

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMPARAISON D'ALGORITHMES DE DÉTECTION D'ANOMALIES
DE CONSOMMATION ÉLECTRIQUE EN CONDITION NORDIQUE

MÉMOIRE PRÉSENTÉ
COMME EXIGENCE PARTIELLE DE LA
MAÎTRISE EN GÉNIE ÉLECTRIQUE

PAR
MOHAMED AMINE KAIBICHE

AOÛT 2026

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire, de cette thèse ou de cet essai a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire, de sa thèse ou de son essai.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire, cette thèse ou cet essai. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire, de cette thèse et de son essai requiert son autorisation.

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES
MAÎTRISE EN GÉNIE ÉLECTRIQUE (M. Sc. A.)

Direction de recherche :

Pr. Kodjo AGBOSSOU

Directeur de recherche

Dr. Nilson HENAO

Codirecteur de recherche

Jury d'évaluation :

Pr. Ahmed CHÉRITI

Évaluateur interne

Pr. Mohsen GHRIBI

Évaluateur externe

Dr. Nilson HENAO

Codirecteur de recherche

Résumé

Le secteur énergétique occupe une place centrale dans le développement des sociétés modernes. Toutefois, il est aujourd'hui confronté à une hausse rapide de la demande, ce qui accentue l'importance stratégique de l'efficacité énergétique. Dans les climats nordiques comme celui du Québec, cet enjeu est particulièrement critique puisque les conditions météorologiques hivernales influencent fortement la consommation électrique, créant des périodes de pointe contraignantes pour le réseau. Dans ce contexte, la détection d'anomalies de consommation représente un levier essentiel pour repérer les usages inefficaces, informer les utilisateurs finaux de leurs écarts et favoriser des comportements plus sobres. Cependant, cette détection demeure un processus complexe en raison de la forte variabilité des données, induite par des comportements aléatoires, et de la difficulté à définir des seuils de décision assurant à la fois fiabilité et rapidité.

Pour répondre à cette problématique, ce mémoire propose une étude comparative de plusieurs algorithmes de détection d'anomalies de consommation électrique en milieu résidentiel. L'objectif est d'identifier les avantages, les limites et les enjeux d'implémentation de ces méthodes, en considérant spécifiquement l'impact des conditions météorologiques et la nature des inefficacités énergétiques. L'approche développée repose sur des techniques d'apprentissage non supervisé exploitant l'historique de consommation, couplées à l'analyse analytique des résidus de prédiction pour caractériser les déviations par rapport à un profil de référence. La méthodologie est validée sur des données réelles provenant d'une résidence instrumentée à Trois-Rivières (Québec). Afin de reproduire des scénarios de surconsomma-

tion contrôlés et réalistes, des anomalies artificielles de type décalage positif de niveau est volontairement injectées.

Les performances des approches retenues sont ensuite évaluées et comparées à l'aide de métriques portant sur la fiabilité de la détection, la robustesse du paramétrage et le coût computationnel. Dans le scénario expérimental considéré, les résultats mettent en évidence la capacité des systèmes étudiés à détecter l'anomalie injectée, tout en faisant ressortir les limites propres à chaque algorithme. Cette étude souligne ainsi l'importance d'adapter le choix de la méthode au type d'anomalie ciblé, aux caractéristiques des données et aux contraintes opérationnelles, afin d'améliorer la précision de la détection et la pertinence des notifications transmises aux consommateurs.

Avant-propos

Je tiens tout d'abord à exprimer ma profonde gratitude à mon directeur de recherche, Professeur Kodjo Agbossou, pour la qualité de son encadrement, son soutien constant et la confiance qu'il m'a accordée tout au long de ce travail. J'adresse également mes sincères remerciements à mon codirecteur de recherche, Monsieur Nilson Henao, pour son accompagnement, ses conseils scientifiques et son appui. Je remercie aussi Monsieur Juan Dominguez pour son aide, sa disponibilité et ses précieuses contributions. Ce travail n'aurait pas été possible sans l'appui de différents instituts et organismes qui, à travers leur soutien financier, ont reconnu l'importance de ce projet. Je tiens ainsi à remercier le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), l'Institut de recherche sur l'hydrogène (IRH) ainsi que la Fondation de l'Université du Québec à Trois-Rivières pour leur soutien financier et institutionnel.

Je voudrais également exprimer toute ma reconnaissance à ma mère pour son amour et son soutien moral. Je remercie profondément ma conjointe, Asmaa, pour sa patience, sa compréhension, son soutien quotidien et ses encouragements constants. J'adresse également une pensée affectueuse à mes deux enfants, Yasser et Elina, qui ont été pour moi une grande source d'énergie, d'espoir et de motivation tout au long de cette aventure.

Je tiens enfin à remercier mes amis Moussa Moussatta, Hani Mehaba, Belkacem Bekkour, Abdelhadi Hadj-Djilani et Amine Saci pour leur soutien, leurs encouragements et leur amitié sincère pendant ces années d'études.

Table des matières

Résumé	iii
Avant-propos	v
Table des matières	vi
Liste des tableaux	x
Liste des figures	xi
Liste des symboles	xii
Liste des acronymes	xv
Chapitre 1 – Introduction.	1
1.1 Contexte et problématique	1
1.2 Problématique spécifique.	5
1.3 Objectifs de recherche	7
1.4 Contributions	8
1.5 Structure du mémoire	9
Chapitre 2 – État de l’art	10
2.1 Performance énergétique et comportement thermique.	10
2.1.1 Facteurs d’influence	11
2.1.2 Anomalies thermiques	13

2.2	Détection d'anomalies	15
2.2.1	Paradigmes d'apprentissage	15
2.2.2	Algorithmes de détection	17
2.2.3	Modélisation statistique et paramétrisation	26
2.2.4	Outils logiciels et frameworks Open Source	29
2.2.5	Analyse des lacunes	33
Chapitre 3 – Méthodologie		37
3.1	Architecture du système	37
3.1.1	Acquisition des données	40
3.2	Acquisition et traitement des données	43
3.2.1	Intégration et alignement temporel	43
3.2.2	Nettoyage des données.	45
3.3	Modélisation du comportement normal.	47
3.3.1	Sélection des modèles	47
3.3.2	Extraction de caractéristiques	50
3.4	Identification des anomalies	52
3.4.1	Calcul des résidus	52
3.4.2	Analyse statistique.	54
3.4.3	Règles de détection et paramétrisation	57

Chapitre 4 – Évaluation et Résultats	61
4.1 Introduction	61
4.1.1 Définition des scénarios et vérité terrain	62
4.2 Cas d'étude et implémentation expérimentale.	64
4.2.1 Environnement de test et profil du bâtiment	64
4.2.2 Environnement logiciel et ressources Open Source	64
4.3 Scénarios de test	65
4.3.1 Anomalies réversibles	65
4.4 Entraînement des modèles	67
4.4.1 Support Vector Regression (SVR)	71
4.4.2 eXtreme Gradient Boosting (XGBoost)	72
4.4.3 Convolutional Neural Networks and Long Short-Term Memory (CNN-LSTM)	74
4.4.4 Time-series Dense Encoder (TiDE)	77
4.5 Prédiction du profil de référence	81
4.5.1 Root Mean Square Error (RMSE)	81
4.5.2 Mean Absolute Error (MAE).	82
4.5.3 Mean Absolute Percentage Error (MAPE)	82
4.5.4 Coefficient de détermination (R^2)	83
4.6 Résultats de détection	84
4.6.1 Analyse du compromis (Réactivité vs Faux positifs)	84

4.6.2	Évaluation comparative des algorithmes	92
4.7	Discussion	98
Chapitre 5 – Conclusion		102
5.1	Synthèse	102
5.2	Contributions principales.	105
5.3	Limites de l'étude	106
5.4	Travaux futurs	107
Références.		109

Liste des tableaux

Tableau 3-1	Variables temporelles cycliques construites par encodage trigonométrique	52
Tableau 4-1	Plages d'hyperparamètres considérées lors de l'optimisation des modèles	68
Tableau 4-2	Hyperparamètres du modèle SVR	72
Tableau 4-3	Hyperparamètres du modèle XGBoost	73
Tableau 4-4	Hyperparamètres du modèle CNN-LSTM	76
Tableau 4-5	Hyperparamètres du modèle TiDE	79
Tableau 4-6	Comparaison des performances des modèles de prévision	83
Tableau 4-7	Paramètres du banc d'essai expérimental et caractéristiques de l'anomalie injectée	88
Tableau 4-8	Synthèse comparative des modèles à leur configuration optimale (meilleur F1-score)	93

