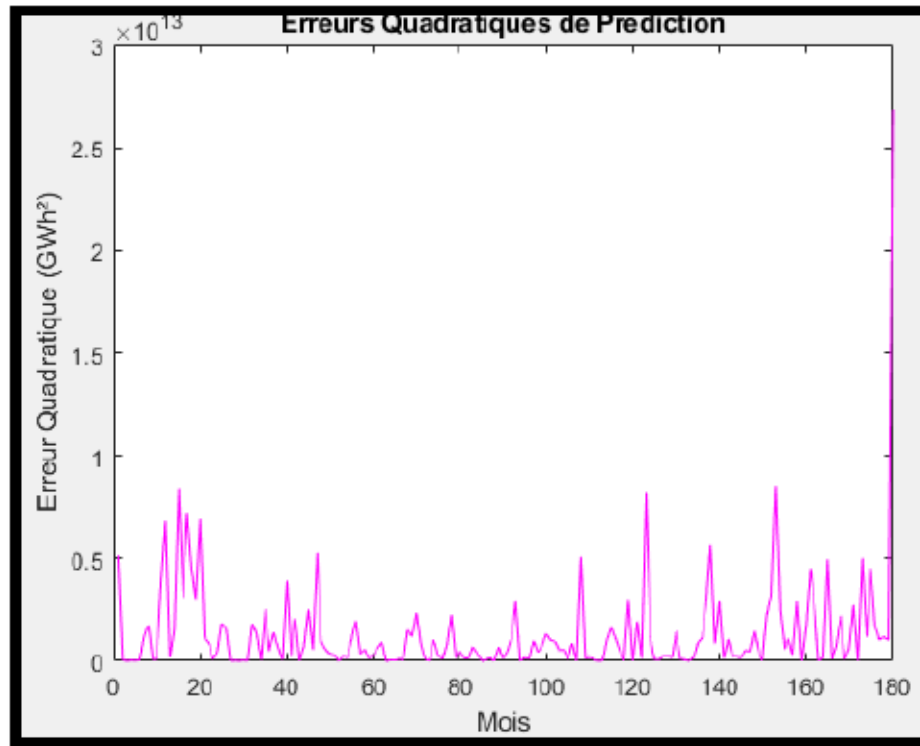


Figure 40 Les Erreurs Quadratiques

Figure 40 : Ce graphique montre les erreurs quadratiques des prédictions du SVM sur la période de janvier 2008 à décembre 2022, Les erreurs quadratiques restent relativement faibles pendant la majorité de la période, avec quelques pics notables.

Les erreurs quadratiques élevées peuvent indiquer des périodes où les prévisions du modèle SVM diffèrent considérablement des valeurs réelles.

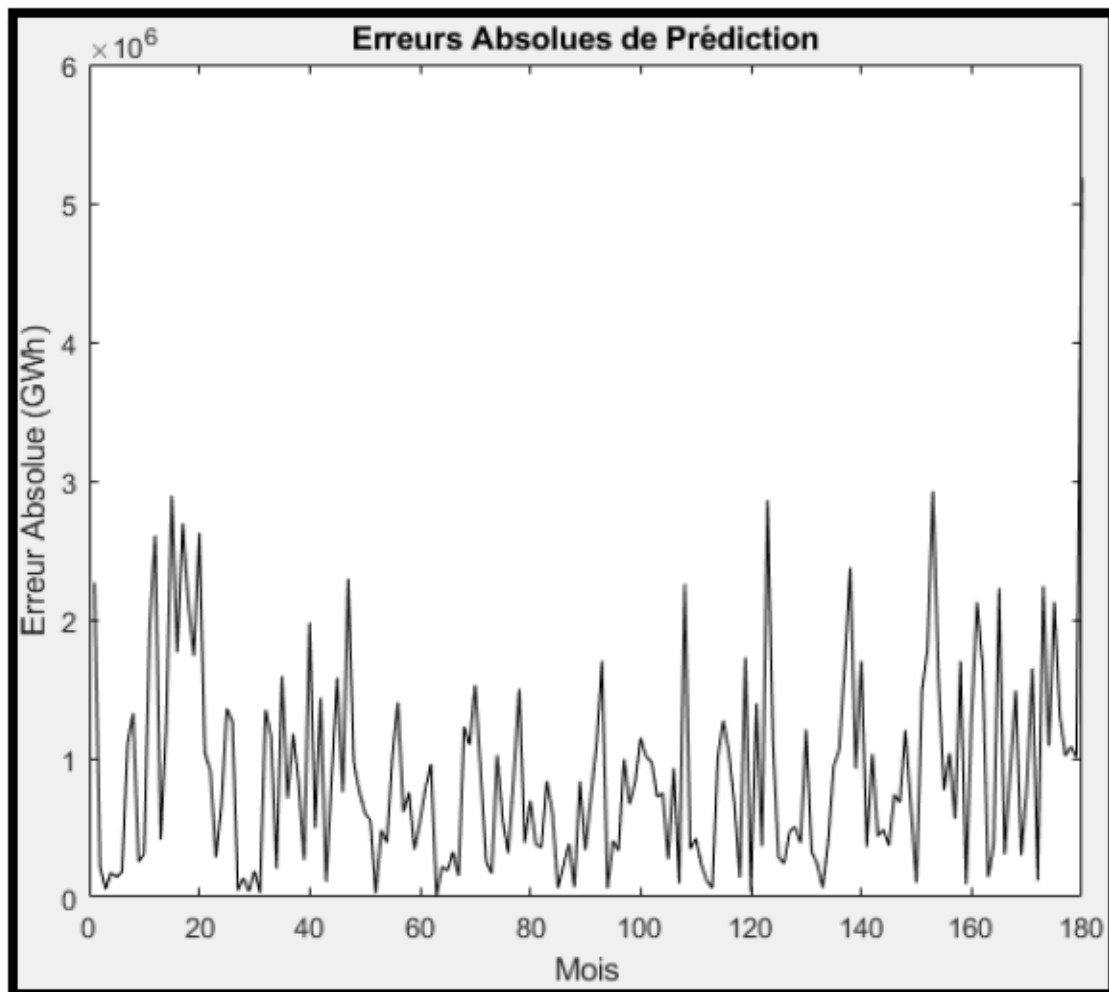


Figure 41 Absolues de Prédiction

Figure 41 : Ce graphique montre les erreurs absolues des prédictions du SVM sur la même période. Les erreurs absolues varient tout au long de la période, avec des valeurs souvent inférieures à 1×10^6 GWh.

Les erreurs absolues montrent une tendance à augmenter légèrement au fil du temps, ce qui peut être dû à des changements dans les conditions sous-jacentes ou à des limitations du modèle.

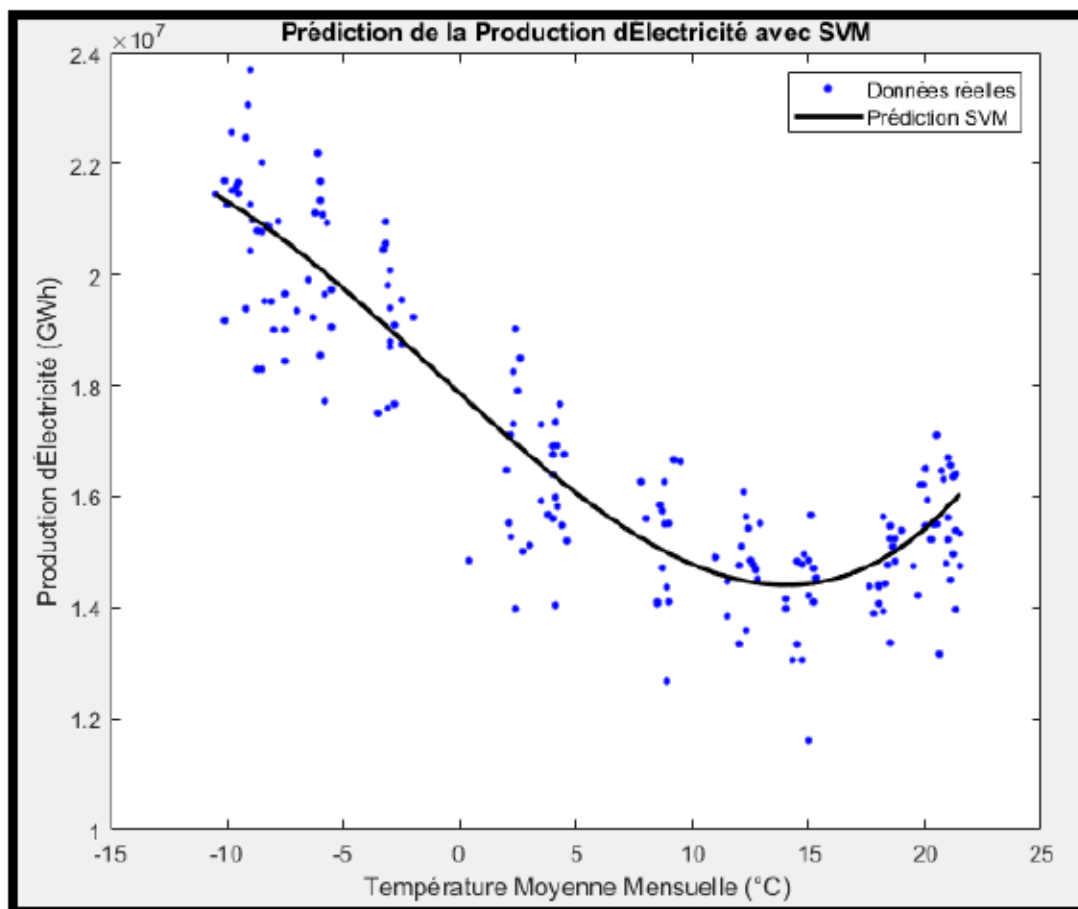


Figure 42 Prédiction de la Production d'Électricité en Fonction de la Température Moyenne Mensuelle Québec 2008-2022

Ce graphique compare les données réelles de production d'électricité (points bleus) et les prédictions du modèle SVM (ligne noire) en fonction de la température moyenne mensuelle.

➤ **Observations :**

- Il y a une relation non linéaire évidente entre la température moyenne mensuelle et la production d'électricité.
- Le modèle SVM avec noyau polynomial capture bien cette relation non linéaire, comme le montre la courbe ajustée qui suit de près les tendances des données réelles.

- Les prédictions sont particulièrement précises pour des températures entre -15°C et 15°C , avec une légère sous-estimation des valeurs extrêmes (très basses ou très hautes températures).

```

Erreur absolue moyenne (MAE) : 900546.22
Erreur quadratique moyenne (RMSE) : 1178523.29
Meilleurs paramètres :
- Box Constraint : 1.00
- Polynomial Order : 4.00

```

Figure 43 Les Hyperparamètres dans cette étude de SVM améliorer

Le modèle SVM amélioré par un noyau polynomial montre des performances robustes pour la prédiction de la production d'électricité, en capturant efficacement les relations non linéaires entre la température moyenne mensuelle et la production. Les erreurs quadratiques et absolues restent généralement faibles, bien qu'il y ait des pics notables vers la fin de la période analysée. Ces résultats suggèrent que le noyau polynomial aide à améliorer la précision des prévisions, mais que des ajustements supplémentaires pourraient être nécessaires pour minimiser les erreurs dans certaines conditions. Globalement, cette approche offre une méthode fiable et précise pour la modélisation et la prédiction de la production d'électricité en fonction des variations de température.

4.3. Forêt aléatoire (Random Forest) -RF- :

Dans cette partie, nous explorons l'utilisation de la Forêt Aléatoire (Random Forest) pour prédire la consommation d'électricité en utilisant des données de température qui sont donnée à titre explicative. Les données utilisées couvrent les années 2008 à 2021.

La Forêt Aléatoire est un modèle d'apprentissage supervisé basé sur des arbres de décision, qui permet de gérer efficacement des ensembles de données complexes et de grandes dimensions [24].

Les données utilisées dans cette étude comprennent les moyennes mensuelles des températures et les consommations d'électricité correspondantes pour chaque mois de

chaque année de 2008 à 2021. Les données sont présentées dans les tableaux suivants :

Tableau 9 Donnée utiliser dans le model RF-voir le fichier de programme pour l'ensemble des donnée-

Mois	Température Moyenne (°C)	Production Totale (MWh)
Janvier 2008	-10.1	19170108
Février 2008	-9.2	19387186
Mars 2008	-2.5	18745780
Avril 2008	3.8	15671767
Mai 2008	11.5	13834635
Juin 2008	17.6	14378476
...
Juillet 2021	21.3	16412180
Août 2021	20.5	17112685
Septembre 2021	15.0	14852517
Octobre 2021	8.7	15739034
Novembre 2021	2.3	17310658
Décembre 2021	-6.2	21114475

Les données sont divisées en deux ensembles : l'ensemble d'entraînement et l'ensemble de test. L'ensemble d'entraînement comprend 80 % des données totales, tandis que l'ensemble de test comprend les 20 % restants. Cette division permet d'évaluer la performance du modèle sur des données qu'il n'a pas vues lors de l'entraînement.