Liste des figures

Figure 1.1	Consommation des différentes formes d'énergie par secteur d'activité au Québec, 2020 [5].	2
Figure 1.2	Comparaison des besoins de puissance à la pointe et puissance installée au Québec (2015–2023) [12].	3
Figure 3.1	Architecture de la méthode de détection des anomalies.	39
Figure 3.2	Système de surveillance de charge en temps réel	41
Figure 3.3	Système de récupération de données météorologiques	43
Figure 3.4	Relation entre la puissance consommée et la température extérieure avec régression	46
Figure 3.5	Évolution de la puissance consommée et de la température extérieure après prétraitement	46
Figure 3.6	Organigramme du système de détection des anomalies	56
Figure 4.1	Exemple d'anomalie injectée de type décalage positif	63
Figure 4.2	Temps d'exécution moyen du modèle SVR	70
Figure 4.3	Architecture du modèle CNN-LSTM	74
Figure 4.4	Courbe de convergence du modèle CNN-LSTM	77
Figure 4.5	Architecture du modèle TiDE	78
Figure 4.6	Courbe de convergence du modèle TiDE	80
Figure 4.7	Évolution du nombre de faux positifs en fonction des seuils $Band_R$ et S_q	89
Figure 4.8	Évolution du rappel et du F1-score en fonction du seuil $Band_R$	90
Figure 4.9	Compromis précision-délai en fonction du seuil S_q	91
Figure 4.10	Synthèse comparative des performances optimales des quatre algorithmes	94
Figure 4.11	Profil multi-métriques des quatre algorithmes au seuil optimal global	97

Liste des symboles

Δt	Pas d'échantillonnage réel des données, exprimé en minutes
$\hat{y}_t$	Consommation énergétique prédite ou attendue par le modèle à l'instant t
$\mathbf{X}_t$	Vecteur des variables explicatives à l'instant t
$\mathcal{N}(\mu_t, \sigma_t)$	Loi normale conditionnelle caractérisée par une moyenne μ_t et un écart-type σ_t
μ_t	Moyenne conditionnelle estimée de y_t sachant $\mathbf{X}_t$
σ_t	Écart-type conditionnel estimé de y_t sachant $\mathbf{X}_t$
τ^*	Moment de détection, défini comme le premier instant où la surface cumulée franchit strictement S_q
$Band_R$	Seuil sur la distribution des résidus normaux, calibré au quantile $Band_R\%$
E	Énergie consommée
E_i	Énergie élémentaire associée à l'observation i , exprimée en kilowattheures
$E_k^{(15)}$	Énergie consommée sur le k -ième intervalle d'agrégation de 15 minutes
$E_k^{(T)}$	Énergie consommée sur le k -ième intervalle d'agrégation de durée T
e_t	Résidu de prédiction
$f(\mathbf{X}_t)$	Fonction apprise reliant les variables explicatives à la variable cible

i	Indice de l'observation ou de l'échantillon
I_k	Ensemble des échantillons appartenant au k -ième intervalle d'agrégation
k	Indice de l'intervalle d'agrégation
kWh	Kilowattheure, unité de mesure de l'énergie électrique
MW	Mégawatt, unité de mesure de la puissance électrique
N	Nombre de jours optimal
P	Puissance électrique
P_i	Puissance mesurée à l'instant i , exprimée en watts
r_t	Résidu de prédiction à l'instant t , défini par $r_t = y_t - \hat{y}_t$
$S(\tau)$	Surface cumulée depuis le début d'un épisode de dépassement jusqu'à l'instant τ
S_i	Surface de l'épisode de dépassement i , calculée par approximation triangulaire pour un pic isolé ou trapézoïdale pour une séquence consécutive
S_q	Quantile calculé sur la distribution des surfaces de dépassement
T	Durée de l'intervalle d'agrégation, exprimée en minutes
t	Instant temporel considéré
t_{end}	Instant de fin d'un épisode de dépassement
t_{start}	Instant du début d'un épisode de dépassement
TWh	Térawattheure , unité de mesure de la puissance électrique

W	Watt, unité de mesure de la puissance électrique
Wh	Wattheure, unité de mesure de l'énergie électrique
$y_t \mid \mathbf{X}_t$	Variable cible conditionnée par les variables explicatives
y_t	Consommation énergétique réellement observée à l'instant t
FN	Faux négatifs, nombre de points anormaux non détectés par le système
FP	Faux positifs, nombre de points normaux signalés à tort comme anormaux
TN	Vrais négatifs, nombre de points normaux correctement identifiés comme tels
TP	Vrais positifs : nombre de points anormaux correctement détectés
$^{\circ}\text{C}$	Degré Celsius

Liste des acronymes

R^2 Coefficient de détermination. viii, 83

AE Autoencodeur. 21

API Application Programming Interface. 32, 42

ARIMA Autoregressive Integrated Moving Average. 19

CAE Autoencodeur Convolutif. 20

CNN-LSTM Convolutional Neural Networks and Long Short-Term Memory. viii, 21, 74

DeepOD Python Deep Outlier Anomaly Detection. 33

DQR Deep Quantile Regression. 20

EM Expectation-Maximization. 18

EnbPI Ensemble batch Prediction Intervals. 24

GBM Gradient Boosting Models. 20

GMM Gaussian Mixture Model. 18

GNN Graph Neural Networks. 21

IoT Internet of Things. 4

IQR Interquartile Range. 22

JSON JavaScript Object Notation. 42

LODA Lightweight On-line Detector of Anomalies. 23

LOF Local Outlier Factor. 22–24

LPBNN Load Prediction with Neural Network. 24, 25

LPBSVR Load Prediction with Regression. 24, 25

LSTM Long Short-Term Memory. 20, 21

MAE Mean Absolute Error. viii, 82

MAPE Mean Absolute Percentage Error. viii, 82

MCC Matthews Correlation Coefficient. 23, 25

MCFS Multi-Cluster Feature Selection. 18

MIT Massachusetts Institute of Technology. 40

ML Machine Learning. 15

mZ-score modified Z-score. 22

One-Class SVM One Class Support Vector Machines. 18

PyOD Python Outlier Detection. 33

RMSE Root Mean Square Error. viii, 81

SARIMAX Seasonal AutoRegressive Integrated Moving Average with eXogenous variables.

19

SVR Support Vector Regression. viii, 20, 47, 71

TES Time-to-Equal-Severity. 38, 39, 59, 87

TiDE Time-series Dense Encoder. viii, 48, 77

TODS Automated Time-series Outlier Detection System. 32

TWF Temporal width future. 78

TWP Temporal width past. 78

URL Uniform Resource Locator. 42

VAE Variationnel Autoencodeur. 21

XGBoost eXtreme Gradient Boosting. viii, 47, 72

XGBoostLSS Extension of XGBoost to probabilistic modelling and prediction. 24

Chapitre 1 – Introduction

1.1 Contexte et problématique

Le secteur énergétique joue un rôle fondamental dans le fonctionnement de nos sociétés modernes. Pilier du développement économique, social et technologique, il est confronté à des défis majeurs au cours des dernières décennies, notamment le réchauffement climatique, auquel il contribue de manière significative en raison des émissions de gaz à effet de serre [1]. Parallèlement, il doit répondre à une demande énergétique mondiale en forte croissance. Ainsi, en 2022, la consommation mondiale d'électricité a atteint plus de 25 000 TWh¹, portée notamment par l'urbanisation rapide et l'évolution des modes de vie vers un confort accru, impliquant l'usage de technologies de plus en plus énergivores [2]. Au Québec, la situation est également préoccupante, car la demande d'électricité continue d'augmenter. Selon le fournisseur d'électricité, une hausse de 12,5 % est prévue entre 2019 et 2029 [3]. Dans ce contexte, il devient impératif de s'engager dans une transition énergétique, en favorisant des sources d'énergie plus vertes et propres, tout en mettant en œuvre des stratégies visant une utilisation plus efficace et responsable de l'énergie [4].

Face à ces enjeux, l'analyse de la répartition de la consommation d'énergie par secteur s'avère essentielle. Dans la province de Québec, le secteur du bâtiment joue un rôle prépondérant dans l'augmentation de la demande énergétique, représentant environ 35 % de la consommation totale d'énergie en 2020, soit le même pourcentage que le secteur industriel, tandis que le secteur résidentiel représente à lui seul 20 % de cette consommation globale [5].

1. Enerdata, *World Energy & Climate Statistics – Yearbook 2022*, <https://yearbook.enerdata.net/electricity/electricity-domestic-consumption-data.html>, consulté le 22 février 2026.

De plus, l'électricité constitue la principale source d'énergie dans ce secteur. La Figure 1.1 illustre la consommation des différentes formes d'énergie par secteur d'activité au Québec et révèle que l'électricité est la principale source d'énergie dans le secteur résidentiel, représentant 74 % de sa consommation.