4.3.1. Modélisation avec la Forêt Aléatoire :

La Forêt Aléatoire est composée de multiples arbres de décision. Chaque arbre est entraîné sur un sous-ensemble aléatoire des données, ce qui permet de réduire le surapprentissage et d'améliorer la généralisation du modèle. Les hyperparamètres

principaux de la Forêt Aléatoire incluent le nombre d'arbres 'NumTrees' et la taille minimale des feuilles 'MinLeafSize' sur MATLAB.

Pour optimiser ces hyperparamètres, nous utilisons la recherche de grille (Grid Search) avec validation croisée. La validation croisée permet de diviser les données d'entraînement en plusieurs sous-ensembles (folds) et d'entraîner le modèle sur différents sous-ensembles pour obtenir une estimation plus robuste de la performance du modèle [25].

4.3.2. Exemple de Calcul Manuel :

Pour illustrer le fonctionnement d'une Forêt Aléatoire en utilisant l'exemple de prédiction de la consommation d'électricité basé sur les données de température. On va montrer comment les différents arbres de décision sont construits à partir de sous-ensembles des données et comment leurs prédictions sont combinées pour obtenir le résultat final.

Schéma de la Forêt Aléatoire pour la Prédiction de la Consommation d'Électricité :

- **Ensemble de données (Dataset)** : Les données de température et de consommation d'électricité sont divisées en sous-ensembles pour entraîner plusieurs arbres de décision.
- **Arbres de Décision** : Chaque arbre de décision est entraîné sur un sous-ensemble différent des données.
- **Résultats des Arbres de Décision** : Les prédictions de chaque arbre sont obtenues en utilisant les températures pour prédire la consommation d'électricité.
- **Combinaison des Prédictions** : Les prédictions de tous les arbres sont combinées par moyenne pour obtenir la prédiction finale de la consommation d'électricité.

4.3.3. Données d'Exemple :

Tableau 10 donnée d'exemple RF

Mois	Température Moyenne (°C)	Production Totale (MWh)
Janvier 2008	-10.1	19170108
Février 2008	-9.2	19387186
Mars 2008	-2.5	18745780
Avril 2008	3.8	15671767
Mai 2008	11.5	13834635
Juin 2008	17.6	14378476
Juillet 2008	20.9	14790878
Août 2008	19.7	14217588
Septembre 2008	14.3	13051507
Octobre 2008	8.5	14066823
Novembre 2008	2.1	15518440
Décembre 2008	-5.8	19652234

- Division en Sous-ensembles pour les Arbres de Décision :
 - ❖ **Arbre de Décision 1** : Utilise les données de Janvier à Juin 2008.
 - ❖ **Arbre de Décision 2** : Utilise les données de Juillet à Décembre 2008.
- Construction des Arbres de Décision :

Arbre de Décision 1 :

- Noeud 1 :

Si *Température* ≤ -5

\Rightarrow *Consommation* = *moyenne*(19170108, 19387186, 19652234)

- Températures: -10.1, -9.2, -5.8

- Consommations: 19170108, 19387186, 19652234

- Noeud 2 :

Si *Température* > -5

⇒ *Consommation* = *moyenne*(Noeud 1. Consommations)

- Températures : -2.5, 3.8, 11.5, 17.6, 20.9, 19.7, 14.3, 8.5, 2.1
- Consommations : 18745780, 15671767, 13834635, 14378476, 14790878, 14217588, 13051507, 14066823, 15518440

Arbre de Décision 2 :

- Noeud 1 :

Si *Température* ≤ 10

⇒ *Consommation* = *moyenne*(Noeud 2. Consommations)

- Températures: -10.1, -9.2, -2.5, 3.8, 8.5, 2.1, 11.5, 14.3
- Consommations: 19170108, 19387186, 18745780, 15671767, 13834635, 14066823, 15518440, 13051507, 13834635

- Noeud 2 :

Si *Température* > -5

⇒ *Consommation* = *moyenne*(14378476, 14790878, 14217588)

- Températures : 17.6, 20.9, 19.7
- Consommations : 14378476, 14790878, 14217588

4.3.4. Calcul des Prédictions Manuelles :

Pour une température de -7.5°C, nous allons faire les prédictions avec les deux arbres de décision et les combiner.

Arbre de Décision 1 :

Température = $-7.5^{\circ}\text{C} \leq -5^{\circ}\text{C}$

$$\Rightarrow \text{Consommation} = \text{moyenne}(19170108, 19387186, 19652234)$$

$$\Rightarrow \text{Consommation} = \frac{19170108, 19387186, 19652234}{3}$$

$$\Rightarrow \text{Consommation} = 19403176$$

Arbre de Décision 2 :

Température = $-7.5^{\circ}\text{C} \leq 10^{\circ}\text{C}$

$$\Rightarrow \text{Consommation}$$

$$= \text{moyenne}(19170108, 19387186, 18745780, 15671767,$$

$$14066823, 15518440, 13834635, 13051507)$$

$$\Rightarrow \text{Consommation}$$

$$= \frac{19170108, 19387186, 18745780, 15671767, 14066823, 15518440, 13834635, 13051507}{8}$$

$$\Rightarrow \text{Consommation} = 16168593.375$$

Combinaison des Prédictions :

La prédiction Finale pour *Température* = -7.5°C est :

$$\frac{\text{Prediction Arbre 1} + \text{Prediction Arbre 2}}{2}$$

$$\frac{19403176 + 16168593.375}{2} = 17785884.6875$$

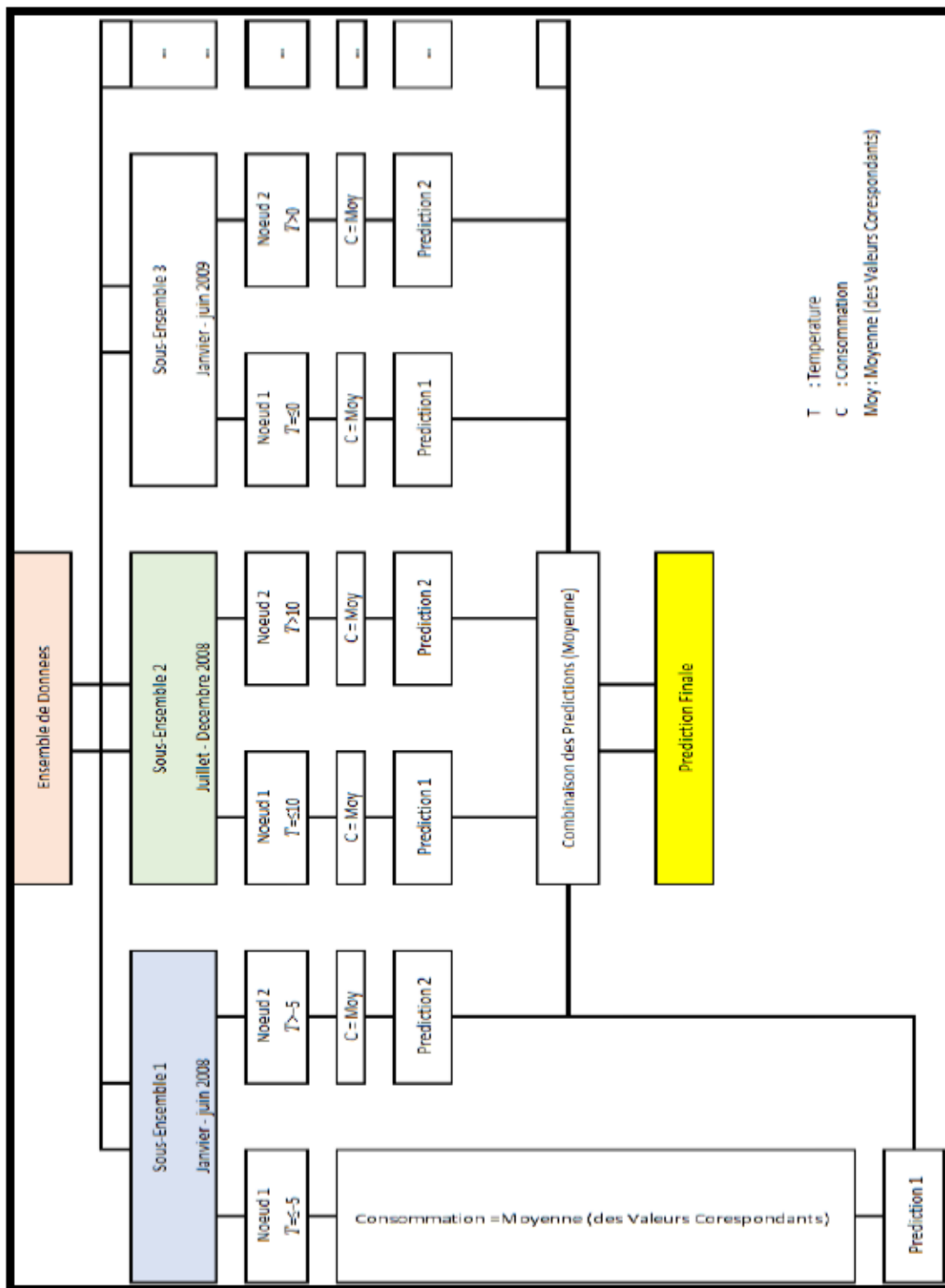


Figure 44 Le processus de la Forêt Aléatoire

4.3.5. Résultats sur le logiciel MATLAB :

➤ Graphique 1 : Prédications de Consommation d'Électricité :

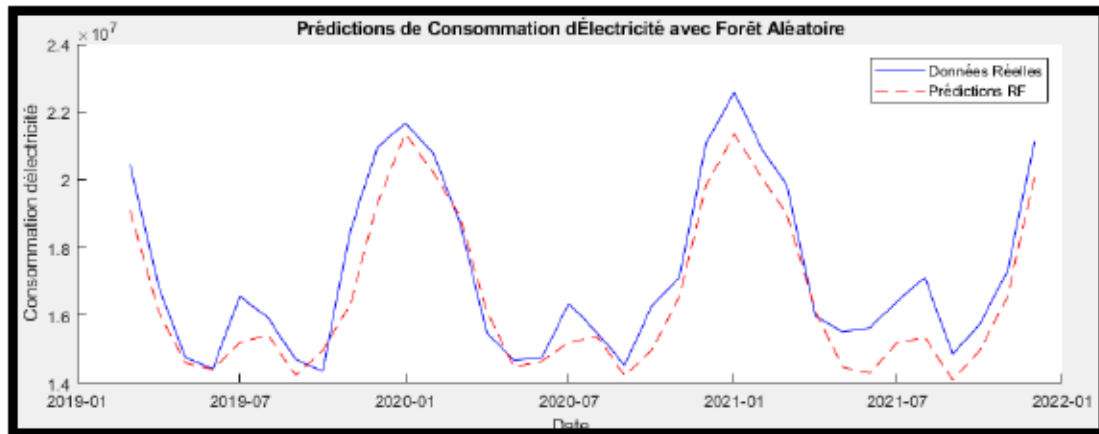


Figure 45 Prédications de Consommation d'Électricité par RF

Dans la figure 45 la ligne bleue représente les données réelles de consommation d'électricité sur une période donnée. Les variations saisonnières sont clairement visibles, avec des pics de consommation en hiver (autour de janvier) et des baisses de consommation en été (autour de juillet). Ces variations peuvent être attribuées à l'augmentation de l'utilisation du chauffage en hiver et une diminution de l'utilisation des appareils électroménagers en été.

La ligne rouge en tirets représente les prédictions de consommation d'électricité générées par le modèle de Forêt Aléatoire (RF). Le modèle semble suivre de près les tendances générales des données réelles, ce qui indique que la Forêt Aléatoire capture efficacement les motifs saisonniers de la consommation d'électricité. Cependant, on peut observer quelques décalages entre les prédictions et les valeurs réelles, particulièrement lors des pics de consommation. Ces décalages peuvent être dus à des facteurs non inclus dans le modèle ou à des limitations inhérentes aux données utilisées pour l'entraînement.

- **Précision des Prédictions :**

La comparaison visuelle entre les lignes bleues et rouges montre que, bien que les prédictions du modèle de Forêt Aléatoire soient globalement précises, il y a encore des marges d'amélioration. Par exemple, le modèle pourrait bénéficier de l'incorporation de variables explicatives supplémentaires telles que des données climatiques plus détaillées ou des variables économiques.

- **Améliorations Potentielles :**

Pour améliorer la précision des prédictions, plusieurs stratégies peuvent être envisagées :

1. Affiner les hyperparamètres de la Forêt Aléatoire par des méthodes comme la recherche de grille (Grid Search) ou l'optimisation bayésienne.
2. Ajouter des variables explicatives supplémentaires pour capturer des variations non expliquées par les données actuelles.
3. Envisager l'intégration de modèles hybrides combinant la Forêt Aléatoire avec d'autres techniques de modélisation comme les réseaux de neurones artificiels (ANN) ou les machines à vecteurs de support (SVM).

➤ **Graphique 2 : Erreur OOB (Out-of-Bag) en Fonction du Nombre d'Arbres :**

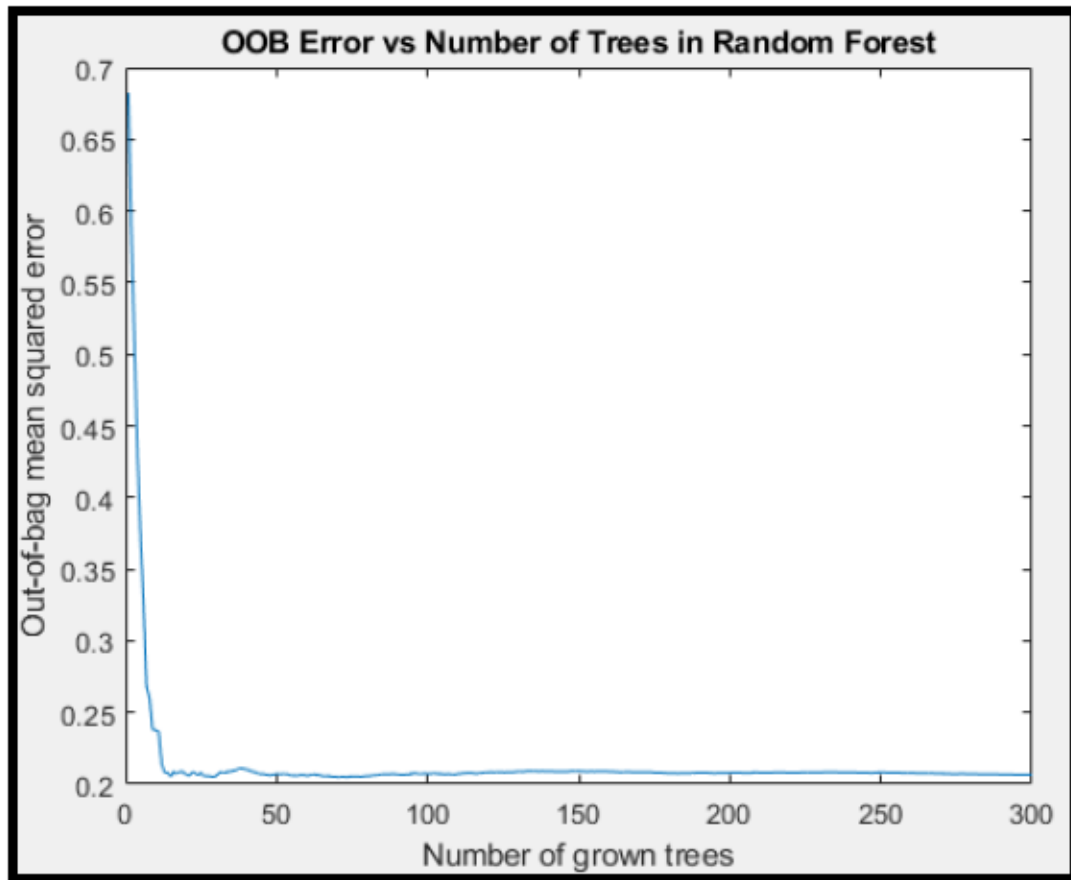


Figure 46 Tendence de l'Erreur OOB

L'axe des X du graphique représente le nombre d'arbres construits dans la Forêt Aléatoire, tandis que l'axe des Y montre l'erreur quadratique moyenne OOB (Out-of-Bag). L'erreur OOB (Figure 46) diminue rapidement avec l'ajout des premiers arbres, indiquant une amélioration rapide de la performance du modèle. Après environ 50 arbres, l'erreur OOB se stabilise et diminue très lentement avec l'ajout de nouveaux arbres.

Le graphique montre qu'après environ 50 arbres, l'amélioration de la précision est marginale. Cela suggère que le modèle atteint un bon compromis entre performance et complexité avec environ 50 arbres. Ajouter plus d'arbres au-delà de ce point n'apporte pas de gains significatifs en termes de réduction de l'erreur, mais augmente la

complexité computationnelle. Par conséquent, pour une implémentation pratique, il est recommandé de limiter le nombre d'arbres à environ 50 pour équilibrer la précision et l'efficacité.

➤ **Interprétation Générale:**

L'analyse des résultats de la Forêt Aléatoire en MATLAB montre que ce modèle est efficace pour prédire les tendances de consommation d'électricité, en capturant bien les variations saisonnières. Cependant, pour optimiser davantage les prédictions, on affine les hyperparamètres et d'incorporer des variables explicatives supplémentaires. En outre, la stabilisation rapide de l'erreur OOB avec un nombre limité d'arbres suggère que la Forêt Aléatoire est une méthode de modélisation robuste qui peut être optimisée sans nécessiter une complexité excessive.

En conclusion, l'utilisation de la Forêt Aléatoire pour prédire la consommation d'électricité se révèle prometteuse, offrant une bonne précision dans la capture des tendances saisonnières. Cependant, il reste des marges d'amélioration, notamment par l'optimisation des hyperparamètres et l'enrichissement des données. L'analyse des erreurs OOB montre également que le modèle peut être efficace avec un nombre raisonnable d'arbres, évitant ainsi une complexité computationnelle inutile.

5. Modèles Hybrides :

Dans les études et les analyses où les données jouent un rôle de plus en plus fondamental, les méthodes de prévision doivent évoluer pour répondre aux défis croissants de complexité et de variabilité. Les modèles hybrides émergent comme une solution efficace, combinant différentes techniques pour améliorer la précision et la robustesse des prévisions.

Les modèles hybrides reposent sur l'intégration de plusieurs méthodes de modélisation. Cette approche permet de tirer parti des points forts de chaque méthode tout en atténuant leurs faiblesses. Par exemple, les modèles statistiques comme ARIMA

(AutoRegressive Integrated Moving Average) sont très efficaces pour capturer les tendances linéaires et les saisons dans les séries temporelles. Cependant, ils peuvent rencontrer des difficultés à modéliser des relations non linéaires complexes. C'est là qu'interviennent les modèles d'apprentissage Machine tels que les réseaux de neurones artificiels (ANN) et les machines à vecteurs de support (SVM), qui excellent dans la capture de ces non-linéarités.

❖ **Différence de Nature des Modèles :**

L'intégration directe d'un modèle de réseau de neurones artificiels (ANN) à un modèle linéaire comme SARIMA peut poser plusieurs problèmes, principalement en raison de la nature fondamentalement différente de ces modèles.

- **SARIMA** : Modèle statistique linéaire utilisé pour capturer les tendances, les saisons, et les relations autorégressives dans les données temporelles. Il repose sur l'hypothèse que les données sont stationnaires ou peuvent être rendues stationnaires par des différenciations.
- **ANN** : Modèle non linéaire basé sur l'apprentissage automatique, capable de capturer des relations complexes et non linéaires dans les données.

Les modèles SARIMA génèrent des résidus (ou erreurs) qui sont supposés être des bruits blancs (i.e., des résidus sans corrélation sérielle). En intégrant directement ces résidus dans un ANN, on risque de ne pas utiliser efficacement la capacité de l'ANN à modéliser des relations non linéaires, car les résidus ne contiennent plus d'information structurée que l'ANN pourrait apprendre, ce qu'on l'appelle « **Problème d'Alignement des Caractéristiques** ».

L'utilisation d'un modèle intermédiaire comme SVM ou RF permet de combiner les avantages des deux mondes (linéaire et non linéaire).

SVM (Support Vector Machine) et RF (Random Forest) sont capables de capturer des patterns non linéaires dans les données tout en étant plus interprétables que les ANN. Ils agissent comme une passerelle entre les modèles linéaires et non linéaires, facilitant une meilleure exploitation des relations complexes.

L'une des combinaisons les plus courantes dans les modèles hybrides est l'intégration de SARIMA (ARIMA Saisonnière) avec des méthodes d'apprentissage Machine comme les SVM ou les forêts aléatoires (RF). SARIMA traite les composants saisonniers et les tendances à long terme, tandis que les SVM ou RF modélisent les relations complexes et non linéaires dans les résidus.

Les avantages des modèles hybrides sont multiples. Ils offrent une meilleure adaptabilité aux variations dans les données et peuvent capturer à la fois les relations linéaires et non linéaires. De plus, leur utilisation permet de réduire les erreurs de prévision et d'améliorer la fiabilité des résultats.

Les ANN, en particulier lorsqu'ils ne sont pas correctement configurés, sont susceptibles de sur-ajuster les données (Overfitting). Les modèles intermédiaires comme SVM et RF, avec leurs mécanismes de régularisation intégrés, peuvent aider à réduire ce risque en agissant comme des filtres.

2.3. Modèle Hybride -SARIMA-SVR- (approche en Cascade) :

Le modèle hybride SARIMA-SVR combine les capacités de deux modèles puissants : le modèle saisonnier autorégressif intégré à moyenne mobile (SARIMA) et les machines à vecteurs de support pour la régression (SVR). Cette combinaison vise à exploiter les forces des deux modèles pour améliorer la précision des prévisions des séries temporelles.

4.3.1. SARIMA :

Le modèle SARIMA est utilisé pour capturer les tendances linéaires et saisonnières dans les données de séries temporelles. Il est particulièrement efficace pour modéliser les patterns saisonniers et les dépendances temporelles.

- **Entrée** : Les données historiques de la série temporelle (par exemple, la production d'électricité).
- **Sortie** : Les prévisions des valeurs futures et les résidus (les différences entre les valeurs observées et les valeurs prévues par le modèle SARIMA).

4.3.2. SVR :

Le modèle SVR est utilisé pour modéliser les résidus non linéaires laissés par le modèle SARIMA. Le SVR aide à capturer les patterns complexes et les variations non linéaires dans les données, et les résidus sont utilisés comme nouvelles observations.

- **Entrée** : Les résidus générés par le modèle SARIMA.

$$e_t = y_t - y'_t$$

Avec : y'_t est la valeur prédite par le modèle SARIMA

- **Sortie** : Les prévisions des résidus, qui sont ensuite ajoutées aux prévisions SARIMA pour obtenir les prévisions finales du modèle hybride.

$$y'_{t,Hybride} = c$$

Avec :

- $y'_{t,SARIMA}$: prévision du modèle SARIMA
- $e'_{t,SVM}$: ajustement des résidus par le modèle SVM

Les résultats sont présentés sous forme de graphiques pour visualiser les différentes étapes et performances du modèle hybride SARIMA-SVM.