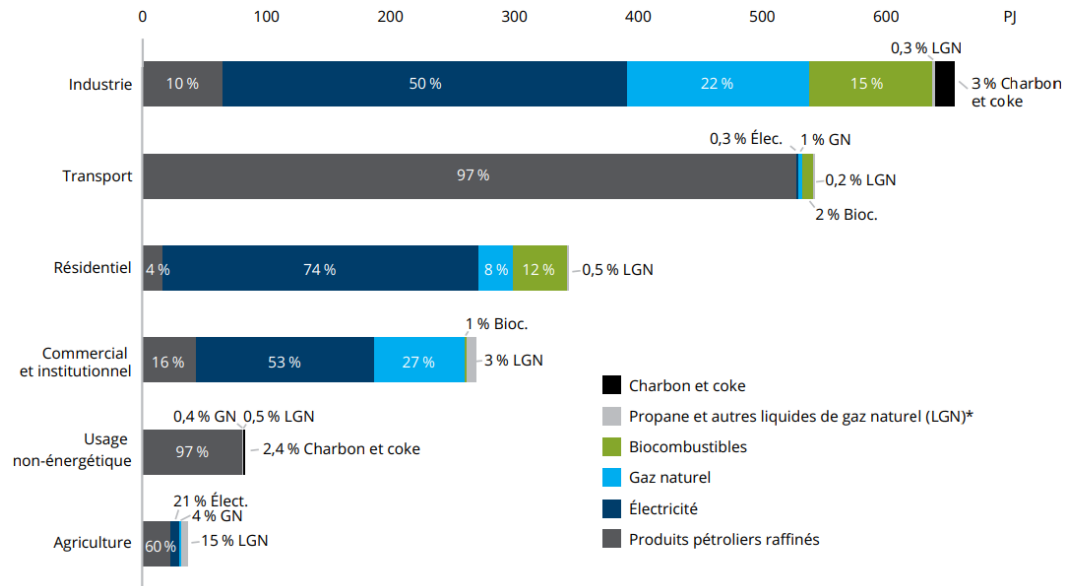


FIGURE 1.1 Consommation des différentes formes d'énergie par secteur d'activité au Québec, 2020 [5].

Au sein du secteur résidentiel, le chauffage des espaces représente la plus grande part de la consommation, avec plus de 54 % de la consommation totale d'électricité, tandis que le chauffe-eau représente jusqu'à 20 % [6]. Cette prédominance du chauffage s'explique par la position géographique du Québec, une région nordique caractérisée par des températures hivernales rigoureuses [7]. Les variations de température extérieure ont un impact direct sur la consommation d'énergie électrique, car les systèmes de chauffage doivent fonctionner à pleine capacité pour maintenir un confort thermique adéquat. Plus précisément, chaque fois que la température extérieure diminue de 1 °C (entre 10 °C et -20 °C), le distributeur d'électricité doit fournir près de 400 MW supplémentaires pour maintenir le confort thermique dans le secteur résidentiel [8]. Ce phénomène est d'autant plus marqué que les plinthes électriques

constituent le mode de chauffage le plus répandu au Québec, représentant 62 % des systèmes utilisés [9]. Bien que leur faible coût d'installation et leur simplicité d'utilisation expliquent cette prévalence, elles figurent parmi les solutions les plus énergivores, en particulier durant les mois d'hiver.

Cette forte dépendance au chauffage électrique place le réseau dans des situations critiques durant les périodes de pointe. La Figure 1.2, qui compare la puissance demandée à la puissance installée au Québec entre 2015 et 2023, montre que la puissance demandée a souvent égalé, voire dépassé, la puissance installée. Ces périodes critiques surviennent principalement en hiver, entre 6 h et 9 h ainsi qu'entre 16 h et 20 h [10], contraignant le fournisseur d'électricité à réaliser des investissements supplémentaires [11].

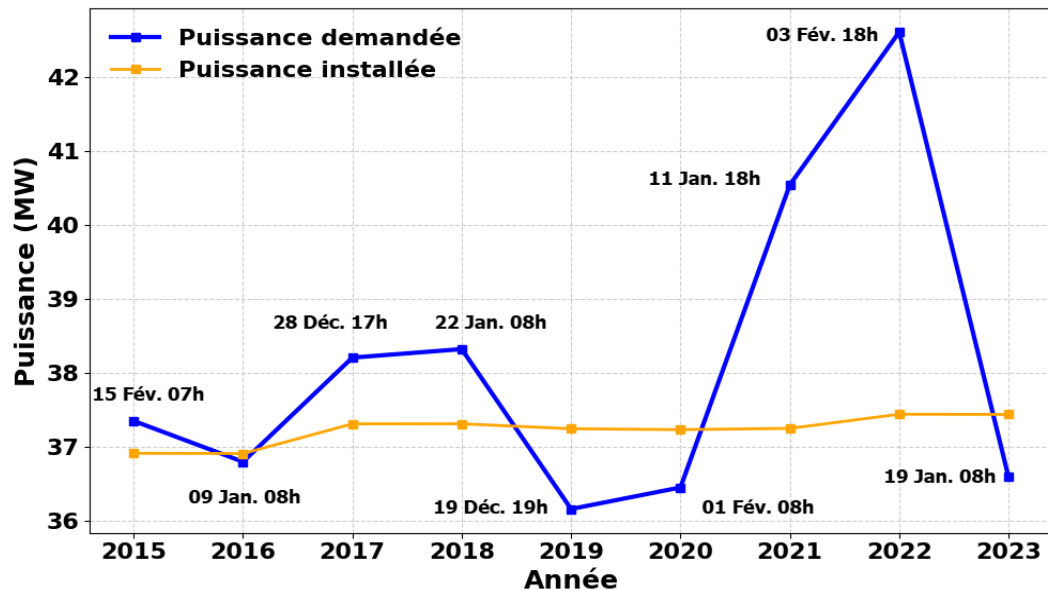


FIGURE 1.2 Comparaison des besoins de puissance à la pointe et puissance installée au Québec (2015–2023) [12].

Une étude estime que 6,5 milliards de dollars seraient nécessaires pour la construction de nouvelles centrales, uniquement pour faire face à des pics de consommation qui n'excèdent généralement pas 5 % de la journée [13]. Toutefois, cette solution coûteuse ne traite pas le problème de manière durable. Une approche plus pérenne consiste à agir sur la demande, en

développant des pratiques relevant de l'efficacité énergétique, afin de réduire les pointes de consommation tout en maintenant le niveau de confort.

Dans le secteur résidentiel, les marges de réduction demeurent importantes, d'après une étude, 30 % de l'énergie consommée est perdue [14], dont 10 % en raison de pertes thermiques liées à l'enveloppe du bâtiment. Face à ce constat, le Québec a mis en place des programmes de rénovation, tels que Rénoclimat, qui offre un accompagnement spécialisé et un soutien financier pour améliorer l'isolation thermique des résidences². D'autre part, les consommateurs seraient responsables de plus de 20 % de ces pertes [14], en raison de pratiques d'usage inefficaces. Des travaux de recherche montrent que l'implication active des occupants dans le suivi de leur consommation, notamment grâce à des retours d'information en temps réel, peut générer des bénéfices significatifs [15].

L'intégration de systèmes de surveillance dans la gestion énergétique des résidences permet d'établir un lien direct entre le fournisseur d'énergie et l'utilisateur [4]. L'adoption de technologies telles que l'internet des objets (Internet of Things (IoT) en anglais), concept introduit en 1999 [16], permet de connecter des capteurs intelligents qui collectent et partagent des données de consommation en temps réel via Internet [17]. L'acquisition de ces données permet d'élaborer des stratégies visant à réduire le gaspillage d'énergie, notamment par l'application d'algorithmes de détection d'anomalies capables d'identifier les comportements de consommation inhabituels ou inefficaces et d'en avertir les occupants en temps réel, leur permettant de prendre des mesures correctives immédiates. Néanmoins, l'efficacité des systèmes de détection dépend de plusieurs paramètres critiques, liés tant au contexte d'usage qu'au choix du modèle. Le développement d'un compromis bien adapté et structuré entre ces paramètres constitue un levier essentiel pour rendre ces systèmes plus performants et

2. Gouvernement du Québec, *Programme Rénoclimat*, <https://transitionenergetique.gouv.qc.ca/residentiel/programmes/renoclimat>, consulté le 22 septembre 2024.

opérationnellement viables.

1.2 Problématique spécifique

La détection d'anomalies de consommation électrique dans le secteur résidentiel, en particulier dans une optique d'efficacité énergétique, représente un défi majeur [18]. Elle vise à informer l'utilisateur final, de manière rapide et fiable, de comportements de consommation anormaux susceptibles d'engendrer du gaspillage. Néanmoins, en contexte nordique, les profils de charge résidentiels sont fortement influencés par les conditions climatiques, puisque les températures hivernales rigoureuses modifient considérablement les usages liés au chauffage et engendrent des variations importantes dans la demande électrique [19]. Cette dépendance climatique rend particulièrement ardue la distinction entre une hausse de consommation normale, induite par une chute de température extérieure, et une surconsommation anormale résultant de pratiques d'utilisation inefficaces de l'énergie par l'occupant.

En effet, l'un des obstacles centraux réside dans la difficulté à isoler, de manière spécifique et automatique, les déperditions thermiques provoquées par des comportements réversibles des occupants, tels que l'ouverture prolongée d'une fenêtre en période de chauffage, au sein du profil énergétique global d'une résidence [19]. Ces événements génèrent une surconsommation ponctuelle qui se superpose aux fluctuations normales du profil de charge, rendant leur repérage complexe sans information contextuelle complémentaire. Cette complexité est amplifiée par la rareté des données étiquetées en milieu résidentiel. Contrairement aux environnements industriels, les données de consommation des ménages sont rarement accompagnées d'annotations indiquant si un comportement observé est normal ou anormal, ce qui limite fortement l'applicabilité des approches d'apprentissage supervisé et impose le recours à des méthodes non supervisées [18].

Par ailleurs, bien que la littérature scientifique recense une grande variété d'algorithmes

de détection d'anomalies, aucun consensus ne se dégage quant aux méthodes non supervisées les plus efficaces pour distinguer les anomalies thermiques du comportement énergétique normal d'une résidence. Cette situation s'explique notamment par l'absence de bancs d'essai standardisés permettant de comparer rigoureusement les différentes approches [20]. Les hypothèses de modélisation, les conditions expérimentales et les métriques d'évaluation varient considérablement d'une étude à l'autre, rendant toute comparaison directe hasardeuse et le choix d'une méthode robuste particulièrement difficile. Au-delà du choix algorithmique, une difficulté opérationnelle majeure s'ajoute à ces constats, celle de la calibration des seuils de détection. Un système efficace doit être en mesure de signaler une anomalie suffisamment tôt pour permettre à l'occupant d'intervenir et de corriger le comportement en cause. Toutefois, une sensibilité trop élevée engendre un nombre excessif de fausses alertes, ce qui risque de décrédibiliser le système et de conduire l'utilisateur à ignorer les notifications. À l'inverse, un paramétrage trop conservateur retarde la détection et laisse persister le gaspillage énergétique. Trouver le juste équilibre entre rapidité de détection et taux de faux positifs acceptable constitue donc un enjeu central pour la viabilité de ces systèmes en contexte résidentiel [21].

Ce mémoire vise à répondre à cette problématique en proposant une étude comparative structurée, basée sur la modélisation du profil de consommation normal, puis sur l'analyse du comportement des algorithmes face à des profils anormaux. Plusieurs familles d'approches de détection d'anomalies sont ainsi évaluées afin d'identifier, pour chacune, les compromis et les limites, tant sur le plan de la fiabilité de détection que sur le plan du coût computationnel, afin de mettre en évidence les facteurs et paramètres qui conditionnent la rapidité et la robustesse des systèmes d'alerte, en vue de recommander des méthodes mieux adaptées à une utilisation opérationnelle en contexte résidentiel [20].

1.3 Objectifs de recherche

L'objectif principal de ce mémoire est de concevoir et d'évaluer un système d'apprentissage automatique non supervisé capable de détecter de manière autonome et fiable les anomalies thermiques réversibles dans le chauffage résidentiel, dans une optique d'efficacité énergétique. Cette étude s'appuie sur une base de données réelle collectée dans une résidence située à Trois-Rivières, au Québec [22]. Elle repose sur une évaluation systématique et détaillée de plusieurs approches issues de familles d'algorithmes variées, tout en tenant compte des particularités climatiques et contextuelles propres aux régions nordiques, ainsi que de la typologie spécifique des anomalies observées dans ce domaine. Plutôt que de chercher à identifier une méthode universellement optimale, ce travail propose une démarche unifiée visant à analyser les forces et les limites de chaque technique, en fonction des caractéristiques des données disponibles, des types d'anomalies ciblées et des exigences associées aux modèles envisagés. Il s'agit ainsi de formuler des recommandations pratiques et contextualisées pour orienter le choix des méthodes les plus pertinentes, tout en contribuant à une meilleure compréhension des enjeux liés à la détection d'anomalies énergétiques dans le secteur résidentiel. Afin de réaliser cet objectif général, deux objectifs spécifiques ont été définis :

- Établir un cadre technique comparatif rigoureux permettant d'identifier les algorithmes d'apprentissage non supervisé les mieux adaptés à la détection des anomalies thermiques réversibles en contexte résidentiel nordique
- Optimiser la paramétrisation des méthodes retenues afin de garantir le meilleur compromis possible entre la rapidité de détection et la minimisation du taux de fausses alertes.

1.4 Contributions

Les travaux menés dans le cadre de ce mémoire apportent plusieurs contributions à l'avancement des connaissances dans le domaine de la détection d'anomalies thermiques en milieu résidentiel. En s'appuyant sur des données réelles collectées en contexte nordique, cette étude propose des apports méthodologiques et pratiques qui se déclinent comme suit :

1. La mise en place d'un cadre comparatif structuré permettant d'évaluer, dans des conditions expérimentales cohérentes, les performances de plusieurs familles d'algorithmes d'apprentissage automatique non supervisé appliqués à la détection d'anomalies thermiques. Cet environnement de test unifié permet d'identifier, de manière rigoureuse, les méthodes les mieux adaptées aux spécificités du profil de consommation résidentiel en climat nordique.
2. Établissement précis des hyperparamètres et des seuils statistiques les plus appropriés pour le contexte résidentiel, en tenant compte des caractéristiques propres aux données de consommation et de la nature des anomalies ciblées. Cette étape est essentielle pour assurer la fiabilité et la reproductibilité des résultats obtenus.
3. Optimisation du compromis entre la réactivité de la détection et la minimisation du taux de fausses alertes. Ce travail cherche à atteindre le meilleur équilibre possible entre une détection suffisamment rapide pour permettre aux occupants d'entreprendre des actions correctives immédiates, et un taux de faux positifs suffisamment faible pour préserver la crédibilité du système auprès de l'utilisateur final.

L'ensemble de ces contributions est validé par des simulations numériques réalisées sur des données réelles de consommation résidentielle, mettant en évidence les gains obtenus en termes de rapidité et de fiabilité dans la détection des anomalies thermiques liées au chauffage

1.5 Structure du mémoire

Ce document est structuré en cinq chapitres, organisés de manière à présenter progressivement les étapes et les contributions de ce travail de recherche. Le premier chapitre introduit le sujet en exposant la problématique, les objectifs visés ainsi que les contributions attendues, afin de poser le cadre général de l'étude. Le deuxième chapitre dresse un état de l'art approfondi du domaine. Il examine d'abord les facteurs influençant la performance énergétique des bâtiments résidentiels et la nature des anomalies thermiques qui en découlent. Il analyse ensuite les solutions de télémétrie résidentielle, telles que le sous-mesurage et les thermostats intelligents, puis synthétise les approches algorithmiques et statistiques de détection d'anomalies, tout en dégagant les lacunes et les axes d'amélioration qui motivent la présente recherche. Le troisième chapitre détaille l'approche méthodologique adoptée pour répondre à la problématique identifiée. Il décrit l'architecture globale du système proposé, le traitement des données hétérogènes issues de sources multiples, la conception du cadre comparatif des modèles d'apprentissage non supervisé, ainsi que la définition des intervalles de confiance statistiques visant à optimiser la paramétrisation algorithmique. Le quatrième chapitre présente les résultats expérimentaux obtenus. À travers une étude de cas et divers scénarios de simulation, il évalue la capacité des modèles à identifier les anomalies thermiques réversibles. Il analyse également le compromis entre la réactivité de détection et l'occurrence de faux positifs, aboutissant à une évaluation comparative rigoureuse des algorithmes testés. Le cinquième chapitre conclut ce mémoire en synthétisant les principaux résultats obtenus au regard des objectifs initiaux. Il énonce les limites inhérentes à cette étude et propose des perspectives d'amélioration pour des travaux futurs, notamment en matière de contrôle intelligent et automatisé de la consommation énergétique résidentielle.

Chapitre 2 – État de l’art

Ce chapitre constitue le socle théorique de ce mémoire. Il examine les concepts fondamentaux et les éléments pertinents devant être pris en compte lors de la conception de systèmes de détection non supervisés appliqués au domaine résidentiel. La réflexion s’articule depuis les fondements algorithmiques et leur paramétrisation, jusqu’aux différents types d’anomalies thermiques rencontrées dans ce contexte. Une attention particulière est portée aux anomalies réversibles, dont la nature transitoire soulève des défis techniques distincts en matière de détection. Sur cette base, ce chapitre élabore des remarques importantes sur les lacunes critiques des études existantes, justifiant ainsi la conduite de ce travail.