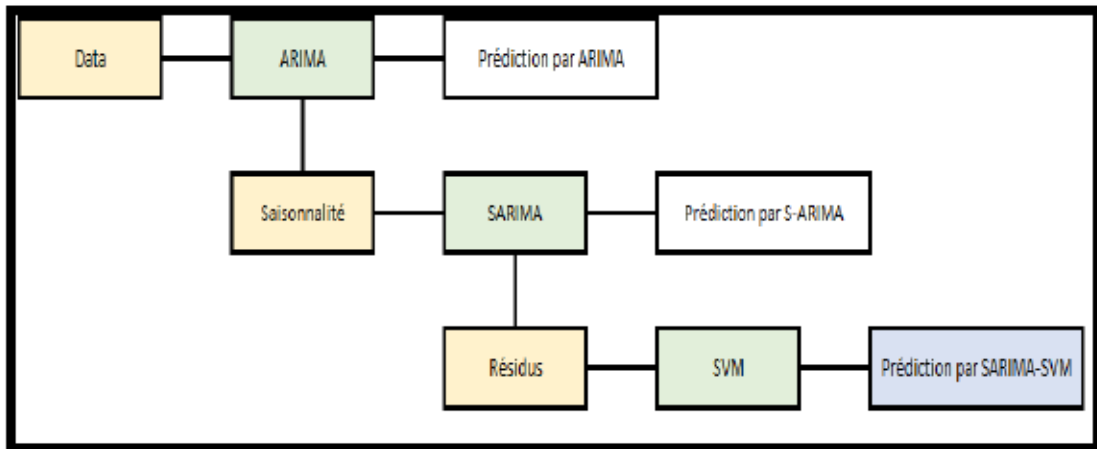


Figure 47 Schéma du Modèle Hybride approche en Cascade

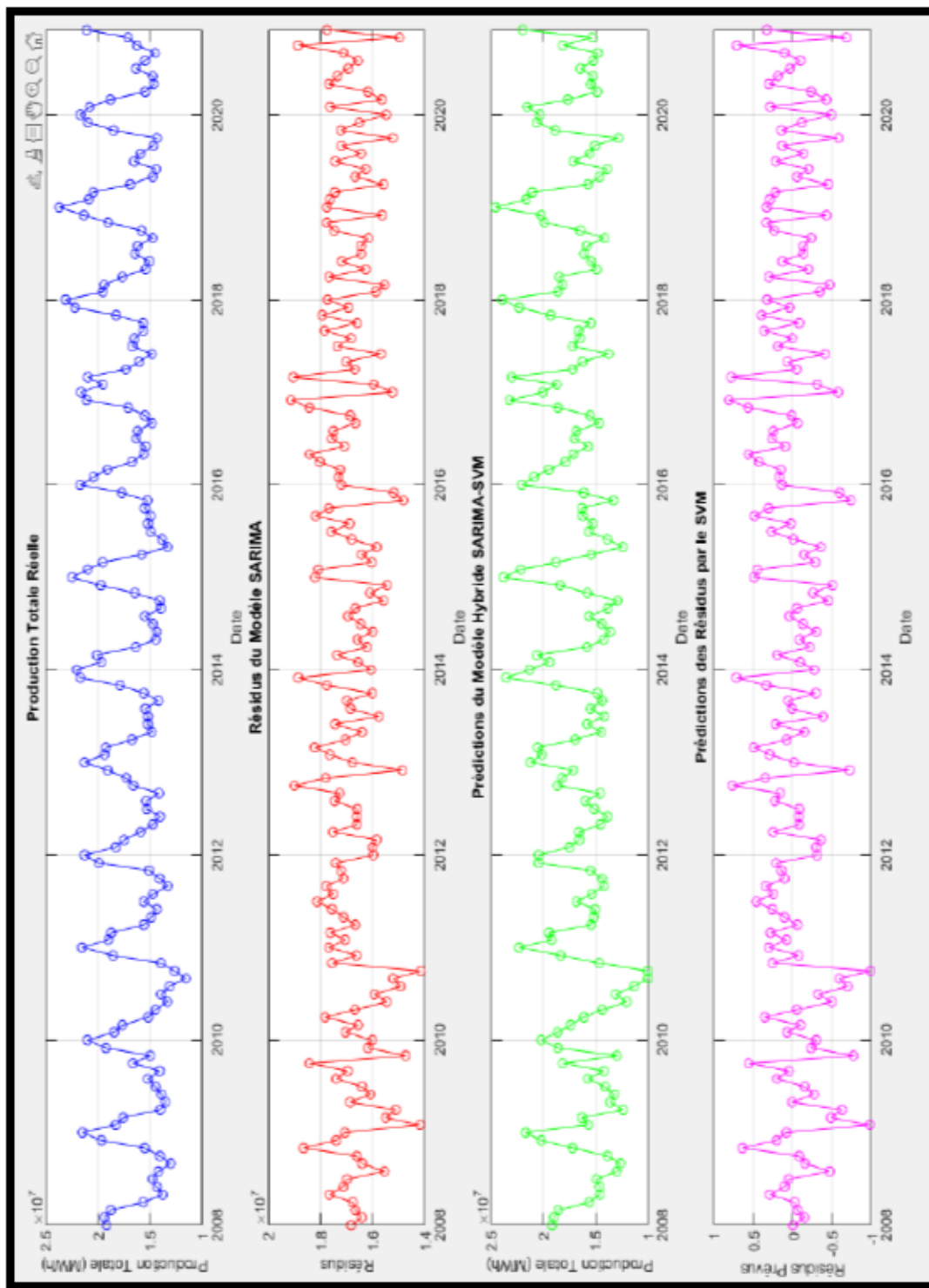


Figure 48 Modèle Hybride SARIMA-SVM

✓ **Production Totale Réelle (Graphique 1 : Bleu) :**

La figure 48 montre la production totale réelle d'électricité en MWh de 2008 à 2020, les points bleus représentent les valeurs observées mensuelles et on remarque que Les fluctuations saisonnières et les tendances au fil du temps sont visibles.

✓ **Résidus du Modèle SARIMA (Graphique 2 : Rouge) :**

Ce graphique montre les résidus obtenus après l'ajustement du modèle SARIMA aux données de production normalisées, les points rouges représentent les résidus calculés. Les résidus révèlent les différences entre les valeurs observées et les valeurs prévues par le modèle SARIMA, indiquant des erreurs et des structures non capturées par SARIMA.

✓ **Prédictions du Modèle Hybride SARIMA-SVM (Graphique 3 : Vert) :**

Ce graphique montre les prédictions finales obtenues en combinant le modèle SARIMA et le modèle SVM, Les points verts représentent les valeurs prévues par le modèle hybride SARIMA-SVM. Les prédictions suivent globalement les tendances des observations réelles avec des ajustements par le SVM pour corriger les erreurs du modèle SARIMA.

✓ **Prédictions des Résidus par le SVM (Graphique 4 : Magenta) :**

Ce graphique montre les prédictions des résidus obtenues par le modèle SVM, les points magenta représentent les résidus prédits par le SVM. Les prédictions des résidus montrent la capacité du SVM à capturer les fluctuations résiduelles non modélisées par SARIMA.

✓ **Comparaison des Observations Réelles, Prédictions SARIMA-SVM et Observations d'entraînement SVM (Graphique 5) :**

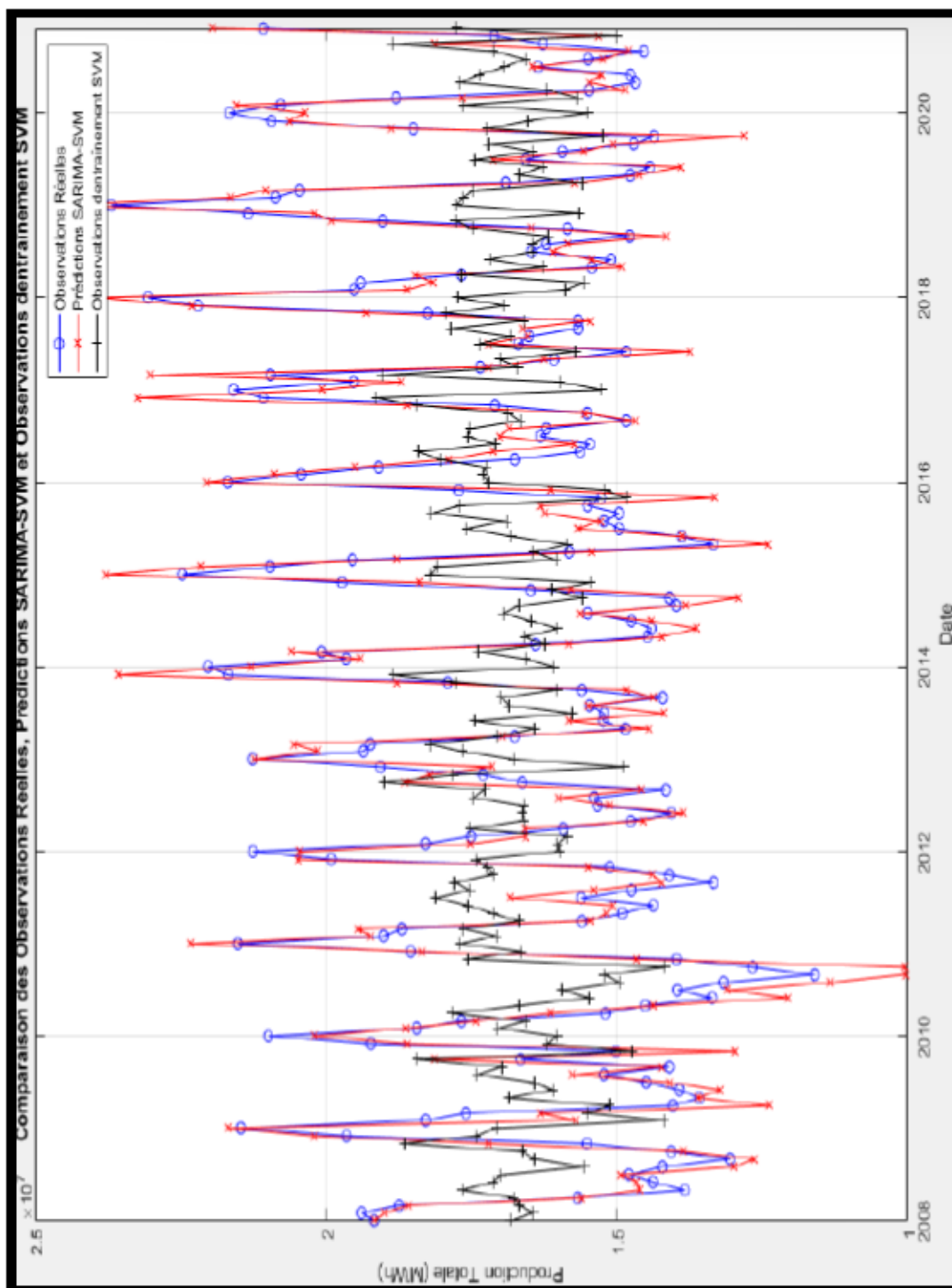


Figure 49 Comparaison des Observations Réelles, Prédictions SARIMA-SVM et Observations d'entraînement SVM

Ce graphe (Figure 49) compare les observations réelles, les prédictions du modèle hybride SARIMA-SVM et les observations d'entraînement SVM :

- Les points bleus représentent les observations réelles.
- Les points rouges représentent les prédictions SARIMA-SVM.
- Les points noirs représentent les observations d'entraînement SVM (les résidus du modèle SARIMA).

La comparaison montre que les prédictions SARIMA-SVM suivent de près les observations réelles, indiquant une bonne performance du modèle hybride, tandis que les observations d'entraînement SVM montrent les résidus ajustés par le SVM.

La **production totale réelle** montre des variations saisonnières claires avec des pics et des creux récurrents chaque année, ainsi que des tendances à la hausse et à la baisse observées sur plusieurs années. Les **résidus du modèle SARIMA** révèlent des fluctuations non capturées par le modèle SARIMA seul, avec des pics et des vallées indiquant des erreurs systématiques non modélisées.

Les **prédictions du modèle hybride SARIMA-SVM** démontrent une amélioration par rapport au modèle SARIMA seul, capturant mieux les tendances globales des données réelles. Les ajustements effectués par le SVM réduisent les erreurs résiduelles et améliorent la précision des prévisions.

En ce qui concerne les **prédictions des résidus par le SVM**, elles montrent que le SVM capture efficacement les variations non linéaires et les patterns complexes dans les résidus, permettant ainsi de corriger les erreurs non modélisées par SARIMA.

La **comparaison des observations et des prédictions** indique que les prédictions du modèle hybride SARIMA-SVM suivent de près les observations réelles, démontrant une bonne performance globale et une amélioration significative de la précision des prévisions grâce aux ajustements apportés par le SVM.

4.3.3. Fusion Tri-Modèle -SARIMA-SVM-RF-(Meta-Apprentissage):

Dans cette partie, nous explorons l'intégration du modèle Forêt aléatoire (RF) dans un modèle hybride existant SARIMA-SVM.

Le modèle hybride que nous construisons vise à améliorer la performance des prévisions en combinant le modèle AutoRégressif Intégré et Mobile Saisonnière (SARIMA), la Machine à Vecteurs de Support (SVM) et l'algorithme Random Forest (RF) avec une approche de Meta-Apprentissage.

La complémentarité des modèles est un facteur clé. SARIMA est efficace pour capturer les tendances saisonnières et linéaires, tandis que SVM excelle dans la modélisation des relations non linéaires. RF, quant à lui, renforce l'ensemble en réduisant la variance et en capturant des interactions complexes qui pourraient échapper aux autres modèles.

Nous effectuerons les calculs sur un petit échantillon pour que le processus soit gérable et compréhensible seulement dans l'étude théorique pour avoir l'idée comment est le fonctionnement global.

Cette partie de recherche a pour objectif d'illustrer comment chaque modèle contribue à la prévision finale dans le Meta-apprentissage, en utilisant des données réelles et approximative.

4.3.4. Étape 1 : Prétraitement des Données :

Avant de modéliser les données, il est essentiel de les prétraiter pour s'assurer qu'elles sont adaptées à la modélisation. Nous commencerons par normaliser les données de température et de production d'électricité.

➤ **Données d'Échantillon (Janvier 2008 à Mars 2008):**

- Température : [-10.1, -9.2, -2.5, 3.8, 11.5]
- Production d'électricité :

[19170108, 21466207, 20994084, 21519681, 21261463]

Normalisation :

La normalisation consiste à ajuster les valeurs des données pour qu'elles aient une moyenne de zéro et un écart type de 1. Cela facilite la modélisation et améliore la performance des modèles.

Supposons :

- Moyenne des données de température : $\mu_{temp} = -1.3$
- Écart type des données de température : $\sigma_{temp} = 8.62$
- Moyenne des données de production : $\mu_{prod} = 20802208.6$
- Écart type des données de production : $\sigma_{prod} = 894065.3$

Valeurs normalisées :➤ **Température Normalisée :**

$$\text{Temperature Normalisee } (TN_t) = \frac{\text{Temperature} - \mu_{temp}}{\sigma_{temp}}$$

$$TN_1 = \frac{-10.1 - (-1.3)}{8.62} = -1.02$$

$$TN_4 = \frac{3.8 - (-1.3)}{8.62} = 0.59$$

$$TN_2 = \frac{-9.2 - (-1.3)}{8.62} = -0.92$$

$$TN_5 = \frac{11.5 - (-1.3)}{8.62} = 1.49$$

$$TN_3 = \frac{-2.5 - (-1.3)}{8.62} = -0.14$$

➤ **Production Normalisée :**

$$\text{Production Normalisee } (PN) = \frac{\text{Production} - \mu_{prod}}{\sigma_{prod}}$$

$$PN_1 = \frac{19170108 - 20802208.6}{894065.3} \\ = -1.83$$

$$PN_2 = \frac{21466207 - 20802208.6}{894065.3} \\ = 0.74$$

$$PN_3 = \frac{20994084 - 20802208.6}{894065.3}$$

$$= 0.21$$

$$PN_5 = \frac{21261463 - 20802208.6}{894065.3}$$

$$= 0.51$$

$$PN_4 = \frac{21519681 - 20802208.6}{894065.3}$$

$$= 0.8$$

4.3.5. Étape 2 : Modèle SARIMA :

Une fois les données normalisées, nous appliquons le modèle SARIMA pour capturer les tendances saisonnières et linéaires des données. Le modèle SARIMA est un modèle de série temporelle qui combine des composants autorégressifs (AR), de moyennes mobiles (MA) et de différenciation saisonnière.