2.1 Performance énergétique et comportement thermique

Les performances énergétiques d’un bâtiment résidentiel sont fortement influencées par l’interaction entre les variables climatiques et les caractéristiques de son enveloppe [23]. Les conditions météorologiques, en particulier la température extérieure, le vent et le rayonnement solaire, constituent des variables exogènes qui déterminent les échanges thermiques entre l’intérieur et l’extérieur du bâtiment [24]. Dans un contexte nordique comme celui du Québec, les écarts importants de température en période hivernale engendrent des pertes thermiques significatives, proportionnelles au gradient thermique entre l’air intérieur et extérieur [25]. Ces pertes sont directement modulées par les propriétés de l’enveloppe du bâtiment, constituée des parois opaques, des vitrages, de la toiture et des planchers. Son niveau de performance est généralement caractérisé par la résistance thermique, qui traduit la capacité des matériaux à s’opposer au flux de chaleur entre le bâtiment et son environnement extérieur

[26]. La résistance thermique équivalente d'un assemblage est calculée à partir des valeurs individuelles de chaque couche constitutive [27]. Toutefois, la résistance thermique seule ne suffit pas à décrire complètement le comportement de l'enveloppe. En effet, les éléments constructifs possèdent également une capacité thermique, qui traduit leur aptitude à stocker et à restituer l'énergie sous forme de chaleur [28]. Cette propriété, intimement liée à la densité et à la chaleur spécifique des matériaux, confère au bâtiment une inertie thermique qui atténue les fluctuations rapides de température intérieure en réponse aux sollicitations climatiques extérieures [29]. Par ailleurs, afin de quantifier l'influence du climat sur les besoins énergétiques, le concept de degrés-jours de chauffage constitue un indicateur clé largement utilisé. Il permet à la fois d'estimer les besoins en chauffage et de dimensionner les équipements énergétiques¹, tout en établissant des exigences réglementaires en matière d'isolation [30]. À ce titre, les zones climatiques définies par la norme ENERGY STAR® pour le Canada fournissent un cadre de référence pour spécifier les niveaux minimaux de performance de l'enveloppe². Il convient de souligner que la majorité des villes québécoises appartiennent à une zone climatique froide caractérisée par des besoins élevés en chauffage, ce qui reflète l'importance critique de l'isolation thermique dans ce contexte [31].

2.1.1 Facteurs d'influence

Les conditions météorologiques constituent la première catégorie de facteurs exogènes. La température extérieure est la principale variable climatique qui module les échanges thermiques entre l'enveloppe du bâtiment et son environnement [32]. En période hivernale, la diminution de la température extérieure accentue le gradient thermique, entraînant une augmentation de la demande en chauffage. Cette dépendance est particulièrement marquée dans les régions à climat rigoureux comme le Québec, où la composante météorologique occupe

1. Environnement et Changement climatique Canada, *Glossaire climatique*, <https://www.canada.ca/fr/environnement-changement-climatique.html>, consulté en septembre 2024.

2. Ressources naturelles Canada, <https://www.nrcan.gc.ca>, consulté en décembre 2025.

une part significative de la demande totale en raison des longues saisons froides [33]. Par ailleurs, les événements climatiques extrêmes et les effets de microclimat urbain peuvent induire des variations importantes des degrés-jours de chauffage, introduisant ainsi une dynamique non stationnaire dans les profils de consommation [34]. Ainsi, une augmentation de la consommation énergétique peut aussi bien résulter d'une réaction normale du bâtiment à une chute de la température extérieure que d'une intervention des occupants sur les consignes de chauffage. En effet, les profils d'occupation et les consignes de chauffage constituent une deuxième catégorie de facteurs, à forte composante humaine et stochastique. La présence des occupants influence la consommation à travers l'utilisation des équipements et les gains thermiques internes générés par les activités humaines [35]. Les consignes de chauffage, définies par les températures de consigne des thermostats, représentent une forme d'intervention directe des occupants sur le système énergétique. Bien que les standards usuels fixent généralement ces consignes autour de 20 °C [36], ces valeurs varient selon les préférences individuelles et les habitudes de confort propres aux ménages nordiques, constituant ainsi une source significative de variabilité comportementale. Le comportement humain étant un processus stochastique et non déterministe, il n'existe pas de température précise à laquelle tout le monde adopte un même comportement, mais une probabilité croissante en fonction des conditions environnementales [37]. Le nombre d'occupants, leurs habitudes de vie et leurs horaires introduisent des tendances saisonnières, hebdomadaires et journalières qui caractérisent les variations de la demande entre les fins de semaine, les jours fériés et les jours ouvrables [38]. De ce fait, pour des conditions climatiques identiques, deux ménages peuvent présenter des comportements énergétiques très différents.

2.1.2 *Anomalies thermiques*

Les bâtiments résidentiels constituent des systèmes thermiques complexes, dont le comportement énergétique résulte de l'interaction entre les conditions climatiques, les caractéristiques de l'enveloppe et les usages des occupants [25]. Dans les régions nordiques comme le Québec, cette dynamique est renforcée par la rigueur de l'hiver et par la forte dépendance au chauffage électrique, qui demeure le mode de chauffage le plus répandu dans la province [31]. Dans ce contexte, la littérature montre que la consommation résidentielle est étroitement liée à la température extérieure durant la saison froide, de sorte que certains comportements inefficaces des occupants peuvent se traduire de manière marquée dans les données de consommation électrique [33]. Cette relation confère ainsi aux anomalies thermiques une importance particulière dans l'évaluation de l'efficacité énergétique des bâtiments résidentiels nordiques [39]. Ces anomalies peuvent être définies comme des écarts mesurables entre la consommation électrique observée et celle attendue dans des conditions normales d'exploitation [40]. Selon la nature du phénomène en cause, elles peuvent refléter soit une perturbation transitoire du fonctionnement thermique, soit une dégradation plus durable des propriétés physiques de l'enveloppe [41].

Dans cette perspective, les anomalies thermiques réversibles correspondent à des perturbations temporaires dont les effets disparaissent une fois la cause supprimée, sans modification permanente des propriétés thermiques du bâtiment. Elles sont principalement liées aux comportements des occupants et aux modalités d'exploitation. Lors de l'ouverture d'une fenêtre en période hivernale, la température intérieure chute rapidement et le temps requis pour revenir à la température de consigne augmente en fonction de la durée d'ouverture. Il en résulte une sollicitation accrue du système de chauffage ainsi qu'une surconsommation qui s'écarte du profil énergétique nominal [42]. De manière analogue, une élévation inha-

bituelle de la consigne du thermostat constitue une autre forme d'anomalie réversible [19]. En imposant une température de confort supérieure aux valeurs usuelles, l'occupant accroît la puissance appelée par le système de chauffage; dès que la consigne retrouve sa valeur normale, le profil de consommation tend à revenir à son état nominal [37].

Par ailleurs, les anomalies thermiques irréversibles correspondent à des dégradations durables de l'enveloppe du bâtiment qui altèrent de façon persistante ses propriétés thermiques [43]. Elles trouvent principalement leur origine dans des défauts constructifs, dans le vieillissement des matériaux ou dans des mécanismes de dégradation progressive affectant l'isolation et la continuité thermique de l'enveloppe. Parmi les causes les plus fréquentes figurent la détérioration des matériaux isolants sous l'effet d'une exposition prolongée à l'humidité ou à la condensation [44]. Ces anomalies traduisent donc moins une perturbation temporaire d'usage qu'une altération physique durable du bâtiment, dont l'origine est généralement liée à l'état même de l'enveloppe.

Cette distinction revêt une importance fondamentale dans la conception de systèmes de détection appliqués au secteur résidentiel, dans la mesure où les anomalies réversibles et irréversibles se différencient par leur dynamique temporelle, leurs mécanismes d'apparition et leurs effets sur la performance énergétique du bâtiment. Elle influence ainsi le choix des variables pertinentes, le paramétrage des méthodes de détection, l'interprétation des alertes ainsi que l'orientation des actions correctives. En ce sens, distinguer ces deux catégories d'anomalies permet d'améliorer la pertinence analytique du système de surveillance, en évitant de confondre une perturbation transitoire d'usage avec une dégradation durable de l'enveloppe du bâtiment.

2.2 Détection d'anomalies

2.2.1 *Paradigmes d'apprentissage*

L'apprentissage automatique, Machine Learning (ML) en anglais, désigne un ensemble de méthodes permettant à un système informatique d'apprendre à partir de données, sans être explicitement programmé pour chaque tâche [45]. Plutôt que de suivre des règles fixes définies par un expert, un modèle d'apprentissage automatique identifie des structures, des régularités ou des relations dans les données afin de généraliser ses connaissances à de nouvelles observations. Dans le contexte de la détection d'anomalies dans les séries temporelles, ces méthodes se distinguent principalement selon les informations disponibles sur les données lors de l'entraînement [46].