- Entré : Série temporelle Y_t
- Sortie : Prévion $y'_{t,SARIMA}$ et Résidus $e_{t,SARIMA}$

Le variable Y_t inclut toutes les caractéristiques originales (comme la température normalisée et Production d'Énergie ...) à l'instant t et y_t représente une partie de la série temporelle (Production d'Énergie Éclectique).

En supposant un modèle simplifié SARIMA (2,1,2) (1,1,1), le modèle peut être représenté comme suit :

$$(1 - \phi_1 L - \phi_2 L^2)(1 - \phi_{12} L^{12})(1 - L)(1 - L^{12})Y_t$$

$$= c + (1 + \theta_1 L + \theta_2 L^2)(1 + \theta_{12} L^{12})\varepsilon_t$$

Pour les besoins de cette démonstration, nous utiliserons les paramètres hypothétiques suivants :

- $\phi_1 = 0.0632$
- $\phi_2 = 0.4470$
- $\theta_1 = -0.4583$
- $\theta_2 = 0.2235$
- $\phi_{12} = 0.5122$
- $\theta_{12} = 0.3806$
- Période saisonnière $s = 12$
- $\varepsilon_t = 0.2$

Pour Avril et Mai:

$$y'_{t,SARIMA} = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \phi_{12} y_{t-s} + \theta_{12} \varepsilon_{t-s}$$

Calculer $y'_{t,SARIMA}$ pour février et mars :

- Pour Avril $t = 4$:

$$y'_{4,SARIMA} = \Phi_1 y_3 + \Phi_2 y_2 + \theta_1 \varepsilon_3 + \theta_2 \varepsilon_2$$

$$\begin{aligned} y'_{4,SARIMA} &= 0.0632 \times 0.21 - 0.4470 \times 0.74 - 0.4583 \times 0.2 + 0.2235 \times 0.2 \\ &= -0.2358 \end{aligned}$$

- Pour Mai $t = 5$:

$$\begin{aligned} y'_{5,SARIMA} &= 0.0632 \times 0.8 - 0.4470 \times 0.21 - 0.4583 \times 0.2 + 0.2235 \times 0.2 \\ &= -0.0903 \end{aligned}$$

4.3.6. Étape 3 : Modèle SVM :

Pour cette étape, nous allons utiliser un modèle SVM polynomial pour capturer les relations non linéaires dans les résidus SARIMA.

- **Entré** : Résidus $e_{t,SARIMA}$
- **Sortie** : Préviation des résidus par SVM $e'_{t,SVM}$

Étant donné les résidus SARIMA :

$$e_{t,SARIMA} = y_t - y'_{t,SARIMA}$$

- $e_{4,SARIMA} = 0.80 - (-0.2358) = 1.0358$
- $e_{5,SARIMA} = 0.51 - (-0.0903) = 0.6003$

Supposons un modèle SVM polynomial avec les paramètres suivants $w = 0.6$ et $b = 0.1$ et un degré de 2, donc l'Équation du modèle SVM polynomial :

$$e'_{t,SVM} = (we_{t,SARIMA} + b)^d$$

$$e'_{4,SVM} = (0.6 \times 1.0358 + 0.1)^2 = (0.6215 + 0.1)^2 = 0.7215^2 = 0.5205$$

$$e'_{4,SVM} = (0.6 \times 0.6003 + 0.1)^2 = (0.3602 + 0.1)^2 = 0.4602^2 = 0.2118$$

4.3.7. Étape 4 : Modèle Forêt aléatoires (Random Forest) :

Entraîner le modèle RF sur les prévisions faites par SARIMA et SVM et les données originales. Cette méthode traite RF comme une étape finale, affinant la prévision en apprenant à partir de la sortie combinée des modèles précédents.

- Entré : Série temporelle Y_t et Prévision du Model Hybride $y'_{t,SARIMA_SVM}$
- Sortie : Prévision finale $y'_{t,SARIMA_SVM_RF}$

Avec :

$y'_{t,Hybride} = y'_{t,SARIMA_SVM} = f(y'_{t,SARIMA}, e'_{t,SVM})$: Les prédictions SARIMA-SVM

La prédiction finale en utilisant Random Forest (RF) est une combinaison des prédictions de SARIMA-SVM et des caractéristiques originales. L'équation finale du Meta-apprentissage peut être représentée comme suit :

$$y'_{t,SARIMA_SVM_RF} = RF(Y_t, y'_{t,Hybride})$$

Cette équation capture l'approche en cascade où chaque modèle contribue à affiner la prédiction finale en intégrant différentes dynamiques des données. Cette méthode illustre la puissance du Meta-apprentissage pour améliorer la précision des prévisions de séries temporelles.

Dans notre exemple de calcul manuel, pour simplifier la compréhension, Supposons que le modèle RF utilise une moyenne simple de la température et des prédictions SARIMA-SVM pour simplifier l'illustration (en réalité, RF utiliserait une approche

basée sur des arbres plus complexe comme illustre dans les programmes informatiques sur le logiciel MATLAB), n représente nombre d'arbre :

$$y'_{t,RF} = \frac{\text{Temperature Normalisee } (TN_t) + y'_{t,SARIMA_SVM}}{n}$$

$$y'_{4,RF} = \frac{0.59 + 0.2847}{2} = 0.4374$$

$$y'_{5,RF} = \frac{1.49 + 0.1215}{2} = 0.80575$$

4.3.8. ETAPE 5 : Prédiction Finales :

Prédiction finale du modèle hybride qu'on va l'appeler « Fusion Tri-Modèle » pour avril et mai :

$$y'_{t,Tri_Modele}(Normalisee) = y'_{t,SARIM_SVM_RF} = 0.2847 + 0.4374 = 0.7221$$

$$y'_{t,Tri_Modele}(Normalisee) = 0.2847 + 0.4374 = 0.7221$$

4.3.9. ETAPE 6 : Dénormalisation des prédictions finales :

$$Prediction\ Final_t = y'_{t,Tri_Modele} = y'_{t,SARIM_SVM_RF} \times \sigma_{prod} + \mu_{prod}$$

$$y'_{4,Tri_Modele} = 0.7221 \times 894065.3 + 20802208.6 = 21504092.5 \text{ MWh}$$

$$y'_{5,Tri_Modele} = 0.92725 \times 894065.3 + 20802208.6 = 22189642.4 \text{ MWh}$$

4.3.10. ÉTAPE FINAL : Interprétation Des Résultats Des Études Pratique Sur Le Logiciel MATLAB :

Dans cette étape, nous allons interpréter les résultats obtenus à partir de notre modèle SARIMA-SVM-RF en utilisant les graphiques générés par MATLAB.

Le graphique ci-dessus montre une comparaison entre les observations réelles de la production d'électricité (cercles bleus) et les prédictions du modèle hybride SARIMA-SVM-RF (lignes rouges avec des croix).