Parmi ces paradigmes, l'apprentissage supervisé repose sur l'utilisation de jeux de données dans lesquels les observations normales et anormales sont explicitement étiquetées, permettant au modèle d'apprendre à partir d'exemples représentatifs afin de classer de nouvelles observations [47]. Bien que cette stratégie puisse offrir de bonnes performances lorsque des bases annotées de qualité sont disponibles, son application dans le domaine de la gestion de l'énergie électrique se heurte à des obstacles pratiques majeurs [48]. Les anomalies de consommation électrique sont par nature rares, ce qui engendre un déséquilibre de classes important et limite considérablement la capacité du modèle à apprendre des exemples suffisamment représentatifs [49]. De surcroît, la constitution de tels jeux de données requiert un investissement considérable en temps et en expertise métier, puisqu'il faut identifier, interpréter et confirmer chaque événement anormal avec un niveau élevé de fiabilité, ce qui rend cette approche souvent infaisable dans des contextes opérationnels réels [50]. À cela s'ajoute la nature non stationnaire des signaux électriques, soumis à la saisonnalité, aux dérives de concept et aux changements de régime de consommation, qui rendent d'autant plus

difficile la définition stable d'une frontière entre comportement normal et anormal [51]. Ces contraintes pratiques expliquent l'intérêt croissant porté aux paradigmes plus flexibles, qui s'affranchissent de l'intervention humaine dans l'attribution des étiquettes d'anomalies.

C'est précisément dans cette perspective que l'apprentissage non supervisé pallie ces limitations en s'affranchissant totalement des annotations. Les modèles sont entraînés sur l'ensemble des données disponibles, sans distinction préalable entre observations normales et anormales, avec l'objectif d'en capturer la structure et la distribution sous-jacente. Lors de l'inférence, un score d'anomalie est calculé pour chaque nouvel échantillon en mesurant son écart par rapport aux modèles généraux appris, si ce score dépasse un seuil prédéfini ; l'observation est classée comme anormale [52]. Ce principe repose sur l'hypothèse que les anomalies, étant rares, n'influencent que marginalement la distribution globale capturée par le modèle, et se distinguent ainsi naturellement des comportements majoritaires. Cette flexibilité est particulièrement précieuse pour concevoir des applications de détection qui ne dépendent pas de l'intervention humaine.

Parallèlement à cette famille d'approches, l'apprentissage auto-supervisé constitue un paradigme distinct, lui aussi fondé sur l'exploitation de données non annotées, mais reposant sur une logique différente [53]. Son objectif n'est pas de modéliser directement la distribution globale des données pour repérer les écarts, mais d'apprendre des représentations riches à partir de tâches prétextes construites automatiquement à partir des données elles-mêmes, sans annotation humaine. Deux grandes familles de tâches prétextes se distinguent dans la littérature, les tâches auto-prédictives, qui consistent à prédire ou reconstruire une portion des données à partir du reste, et les tâches contrastives, qui visent à rapprocher les représentations d'instances similaires tout en éloignant celles d'instances dissemblables [54].

2.2.2 Algorithmes de détection

Au cours des dernières décennies, l'adoption de nouvelles technologies telles que l'internet des objets a constitué une évolution majeure, en permettant de connecter des objets physiques à Internet afin qu'ils communiquent et interagissent de manière autonome [55]. Introduit pour la première fois en 1999 par Kevin Ashton [56], ce concept regroupe un large éventail d'appareils et de capteurs intelligents capables de collecter et de transmettre des données en temps réel. Dans le secteur de l'énergie, cette transformation offre un cadre favorable à un échange d'informations plus fluide entre le fournisseur d'énergie et les occupants, notamment grâce à l'acquisition continue des mesures de consommation via des systèmes de surveillance de charge. Toutefois, la disponibilité de ces flux en temps réel s'accompagne d'un effet direct, la constitution de bases de données volumineuses et hétérogènes, dont l'analyse manuelle devient rapidement impraticable pour distinguer de manière fiable les comportements usuels des comportements atypiques [57].

Dans ce contexte, l'apprentissage supervisé est fortement limité par la pénurie d'anomalies étiquetées et par le coût de constitution d'une vérité terrain robuste, contrainte fréquemment soulignée dans la littérature sur la détection d'anomalies en séries temporelles [48]. Par conséquent, les travaux récents accordent un intérêt croissant aux approches non supervisées, qui apprennent une représentation du comportement de référence à partir de données non annotées supposées majoritairement normales, puis quantifient l'écart des nouvelles observations par rapport à cette normalité apprise [58]. D'un point de vue opérationnel, ces modèles sont ensuite déployés sur des flux futurs, jamais observés durant l'entraînement et dans la majorité des cas non annotés en temps réel ; la décision repose alors sur un score d'écart plutôt que sur une comparaison à des étiquettes disponibles au moment de l'inférence [59]. La littérature distingue plusieurs familles d'algorithmes, classées selon leur mécanisme de

détection, que les sections suivantes passent en revue en précisant, pour chacune d'elles, le principe sous-jacent et les contributions associées.

• Méthodes à frontière de décision

Le principe de cette famille consiste à délimiter dans l'espace des caractéristiques une région englobant les données normales, tout point situé hors de cette frontière est classifié comme anomalie [60]. En effet, Himeur et al [61] mobilisent un modèle One Class Support Vector Machines (One-Class SVM) qui construit un hyperplan optimal séparant les modèles de consommation normaux de l'origine dans un espace transformé par noyau. Bien qu'efficace pour les données linéairement séparables, cette méthode demeure sensible au choix du noyau, échoue à capturer les dépendances temporelles, et son coût computationnel élevé la rend peu adaptée aux déploiements en temps réel sur des dispositifs à ressources limitées [62, 63].

• Méthodes de clustering

Cette famille repose sur le regroupement non supervisé des données sans étiquettes préalables ; tout point isolé des groupes denses est considéré comme une anomalie. Dans ce cadre, les auteurs de [59] proposent un système combinant un module Multi-Cluster Feature Selection (MCFS), qui réduit l'espace de représentation aux caractéristiques les plus discriminantes, et un modèle Gaussian Mixture Model (GMM) estimé par l'algorithme Expectation-Maximization (EM). La distribution des profils de consommation est modélisée comme une combinaison de gaussiennes, et un utilisateur est signalé comme anormal lorsque sa probabilité d'appartenance au groupe normal est jugée insuffisante.

• Méthodes d'ensemble par isolation

À l'inverse des approches précédentes qui modélisent explicitement la normalité, cette

famille exploite directement la rareté des anomalies. Le principe fondateur de l'Isolation Forest repose sur l'idée que les points anormaux, situés dans des régions peu denses de l'espace des caractéristiques, sont isolables par un nombre réduit de partitions aléatoires récursives [62]. L'algorithme construit un ensemble d'arbres de décision aléatoires dans lesquels un attribut et une valeur de séparation sont choisis aléatoirement à chaque nœud, la longueur du chemin parcouru constituant le score d'anomalie. Sa complexité computationnelle linéaire le rend particulièrement adapté aux applications en temps réel et aux données de grande dimensionnalité [64].

• Méthodes de régression

Le principe général consiste à apprendre une relation fonctionnelle $f(\mathbf{X}_t)$ entre une variable cible y_t et un ensemble de variables explicatives $\mathbf{X}_t$ à partir de données historiques, puis à comparer l'observation réelle à la valeur attendue [65]. Deux cadres complémentaires se distinguent selon la nature de la sortie du modèle. D'une part, la régression déterministe produit une estimation ponctuelle $\hat{y}_t = f(\mathbf{X}_t)$ et l'anomalie est quantifiée via le résidu $e_t = y_t - \hat{y}_t$, lequel sert directement de score de décision. D'autre part, la régression probabiliste modélise la loi conditionnelle de y_t sachant $\mathbf{X}_t$ sous forme distributionnelle $y_t \mid \mathbf{X}_t \sim \mathcal{N}(\mu_t, \sigma_t)$ [66], permettant de construire des intervalles de confiance dont le dépassement signale l'anomalie.

Plusieurs architectures ont été mobilisées selon ce cadre. En effet, Krishna et al. [67] exploitent un modèle Autoregressive Integrated Moving Average (ARIMA) pour prédire la consommation attendue à partir de l'historique de 450 compteurs intelligents, en signalant comme anormales les lectures qui s'écartent d'un intervalle de confiance à 95 %. De manière complémentaire, Abhale et al. [68] proposent une détection de vol d'électricité fondée sur des modèles Seasonal AutoRegressive Integrated Moving Average with eXogenous variables

(SARIMAX), qui estiment la consommation attendue à partir de l'historique et de variables exogènes, l'écart entre la lecture du compteur et la prévision est quantifié par le résidu, dont le dépassement d'un seuil déclenche une alerte de fraude. Par ailleurs, Fahim et al. [69] s'appuient sur une Support Vector Regression (SVR) à noyau pour prédire la charge électrique dans le cadre de la gestion énergétique de bâtiments intelligents, en identifiant comme anomalies les observations qui s'écartent significativement du comportement appris, avec un service de visualisation intégré. Enfin, Shahid et al. [70] proposent un modèle Long Short-Term Memory (LSTM) pour la prédiction séquentielle de la consommation, l'anomalie étant détectée en appliquant un seuil statistique basé sur l'écart-type des résidus. Par ailleurs, Dlamini et al. [71] comparent deux familles de modèles pour la prévision par quantiles, un modèle ensembliste non paramétrique Gradient Boosting Models (GBM) configuré en régression quantile, et un modèle Deep Quantile Regression (DQR) basé sur réseau de neurones, tous deux permettant de construire un intervalle de normalité $[\hat{q}_{0.1}(t), \hat{q}_{0.9}(t)]$ en dehors duquel une observation est considérée comme anormale. de plus, Rituka Jaiswal et al. [72] proposent quant à eux une approche en deux étapes, une phase de prévision probabiliste par Prophet, dont les intervalles de prédiction servent à labelliser les anomalies, suivie d'une phase de classification supervisée distinguant consommation normale et anormale.