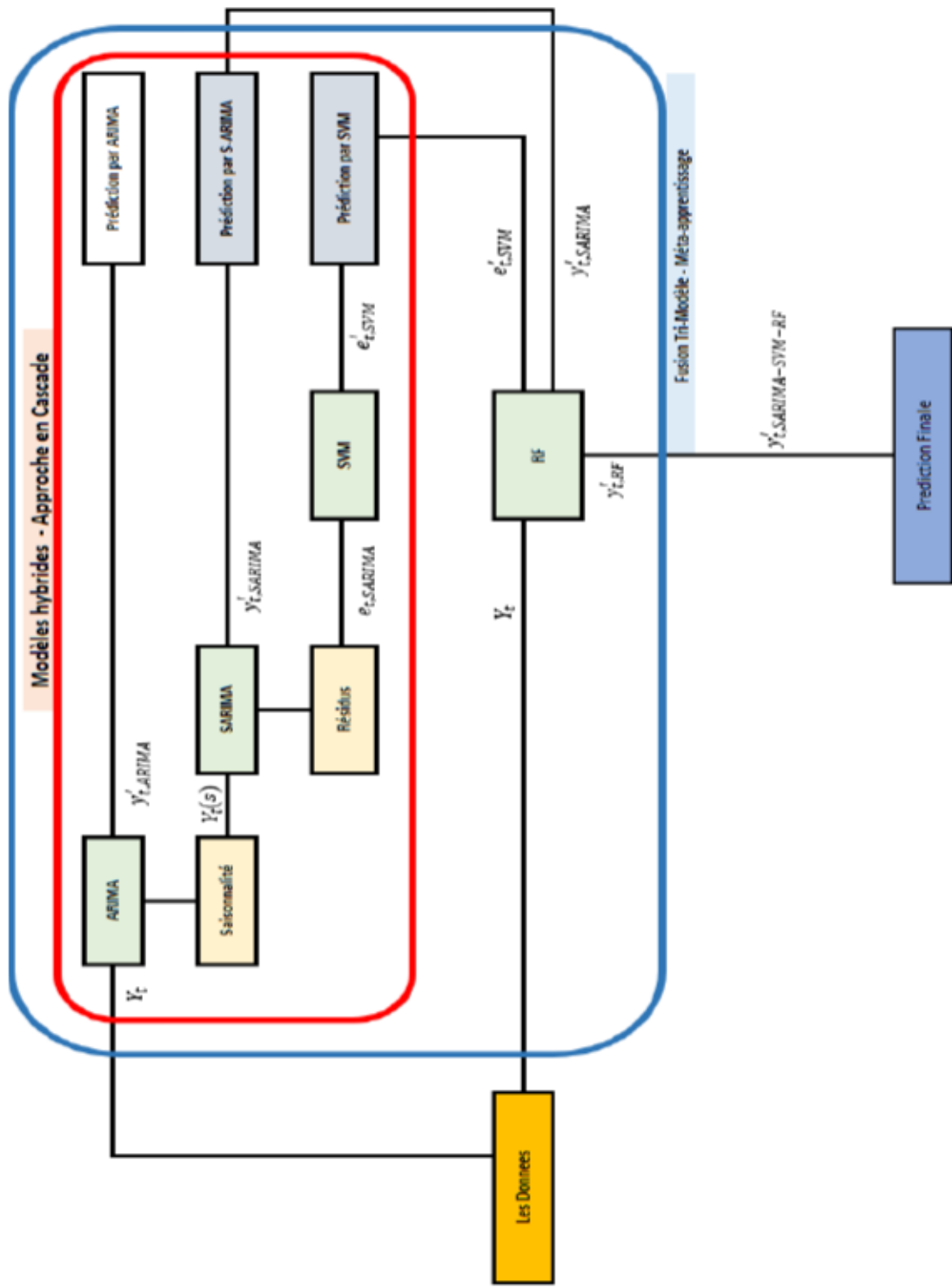


Figure 50 Schémas explicative du Processus d'étude

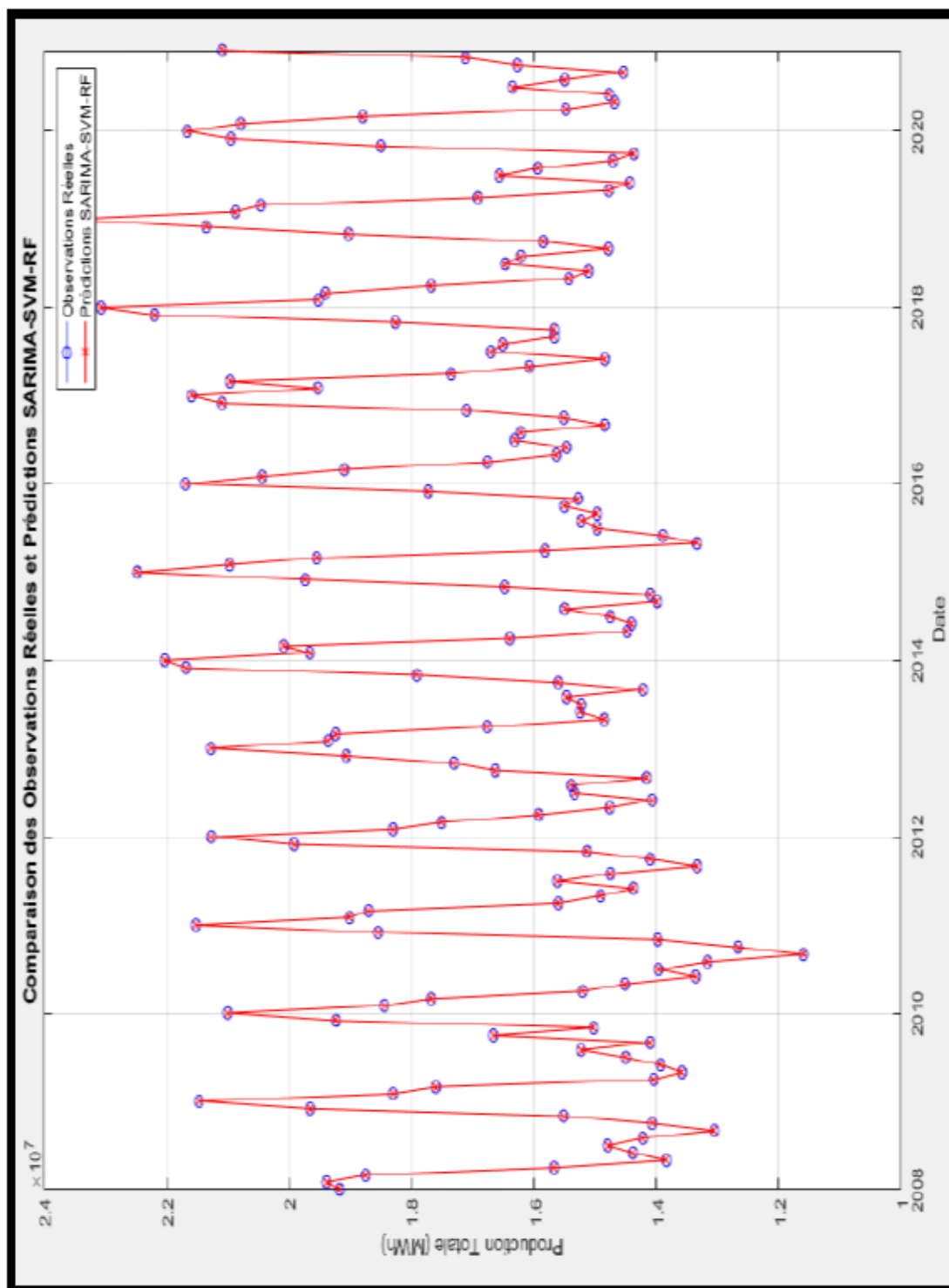


Figure 51 les prédictions du modèle SARIMA-SVM-RF

- **Observations Réelles** : Les points bleus représentent les valeurs réelles de la production d'électricité pour chaque mois de la période considérée.
- **Prédictions SARIMA-SVM-RF** : Les lignes rouges montrent les prédictions faites par le modèle hybride.

Les prédictions (Figure 51) suivent de près les observations réelles, ce qui indique que la Fusion Tri-Modèle (SARIMA-SVM-RF) est capable de capturer les tendances et les variations saisonnières de la production d'électricité.

❖ **Détail des Résultats par Modèle :**

Ce graphique (Figure 52) montre les prédictions et les résidus pour chaque modèle utilisé dans le Meta-apprentissage.

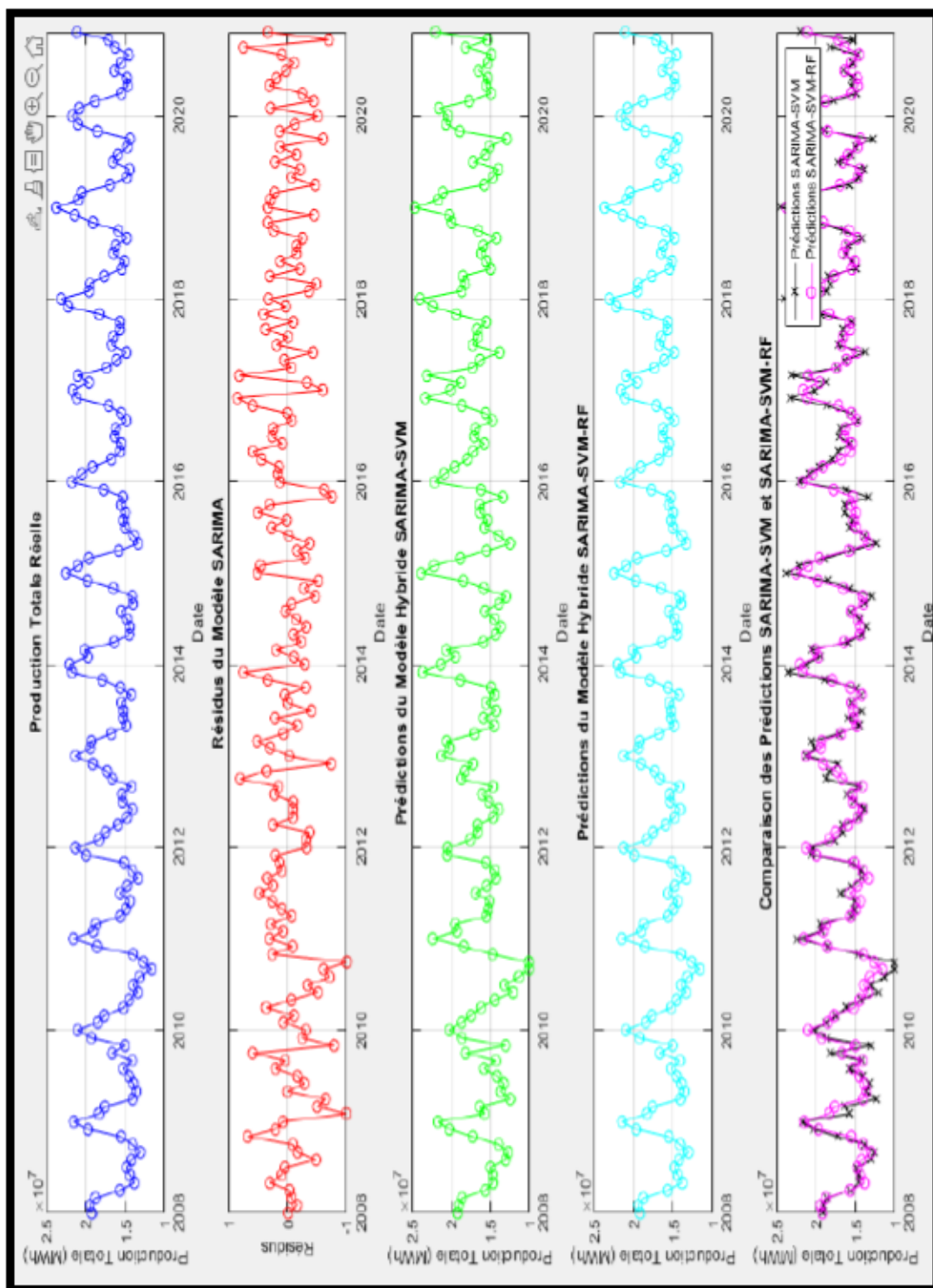


Figure 52 les prédictions et les résidus pour chaque modèle

✓ **Production Totale Réelle :**

La première sous-figure représente les observations réelles de la production d'électricité (points bleus). C'est la référence contre laquelle nous évaluons nos prédictions.

✓ **Résidus du Modèle SARIMA :**

La deuxième sous-figure montre les résidus du modèle SARIMA (points rouges). Les résidus représentent la différence entre les observations réelles et les prédictions du modèle SARIMA. Ces résidus sont utilisés comme entrée pour le modèle SVM.

✓ **Prédictions du Modèle Hybride SARIMA-SVM :**

La troisième sous-figure présente les prédictions du modèle Hybride SARIMA-SVM (points verts). Ce modèle utilise les résidus du modèle SARIMA et les données de température pour affiner les prédictions.

✓ **Prédictions du Modèle Hybride SARIMA-SVM-RF :**

La quatrième sous-figure montre les prédictions du modèle hybride final SARIMA-SVM-RF (points cyan). Ce modèle combine les prédictions SARIMA-SVM avec les données originales de température pour produire des prévisions plus précises.

✓ **Comparaison des Prédictions SARIMA-SVM et SARIMA-SVM-RF :**

La dernière sous-figure compare directement les prédictions des modèles SARIMA-SVM (lignes noires avec croix) et SARIMA-SVM-RF (lignes magenta avec cercles). La comparaison montre que le modèle SARIMA-SVM-RF améliore les prédictions par rapport au modèle SARIMA-SVM seul, en capturant mieux les variations de la production d'électricité.

❖ **Explication de l'Approche en Cascade et du Méta-Apprentissage :**

a- Approche en Cascade :

L'approche en cascade implique l'utilisation séquentielle de plusieurs modèles, où les sorties d'un modèle servent d'entrées pour le modèle suivant. Dans ce projet, nous avons utilisé l'approche en cascade avec les modèles SARIMA et SVM :

- **Modèle SARIMA (Modèle de Base) :** Capture les tendances saisonnières et linéaires des données de production d'électricité.
- **Modèle SVM (Hybride avec SARIMA) :** Utilise les résidus du modèle SARIMA pour capturer les relations non linéaires dans les données.

b- Fusion Tri-Modèle avec Méta-Apprentissage :

La fusion tri-modèle avec méta-apprentissage combine les prédictions des modèles SARIMA-SVM et les données originales pour entraîner le modèle Arbre Aléatoires RF. Ce processus permet d'exploiter les forces de chaque modèle pour améliorer la précision des prévisions.

- **Modèle SARIMA-SVM en Cascade :** Les résidus du modèle SARIMA sont utilisés comme entrée pour le modèle SVM, créant ainsi une combinaison hybride.
- **Modèle Random Forest (Fusion Tri-Modèle) :** Utilise les données originales de température et les prédictions combinées SARIMA-SVM pour fournir des prévisions finales.

Les résultats obtenus montrent que l'approche de méta-apprentissage utilisant les modèles SARIMA, SVM polynomial et Random Forest permet d'améliorer significativement les prévisions de la production d'électricité. Les modèles individuels capturent différentes composantes des données, et leur combinaison permet de tirer parti de leurs forces respectives.

- **Modèle SARIMA** : Capture les tendances saisonnières et linéaires.
- **Modèle SVM Polynomial** : Capture les relations non linéaires dans les résidus du modèle SARIMA.
- **Modèle Random Forest** : Intègre les prédictions combinées avec les données originales pour améliorer la précision globale.

Les graphiques illustrent clairement l'amélioration progressive des prédictions à chaque étape de l'approche hybride, culminant avec le modèle SARIMA-SVM-RF qui offre les meilleures performances.

❖ La validation de la Fusion Tri-Modèle :

Nous allons suivre un processus structuré pour entraîner, valider et évaluer les performances du modèle sur un ensemble de données de consommation énergétique mensuelle. L'objectif est de démontrer l'efficacité de ce modèle hybride et de s'assurer de sa robustesse.

Commençant par la division des données en ensembles d'entraînement et validation (80%) et de test (20%), dans notre cas les données de 2008 à 2016 sont réservées pour l'entraînement et validation (Figure 53) et de 2017 à 2021 pour le test.

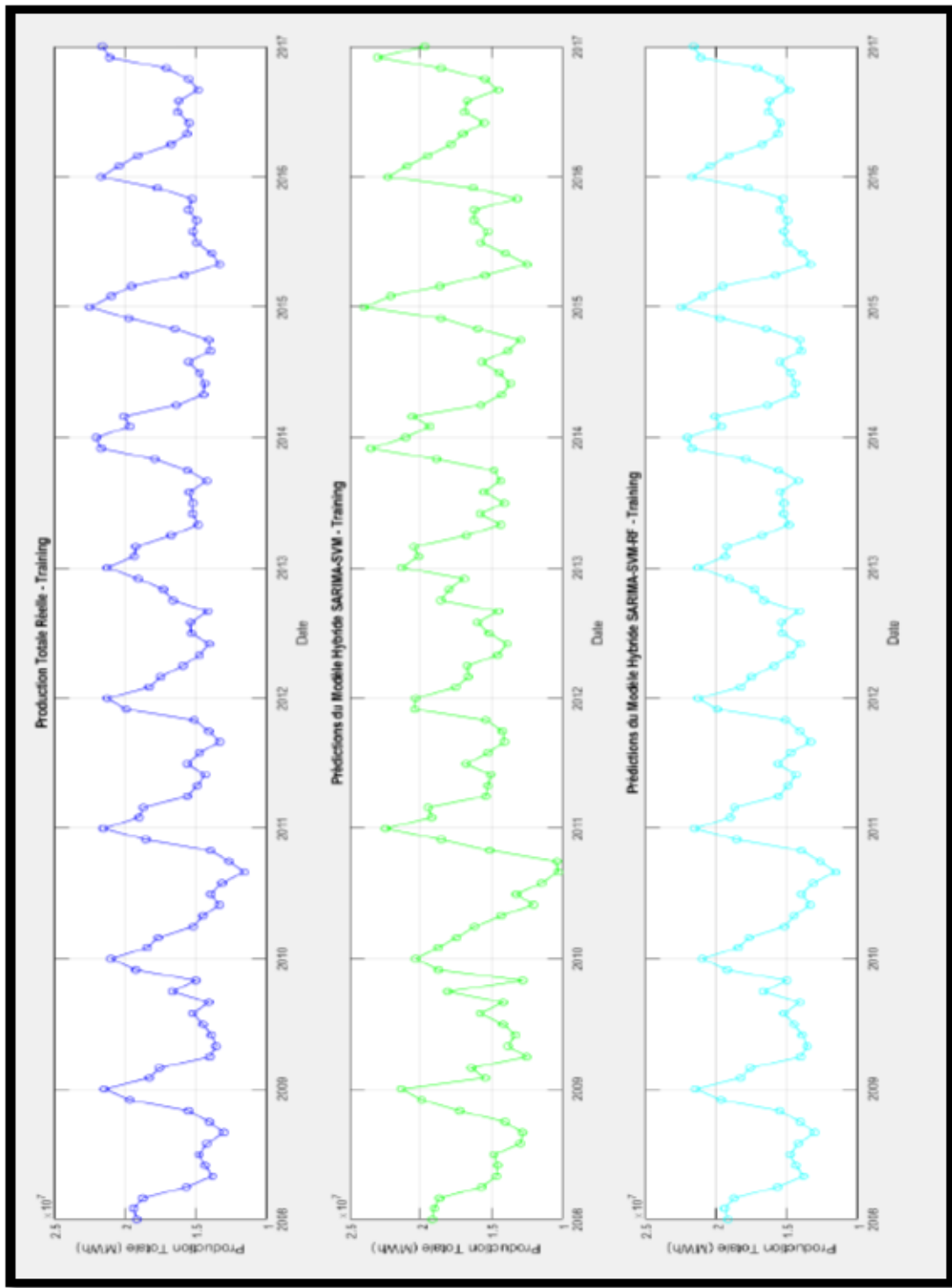


Figure 53 Production Totale Réelle et Prédictions – Entraînement

➤ Production Totale Réelle - Entraînement :

On observe une variation saisonnière avec des pics et des creux réguliers, ce qui est typique pour les données de consommation énergétique.