• Méthodes de reconstruction

Cette famille repose sur un principe d'apprentissage de la normalité par compression et reconstruction, un encodeur projette les données dans un espace latent de dimension réduite, et un décodeur les reconstruit. L'erreur de reconstruction constitue le score d'anomalie, les données anormales, non conformes aux motifs appris, générant des erreurs systématiquement plus élevées [51, 64]. Dans ce contexte, plusieurs variantes architecturales ont été développées. L'Autoencodeur Convolutif (CAE) applique des convolutions unidimensionnelles pour

capturer des motifs locaux avec un nombre réduit de paramètres [73]. L’Autoencodeur (AE) débruiteur renforce la robustesse des représentations apprises en entraînant le modèle à reconstruire les données originales à partir d’entrées volontairement bruitées [50]. Le Variationnel Autoencodeur (VAE) se distingue en modélisant l’espace latent comme une distribution gaussienne plutôt que comme une représentation déterministe, intégrant un terme de régularisation par divergence de Kullback-Leibler qui favorise un espace latent continu et structuré [74]. Les architectures hybrides Convolutional Neural Networks and Long Short-Term Memory (CNN-LSTM) combinent quant à elles la capacité des couches convolutionnelles à détecter les motifs locaux avec celle des couches LSTM à capturer les dépendances temporelles à long terme, démontrant des performances particulièrement solides sur les séries temporelles industrielles [64]. Enfin, dans le cas des systèmes multivariés à forte dimensionnalité et intercorrélations élevées, les Graph Neural Networks (GNN) constituent une extension naturelle de ce paradigme, en modélisant explicitement les relations de causalité entre variables au sein de la structure de reconstruction [51].

Ainsi, la diversité des algorithmes recensés dans la littérature représente à la fois une force et un défi. Si elle témoigne de la richesse des approches développées, elle complique également la sélection et l’adaptation des modèles aux contextes réels d’application. Dans ce cadre, les études comparatives unifiées revêtent une importance particulière, car elles permettent d’analyser de façon cohérente les performances des différentes méthodes et d’identifier celles qui sont les plus adaptées à chaque situation. La littérature recense également plusieurs études comparatives visant à examiner de manière approfondie les limites des différentes familles de méthodes de détection d’anomalies. Toutefois, chaque travail explore une ou plusieurs familles de méthodes sur des jeux de données distincts, ce qui rend les comparaisons directes difficiles, tout en faisant ressortir certaines tendances stables, notamment en ce qui concerne le comportement des faux positifs.

Les premières comparaisons ont naturellement intégré des méthodes statistiques simples comme point de départ. C'est notamment le cas de Basu Hela et al. [74], qui proposent une étude comparative portant sur 200 bâtiments à partir de relevés horaires réels, et qui incluent l'Interquartile Range (IQR) et le modified Z-score (mZ-score) parmi les méthodes évaluées. Ces deux approches définissent un seuil à partir de la distribution statistique des données, l'IQR se fonde sur les quartiles, tandis que le mZ-Score utilise la médiane des écarts absolus afin d'être plus robuste aux valeurs extrêmes. L'idée sous-jacente est qu'un point suffisamment éloigné de la distribution habituelle peut être considéré comme une anomalie. Cependant, cette simplicité constitue également leur principale limite dans le contexte énergétique résidentiel. En effet, la consommation électrique ne suit pas une distribution fixe, elle augmente le matin, diminue la nuit, varie selon les jours de la semaine et évolue au fil des saisons. En ignorant cette structure temporelle, ces méthodes confondent fréquemment une pointe de consommation normale, par exemple un lundi matin, avec une anomalie réelle. Ainsi, la précision maximale rapportée par [74] n'excède pas 52 % pour l'IQR et 46 % pour le mZ-Score, ce qui signifie que près d'un signal sur deux correspond à une fausse alarme.

Par ailleurs, Wenping Zhang et al. [75] comparent plusieurs algorithmes sur une série temporelle réelle de fréquence horaire couvrant une période de deux mois, en incluant notamment le Local Outlier Factor (LOF) comme méthode de référence. Les résultats obtenus mettent en évidence des performances limitées, avec une précision maximale de 24,46 %, le LOF génère environ trois fausses alarmes pour chaque anomalie réelle détectée. Cette faiblesse s'explique par une limite structurelle comparable à celle des méthodes statistiques, le LOF évalue l'isolement d'un point par rapport à ses voisins dans l'espace des caractéristiques, sans prendre en compte le contexte temporel dans lequel ce point apparaît. De plus, Basu Hela et al. [74] confirment également cette limite dans leur propre étude comparative, en rapportant une précision maximale de 48,30 % pour le LOF appliqué à des données énergé-

tiques de bâtiments. Cependant, dans ces deux études comparatives, Isolation Forest obtient des performances supérieures à celles du LOF. En effet, Wenping Zhang et al. [75] rapportent une précision maximale de 63,33 %, tandis que Basu Hela et al. [74] obtiennent une précision de 54,90 %. Une analyse plus nuancée de cette méthode est proposée par [76], qui mènent une étude comparative spécifiquement fondée sur des données résidentielles réelles issues du projet Pecan Street à Austin, au Texas. Leurs résultats montrent que Isolation Forest peut atteindre un coefficient Matthews Correlation Coefficient (MCC) de 0,813, soit l'une des meilleures performances relevées dans la littérature examinée. Toutefois, cette performance n'est obtenue que sous certaines conditions, en particulier lorsque la sélection des variables d'entrée est adéquatement réalisée. Les auteurs montrent notamment que l'association des données historiques de charge avec des variables environnementales, telles que la température et l'humidité, améliore significativement les résultats. Ainsi, cette étude met en évidence un aspect essentiel, souvent moins explicitement documenté dans les autres travaux. Dans le cas d'Isolation Forest, la qualité de la sélection des variables explicatives joue un rôle aussi déterminant que le choix de l'algorithme lui-même. De plus, Aligholian et al. [76] incluent également Lightweight On-line Detector of Anomalies (LODA) dans leur étude comparative. Cette méthode repose sur la projection des données selon des vecteurs aléatoires creux, suivie de la construction d'histogrammes permettant d'estimer un score de vraisemblance. Les résultats montrent que LODA atteint un MCC de 0,545, mais qu'il demeure particulièrement sensible aux périodes de très faible consommation. Cette sensibilité se traduit par la génération de fausses alarmes récurrentes durant les créneaux où la charge est proche de zéro.

Une rupture plus significative dans la littérature comparative est intervenue avec les méthodes basées sur la prédiction, qui apprennent d'abord ce que devrait être la consommation normale avant de décider si une valeur est anormale. Wenping Zhang et al. [75] proposent dans leur étude comparative RobustProphet, qui combine Prophet, un modèle de décomposi-

tion de séries temporelles tenant compte de la tendance, de la saisonnalité et des jours fériés, avec XGBoost pour corriger les erreurs de prédiction [75]. L'anomalie est détectée non pas sur la valeur brute mais sur le résidu entre la valeur prédite et la valeur observée, avec un seuil dynamique non-paramétrique qui s'ajuste automatiquement pour minimiser les fausses alarmes en pénalisant les détections trop nombreuses. C'est cette combinaison qui explique que RobustProphet atteigne une précision de 88.10 % dans leur comparaison, la meilleure parmi toutes les méthodes qu'ils évaluent, là où LOF ne dépassait pas 24.46 % sur le même jeu de données. De plus, Davor Stjelja et al. [77] étendent cette logique comparative en évaluant deux méthodes probabilistes, l'approche Extension of XGBoost to probabilistic modelling and prediction (XGBoostLSS) et l'algorithme Ensemble batch Prediction Intervals (EnbPI), sur des données de bâtiments finlandais couvrant à la fois le chauffage et l'électricité. Contrairement aux approches déterministes fondées sur une valeur prédite unique, ces méthodes génèrent un intervalle de valeurs attendues. Plus précisément, XGBoostLSS modélise directement la distribution complète de la consommation au moyen d'une distribution Beta, tandis que EnbPI repose sur la prédiction conforme avec des résidus leave-one-out. À cet effet, une matrice d'alarme est ensuite utilisée afin de filtrer les anomalies ponctuelles isolées et de ne retenir que les déviations récurrentes et systémiques, ce qui contribue à réduire les fausses alarmes. Dans leur comparaison, XGBoostLSS atteint ponctuellement une précision de 1, sur certains bâtiments, mais avec une forte variabilité selon le contexte. En revanche, EnbPI, dont les intervalles plus étroits le rendent plus sensible aux fluctuations normales, présente une précision maximale de 85 %. Ainsi, ces résultats montrent que les approches probabilistes peuvent améliorer la fiabilité de la détection, mais que leur performance demeure fortement dépendante du contexte d'application et de la largeur des intervalles prédictifs. Par ailleurs, Aligholian et al. [76] incluent également dans leur étude deux méthodes à prédiction déterministe, Load Prediction with Regression (LPBSVR) et Load Prediction

with Neural Network (LPBNN) . Ces méthodes apprennent la consommation normale respectivement par SVR et par réseau de neurones, puis signalent comme anomalie tout résidu dépassant le seuil $\pm 3\sigma$. Toutefois, ce seuil fixe constitue la principale limite identifiée par les auteurs, car il ne s'adapte pas aux changements de régime de consommation et génère des fausses alarmes lors des transitions saisonnières. Par conséquent, les performances restent modérées, avec des valeurs de MCC de 0,493 pour LPBSVR et de 0,475 pour LPBNN.