➤ Prédiction du Modèle Hybride SARIMA-SVM - Entraînement :

Les prédictions suivent globalement la tendance des données réelles, mais avec quelques divergences qui peuvent être attribuées à des résidus capturés par le modèle SVM.

➤ Prédiction du Modèle Hybride SARIMA-SVM-RF - Entraînement :

Ce modèle semble suivre plus étroitement les variations des données réelles par rapport au modèle SARIMA-SVM seul, indiquant une meilleure capture des motifs et des résidus par le modèle RF.

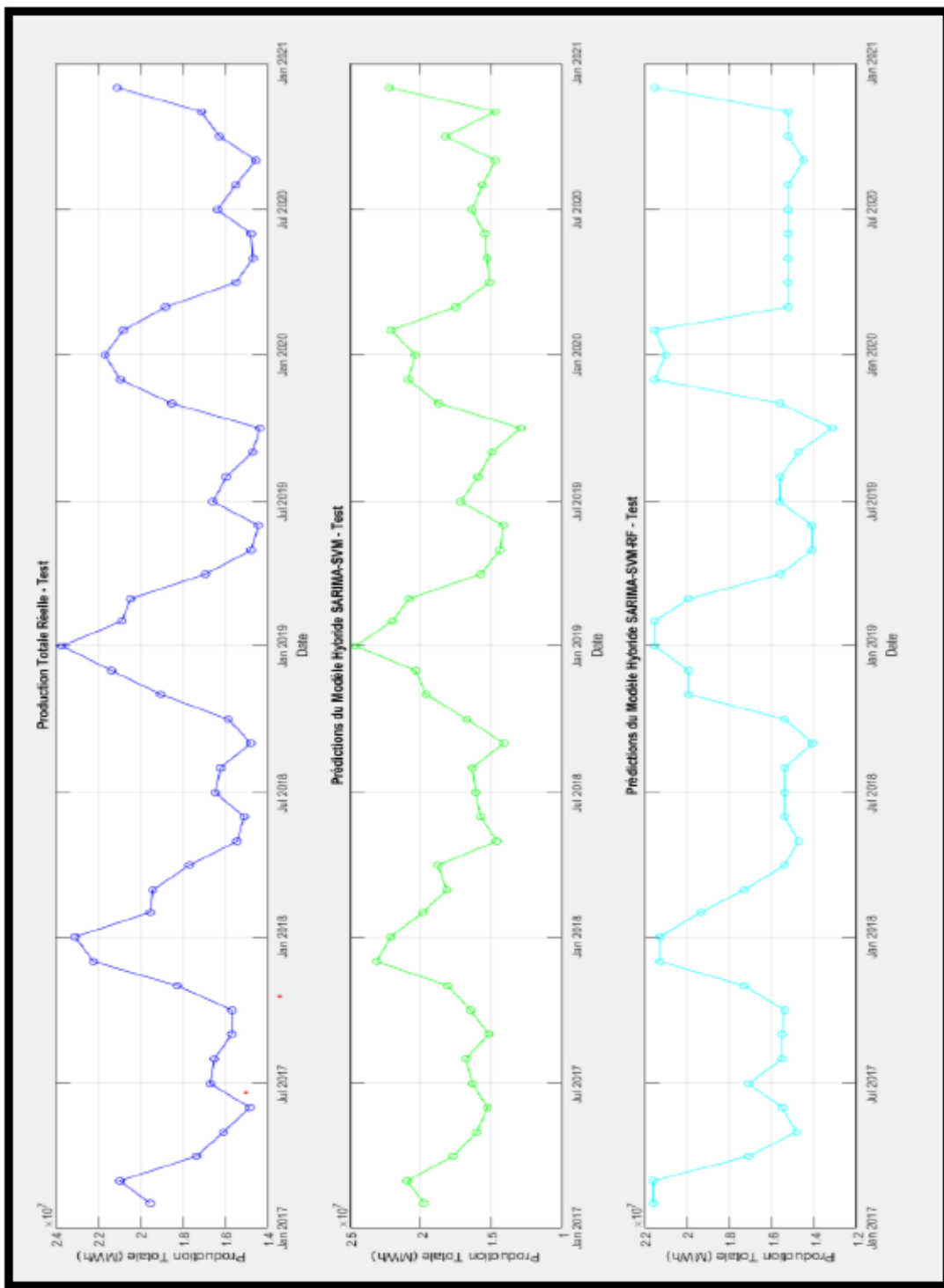


Figure 54 Production Totale Réelle et Prédications – Test

Ce graphique (Figure 54) présente trois sous-graphes montrant la production totale réelle et les prédictions des modèles hybrides SARIMA-SVM et SARIMA-SVM-RF pour la période de test (2017-2020).

➤ Production Totale Réelle - Test :

Les données montrent des fluctuations saisonnières et des pics distincts, notamment autour de janvier 2018 et janvier 2019.

➤ Prédiction du Modèle Hybride SARIMA-SVM - Test :

Les prédictions du modèle SARIMA-SVM suivent globalement la tendance des données réelles. Ce modèle est efficace pour capturer les motifs saisonniers récurrents, ce qui est essentiel pour les prévisions énergétiques à long terme.

➤ Prédiction du Modèle Hybride SARIMA-SVM-RF – Test :

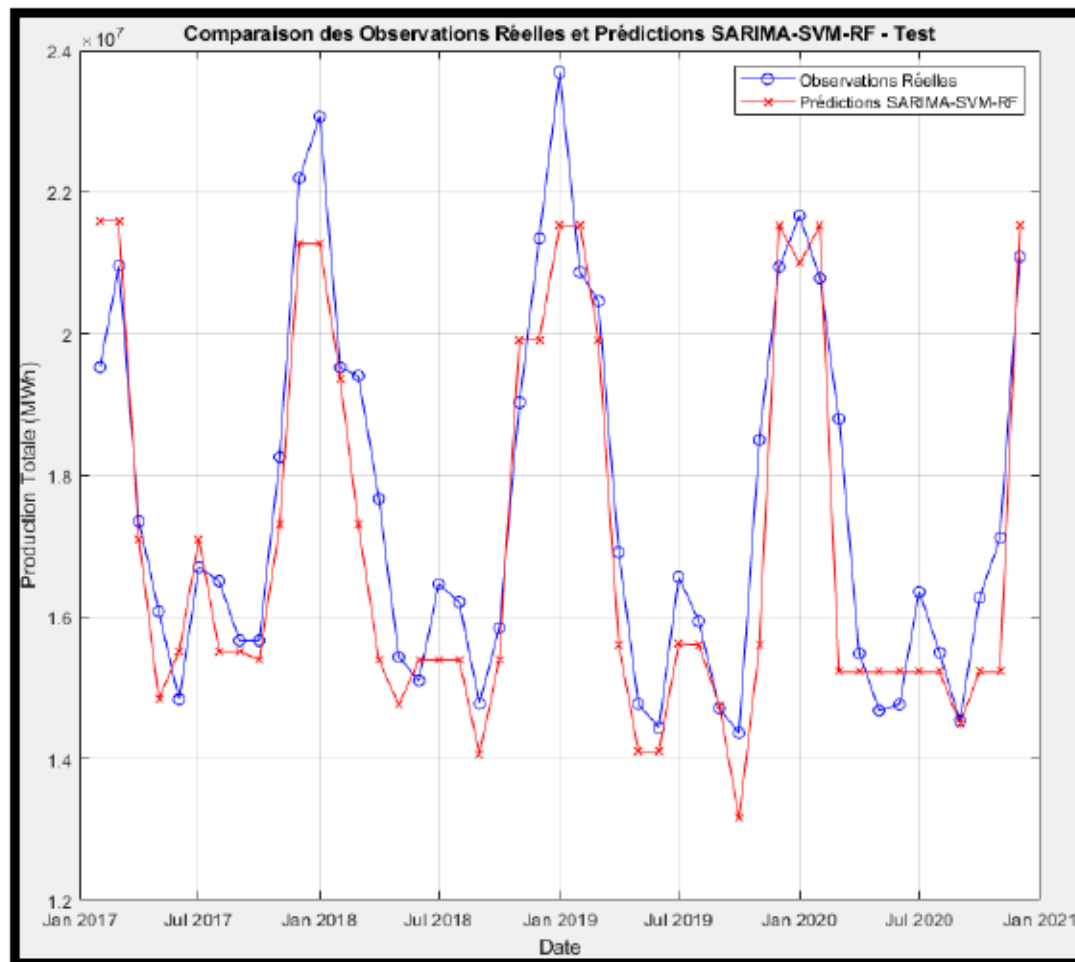


Figure 55 Comparaison des Observations Réelles et Prédictions SARIMA-SVM-RF - Test

Les prédictions du modèle SARIMA-SVM-RF (Figure 55) suivent de manière plus fidèle les variations des données réelles, offrant une meilleure précision par rapport au modèle SARIMA-SVM seul. Cela se manifeste par une réduction significative des divergences lors des pics et des creux extrêmes.

Par exemple, les pics autour de janvier 2018 et janvier 2019 sont mieux capturés, les valeurs prédictives s'approchant plus étroitement des valeurs observées.

Les divergences sont réduites, particulièrement lors des variations extrêmes, grâce à l'intégration du modèle RF. Ce modèle capture efficacement les résidus laissés par le

SVM, comblant les lacunes du modèle précédent. Le RF, en tant que modèle non paramétrique puissant, excelle à modéliser les relations complexes et non linéaires dans les données résiduelles, améliorant ainsi la précision des prédictions.

Le modèle SARIMA-SVM-RF démontre une robustesse accrue face aux variations saisonnières et aux anomalies dans les données. Cette robustesse se traduit par une fiabilité supérieure des prédictions, même dans des conditions de test rigoureuses.

Grâce à la combinaison des modèles SARIMA, SVM et RF, la Fusion Tri-Modèle présente une capacité supérieure à généraliser à de nouvelles données, réduisant ainsi le risque de surajustement et améliorant les performances sur des jeux de données non vus.

L'erreur quadratique moyenne (RMSE) est significativement réduite avec le modèle SARIMA-SVM-RF par rapport au SARIMA-SVM, indiquant une meilleure performance globale.

Conclusion:

Ce chapitre a démontré comment la combinaison de plusieurs approches prédictives améliore la précision et la robustesse des prévisions énergétiques. À partir des modèles linéaires de base, comme SARIMA, qui se concentrent sur les tendances et les saisons, nous avons évolué vers des méthodes plus avancées, capables de capturer les non-linéarités complexes et les variations irrégulières dans les données.

L'introduction de l'apprentissage machine, avec des techniques comme les Machines à Vecteurs de Support (SVM) et la Forêt Aléatoire (Random Forest - RF), a été essentielle pour surmonter les limites des modèles linéaires. En combinant ces méthodes dans un cadre hybride, le modèle SARIMA-SVM a permis de mieux ajuster les prédictions en traitant les erreurs résiduelles laissées par SARIMA.

Toutefois, c'est la **Fusion Tri-Modèle (SARIMA-SVM-RF)** qui a prouvé être la plus performante, en tirant parti des forces des trois modèles. Cette fusion permet de capturer les tendances linéaires et saisonnières (SARIMA), de gérer les non-linéarités (SVM), et de renforcer la stabilité et la précision des prévisions (RF).

Cette approche intégrée et hybride offre une solution prédictive robuste, nécessaire pour une gestion énergétique efficace. La **synergie des modèles** garantit une meilleure anticipation des fluctuations énergétiques, contribuant ainsi à une planification optimale et à des décisions éclairées dans des contextes aussi complexes que la gestion de l'énergie nationale.

Ainsi, la mise en œuvre de ces techniques hybrides, et notamment la **Fusion Tri-Modèle**, constitue un pas en avant vers une **prédiction fiable et adaptable**, particulièrement dans des environnements dynamiques où la précision des prévisions est cruciale.

Conclusion générale:

Ce projet met en lumière l'importance d'intégrer des approches avancées en modélisation prédictive pour analyser des séries chronologiques et des données complexes. L'objectif du projet était d'améliorer la précision des prévisions en appliquant différentes techniques de modélisation, allant des modèles de base (comme ARIMA) à des modèles hybrides et à la fusion tri-modèle, une approche combinant plusieurs algorithmes pour maximiser les performances prédictives.

Les modèles de base, comme ARIMA, ont prouvé leur efficacité dans l'analyse de séries chronologiques simples, notamment en capturant des relations linéaires et des tendances saisonnières. Cependant, ils montrent des limites lorsqu'il s'agit de traiter des données complexes ou des phénomènes non linéaires. C'est ici qu'interviennent les modèles hybrides, qui combinent des modèles traditionnels et des techniques d'apprentissage automatique, comme le SARIMA-SVR. Cette approche permet de tirer parti des forces de chaque méthode : SARIMA pour modéliser les tendances et la saisonnalité, et SVR (Support Vector Regression) pour capturer des relations non linéaires, offrant ainsi une meilleure flexibilité et précision.

Le projet s'est également concentré sur la fusion tri-modèle, qui est une technique plus avancée et prometteuse. Cette approche combine les prédictions de trois modèles — SARIMA, SVM, et Random Forest —, chacun apportant des compétences spécifiques à l'analyse des données. La fusion de ces trois modèles permet non seulement d'améliorer la précision, mais aussi de créer une solution robuste capable de gérer des scénarios complexes où les modèles individuels pourraient échouer. Cette approche s'avère particulièrement utile dans le cadre d'analyses prédictives impliquant de grandes quantités de données volatiles ou incertaines.

Un autre point clé du projet a été la préparation des données, un élément fondamental pour garantir la qualité des prévisions. La gestion des données manquantes, la

normalisation des données, et l'analyse des tendances saisonnières et des cycles ont permis de s'assurer que les modèles utilisés étaient alimentés par des données fiables et bien structurées. Cela a eu un impact direct sur la précision des prédictions et sur la capacité des modèles à généraliser efficacement.

En résumé, ce projet a montré que l'utilisation de techniques avancées de modélisation prédictive, en particulier les modèles hybrides et la fusion tri-modèle, est essentielle pour améliorer la précision des prévisions dans des environnements de données complexes. En combinant des méthodes linéaires et non linéaires, en optimisant la préparation des données, et en tirant parti de la complémentarité des modèles, ce projet a réussi à proposer des solutions prédictives robustes et fiables, adaptées aux défis actuels en matière de gestion des données et de prise de décision stratégique.

Bibliographie:

- [1] IBM, "Évaluation de données avec des modèles de prévision," IBM SPSS Statistics, 2024. <https://www.ibm.com/docs/fr/spss-statistics/saas?topic=system-scoring-data-predictive-models>.
- [2] Université Laval, "MQT-2101 Analyse et modélisation des données," <https://www.ulaval.ca/>.
- [3] Le Digitaliseur, "Analyse prédictive: Définition et exemples concrets," 2024 <https://ledigitaliseur.fr/analyse-predictive-definition-exemples-concrets>.
- [4] Aquila Data, "Mieux comprendre les méthodes de prévision des séries chronologiques," 2024. <https://www.aquiladata.fr>
- [5] GRETTIA, "Approches hybrides Modèles physiques/simulation et modèles données," 2024 <https://grettia.univ-gustave-eiffel.fr>.
- [6] Ipsos, "METHODE > Hybridation et fusion de données," 2011. <https://www.ipsos.com>.
- [7] Claranet, "Data Modeling : modèles de données," 2024. <https://www.claranet.com>.
- [8] A. Montanari, R. Rosso, and M. S. Taqqu, "A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan," *Water Resour.*, 2000.
- [9] P. F. Pai and C. S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, 2005.
- [10] Sameer Al-Dahidi, Hani Muhsen, Ma'en S Sari, Mohammad Alrbai, "Very short-term solar PV generation forecast using SARIMA model: A case study," in *Proc. 7th Int. Conf. Power Syst. (ICPS)*, 2017
- [11] S. Muzaffar and A. Afshari, "Short-term load forecasts using LSTM networks," *Energy Procedia*, vol. 158, 2019.
- [12] <https://www.cer-rec.gc.ca/en/data-analysis/energy-markets/provincial-territorial-energy-profiles/provincial-territorial-energy-profiles-canada.html>
- [13] P. F. Pai and C. S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, 2005.

- [14] W. S. Jevons, "The Coal Question: An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal-Mines," 3rd ed., Macmillan and Co., 1906.
- [15] K. Pearson, "Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia," *Philos. Trans. R. Soc. Lond. A*, 1896.
- [16] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, 1936.
- [17] A. N. Kolmogorov, "Foundations of the Theory of Probability," 2nd English ed., Chelsea Publishing Company, 1956.
- [18] N. Wiener, "Differential space," *J. Math. Phys.*, 1923.
- [19] G. E. P. Box and G. M. Jenkins, "Time Series Analysis: Forecasting and Control," Holden-Day, 1970.
- [20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, 2015.
- [21] H. T. Nguyen, M. N. Le, and N. T. Pham, "A hybrid ARIMA and artificial neural network model for time series forecasting," *Procedia Comput. Sci.* 2021.
- [22] V. Vapnik, "The Nature of Statistical Learning Theory," Springer, 1995.
- [23] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, 1992
- [24] L. Breiman, "Random Forests," *Mach. Learn.* 2001.
- [25] A. Liaw and M. Wiener, "Classification and Regression by randomForest," *R News*, 2002.
- [26] <https://www.cer-rec.gc.ca/fr/donnees-analyse/produits-base-energetiques/electricite/rapport/electricite-renouvelable-canada/provinces/electricite-renouvelable-canada-canada.html>
- [27] Centre canadien d'information sur l'énergie, "Production d'énergie électrique, réceptions, livraisons et disponibilité," *Statistique Canada*, 2024. <https://www150.statcan.gc.ca/t1/tb11/fr/tv.action?pid=2510001601>
- [28] J. Hamilton, "Time Series Analysis," Princeton University Press, 1994.
- [29] R. Hodrick and E. Prescott, "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking*, vol. 29, no. 1, pp. 1-16, 1997.

- [30] Soliman, M., Fatnassi, T., Elgammal, I. and Figueiredo, R., 2023. Exploring the Major Trends and Emerging Themes of Artificial Intelligence in the Scientific Leading Journals amidst the COVID-19 Era. *Big Data and Cognitive Computing*, 7(1), p.12.
- [31] Business Council of Canada, "Canada in the New Energy Landscape," Business Council of Canada, 2023. <http://thebusinesscouncil.ca>.
- [32] Canada Energy Regulator. (2023). "Canada Energy Regulator projects a major transformation to Canada's energy system in a net-zero world." Canada Energy Regulator. Retrieved from www.cer-rec.gc.ca.
- [33] Statistics Canada, "Energy statistics, March 2023," Statistics Canada, 2023. <https://www150.statcan.gc.ca>.
- [34] Natural Resources Canada, "Government of Canada Invests in Energy-Modelling Tools to Support Emissions Reductions," Natural Resources Canada, June 13, 2023. <https://www.canada.ca/en/natural-resources-canada/news/2023/06/government-of-canada-invests-in-energy-modelling-tools-to-support-emissions-reductions.html>.
- [35] Canada Energy Regulator, "Canada's Energy Future 2023: Energy Supply and Demand Projections to 2050," Canada Energy Regulator, 2023. <https://www.cer-rec.gc.ca>.
- [36] "Deep learning integration optimization of electric energy load forecasting and market price based on the ANN-LSTM-transformer method," *Frontiers in Artificial Intelligence*, 2023. Part 1. <https://www.frontiersin.org>.
- [37] "Why AI and energy are the new power couple," International Energy Agency (IEA), November 2, 2023. <https://www.iea.org/commentaries/why-ai-and-energy-are-the-new-power-couple>.
- [38] "Comparison of machine learning and statistical methods in the field of renewable energy power generation forecasting: a mini review," *Frontiers in Energy Research*, 2023. <https://www.frontiersin.org>.
- [39] "Deep learning integration optimization of electric energy load forecasting and market price based on the ANN-LSTM-transformer method," *Frontiers in Artificial Intelligence*, 2023. Part 2 <https://www.frontiersin.org>.
- [40] Y. Dou, S. Tan, and D. Xie, "Comparison of machine learning and statistical methods in the field of renewable energy power generation forecasting: a mini review," *Front. Energy Res.*, vol. 11, Jul. 26, 2023.
- [41] D. McLaughlin and B. Page, "The Canada-US trade and energy relationship," *Policy Options*, June 1, 2010. <https://policyoptions>.

- [42] P. Staudt et al., "Forecasting cross-border power transmission capacities in Central Western Europe using artificial neural networks," *Energy Informatics*, 2023. <https://energyinformatics.springeropen.com/articles/10.1186/s42162-023-00178-6>
- [43] "Hydroelectricity in Canada," *The Canadian Encyclopedia*, <https://www.thecanadianencyclopedia.ca/en/article/hydroelectricity> .
- [44] "Canada's Hydroelectric Potential," *Canadian Geographic*, May 24, 2016. <https://canadiangeographic.ca/articles/canadas-hydroelectric-potential/> .
- [45] "Canada's Energy Future 2023: Energy Supply and Demand Projections to 2050," *Canada Energy Regulator*, <https://www.cer-rec.gc.ca/en/data-analysis/canada-energy-future/2023/> .
- [46] IEA, "Canada 2022," *International Energy Agency*, Paris, 2022. <https://www.iea.org/reports/canada-2022>.
- [47] "Canada Energy Overview," *Energy Council of Canada*, <https://energy.ca/canada-energy-overview/>.
- [48] Priesmann, J., Nolting, L., Kockel, C. et al. (2021). Time series of useful energy consumption patterns for energy system modeling. *Scientific Data*, 8(1), 148. <https://doi.org/10.1038/s41597-021-00907-w>
- [49] *Canada Energy Regulator*. (2021). *Canada's Energy Future 2021 - Key Findings*. <https://www.cer-rec.gc.ca/en/data-analysis/canada-energy-future/2021/key-findings.html>.
- [50] IEA. (Year). "Canada Electricity Security Policy – Analysis." *International Energy Agency*. <https://www.iea.org/articles/canada-electricity-security-policy-2>
- [51] Vine, D. (May 2017). *Interconnected: Canadian and U.S. Electricity*. <https://www.c2es.org/document/interconnected-canadian-and-u-s-electricity/>
- [52] *Government of Canada*. (January 2022). "Canada's bold policies and support for innovation can underpin a successful energy transition, says new IEA policy review." <https://www.canada.ca/en/natural-resources-canada/news/2022/01/canadas-bold-policies-and-support-for-innovation-can-underpin-a-successful-energy-transition-says-new-iea-policy-review.html>
- [53] *Center for Strategic & International Studies*. "Mapping the U.S.-Canada Energy Relationship." <https://www.csis.org/analysis/mapping-us-canada-energy-relationship>
- [54] *Global Energy Institute*. "NAFTA and North American Energy." <https://www.globalenergyinstitute.org/nafta-and-north-american-energy>.

- [55] Jordan B. Kinder (2024) Communicating the “Social Meaning[s] of Shale”, *Environmental Communication*, DOI: 10.1080/17524032.2024.2329654
- [56] Yanez, J., Fuenmayor, R., Srivastava, P., Sadek, N., & Vivas, P. (2024, March). Applying Digital Rapid Response Ranking and Event Detection System in Oriente Basin. Paper presented at the SPE Canadian Energy Technology Conference and Exhibition, Calgary, Alberta, Canada. <https://onepetro.org/specet/proceedings-abstract/24CET/2-24CET/543208>
- [57] Daneshvarjejad, N., Khodaparast, P., Cho, Y., et al. (2024). A Method for Detection of Methane Leaks from Idle and Orphaned Wells Using High-Precision Sensor and a Ventilation Canopy. SPE Canadian Energy Technology Conference. [onepetro.org. https://onepetro.org/specet/proceedings-abstract/24CET/1-24CET/543170](https://onepetro.org/specet/proceedings-abstract/24CET/1-24CET/543170)
- [58] Mwakipunda, G.C., Mgimba, M.M., Ngata, M.R., et al. (2024). A critical review on compressed air energy storage in underground geological media: Advances and future outlook. *Gas Science and Technology*. Elsevier. <https://www.sciencedirect.com/science/article/pii/S2949908924000591>