L'exploration comparative du deep learning non supervisé constitue la direction la plus récente. Kukjin Choi et al proposent la comparaison la plus exhaustive dans ce domaine, évaluant un VAE entraîné spécifiquement sur les données énergétiques, le modèle de fondation MOMENT en versions zéro-shot et fine-tunée, et TimeGPT [74]. Le VAE, qui apprend à reconstruire des sous-séquences de consommation et détecte les anomalies comme les points dont l'erreur de reconstruction est anormalement élevée, atteint une précision de 79.80 %, la meilleure de toute leur comparaison. Ce résultat est d'autant plus remarquable que ce modèle ne compte que 0,168 million de paramètres, contre (375) millions pour MOMENT, tout en étant environ 1,172 fois plus rapide à l'inférence. Par ailleurs, MOMENT fine-tuné, qui adapte ses représentations aux données cibles par masquage auto-supervisé de sous-séquences, améliore la précision à 74.10 % par rapport à sa version zéro-shot à 68.30 %. TimeGPT, en revanche, qui utilise un intervalle de confiance de prédiction à 99 % sans aucune adaptation au domaine, produit le résultat le plus décevant de leur comparaison avec une précision de seulement 29.60 %, environ 70 % de ses alertes sont de fausses alarmes, ce que les auteurs attribuent à une inadéquation fondamentale entre un modèle conçu pour la prévision et la nature des anomalies énergétiques. De plus, Kukjin Choi et al. [51] proposent quant à eux une étude comparative de neuf méthodes deep learning, dont OmniAnomaly, sur des données industrielles multivariées de capteurs. Toutes atteignent des précisions supérieures à 90 %, OmniAnomaly frôlant 99.01 %. Ces résultats sont les meilleurs de toute la littérature

recensée, mais ils concernent des données de systèmes de traitement d'eau et de rovers spatiaux, structurellement différentes de la consommation électrique résidentielle, et ne peuvent pas être transposés directement à ce contexte sans précaution. Ce panorama révèle que, malgré la diversité des approches comparées dans la littérature, aucune étude ne documente une méthode atteignant simultanément une précision et un rappel de 1, un compromis structurel que tous les travaux confirment à leur façon, et que les méthodes les plus récentes cherchent à gérer par des mécanismes de seuillage adaptatif et de post-traitement plutôt qu'à résoudre définitivement.

2.2.3 *Modélisation statistique et paramétrisation*

La littérature récente sur la détection d'anomalies dans les séries temporelles met en évidence une orientation croissante vers les approches fondées sur la prédiction et l'analyse des résidus. Selon PAPARRIZOS et al. [78], cette évolution traduit l'importance grandissante des méthodes capables d'apprendre le comportement normal d'une série, puis d'identifier une anomalie à partir de l'écart entre la valeur prédite et la valeur observée. Dans cette perspective, l'anomalie n'est plus seulement considérée comme une rupture directement perceptible dans le signal, mais comme une divergence par rapport à ce que le modèle reconnaît comme normal. HOWARD et OTOBO [79] confirment également cette orientation dans un cadre appliqué, en s'appuyant sur les résidus issus d'un modèle de prévision comme fondement du processus de détection.

Toutefois, l'intérêt de ces approches ne doit pas occulter leur principale limite. Leur performance dépend étroitement de la manière dont le résidu est interprété, et plus particulièrement du seuil retenu pour décider qu'un écart devient anormal. Dans de nombreux travaux, ce seuil est défini à partir de la moyenne des résidus augmentée d'un multiple de leur écart-type [80]. Une telle stratégie présente l'avantage de la simplicité, mais elle repose

implicitement sur l'hypothèse d'une relative stabilité statistique des résidus. Or, cette hypothèse est rarement vérifiée dans les séries temporelles réelles, où la variabilité du signal évolue au fil du temps sous l'effet de changements de régime, de fluctuations saisonnières ou de dérives progressives. Dès lors que la variance des résidus augmente de manière normale, sans qu'une anomalie soit réellement présente, ce type de seuil tend à produire un nombre excessif d'alertes. La littérature souligne ainsi que le seuillage fixe est particulièrement sensible au bruit, aux décalages temporels et aux variations de distribution des résidus [80]. Cette limite rappelle que, dans les approches fondées sur la prédiction, la détection ne dépend pas seulement du calcul de l'erreur, mais aussi de la manière dont cette dernière est replacée dans son contexte.

C'est précisément sur ce point que le seuil dynamique apparaît comme une réponse plus appropriée. Son principe consiste à recalibrer en continu la limite de décision à partir du comportement local des résidus de prédiction. Le seuil n'est alors plus défini comme une constante globale, mais comme une valeur relative estimée sur une fenêtre récente d'observations à partir des propriétés statistiques locales des résidus. Tout résidu dépassant cette limite est alors interprété comme potentiellement anormal. Une telle approche permet d'ancrer la décision dans l'état courant du système, plutôt que dans une estimation globale souvent trop rigide. PAPARRIZOS et al. [78] montrent que cette logique adaptative est au cœur de plusieurs méthodes récentes, tandis que HOWARD et OTOBO [79] en soulignent l'intérêt dans des contextes où la variabilité du signal évolue fortement au cours du temps. L'apport principal du seuil dynamique réside dans sa capacité à suivre l'évolution des résidus. Lorsque le signal devient plus instable, le seuil s'ajuste à cette nouvelle dispersion et limite le risque de fausses alarmes. À l'inverse, lorsque le système évolue dans un régime plus stable, il devient plus sensible aux écarts faibles mais significatifs. Cette adaptation améliore la cohérence entre la décision de détection et la réalité statistique du signal.

Néanmoins, le seuil dynamique ne fait pas disparaître entièrement la difficulté. Il la déplace vers un défi de paramétrisation algorithmique qui constitue l'un des points les plus épineux de la littérature. La question centrale est la suivante, comment ajuster la sensibilité du seuil de manière à détecter les anomalies le plus rapidement possible, idéalement dès leur apparition, tout en limitant l'augmentation indésirable du taux de faux positifs ? Ce dilemme est structurel, car les résidus de prédiction ne présentent pas une amplitude constante. Leur comportement varie selon le contexte, le régime du signal et les caractéristiques du modèle de prédiction utilisé. Un seuil trop sensible multiplie les alertes déclenchées par de simples erreurs de prédiction non significatives, tandis qu'un seuil trop conservateur retarde la détection au-delà du délai d'intervention utile. Lorsque le volume de fausses alarmes devient excessif, le système engendre ce que les praticiens désignent comme la fatigue d'alerte, les opérateurs finissent par ignorer toutes les alertes, y compris les plus critiques, ce qui annule en pratique les bénéfices attendus du système de détection [79]. PAPARRIZOS et al. [78] confirment cette réalité en soulignant qu'aucune méthode de prédiction ne domine toutes les autres sur l'ensemble des domaines testés, ce qui signifie que le paramétrage du seuil doit nécessairement être adapté au contexte local plutôt que défini de façon universelle.

À ce stade, PAPARRIZOS et al. [78] mettent notamment en avant le lissage temporel du résidu avant son seuillage comme un premier levier. En moyennant l'erreur de prédiction sur une courte fenêtre glissante, il devient possible d'atténuer les fluctuations isolées liées au bruit du modèle, tout en préservant les anomalies réelles lorsqu'elles persistent dans le temps. Ce mécanisme permet ainsi de réduire les alertes injustifiées, à condition que la fenêtre de lissage soit calibrée avec soin afin de ne pas introduire une latence excessive dans la détection. D'autre part, ces mêmes auteurs soulignent qu'aucune valeur de paramétrage n'est universellement optimale et que le réglage du seuil doit nécessairement être adapté au contexte local ainsi qu'aux caractéristiques statistiques des données d'apprentissage. Dans cette perspective,

la sensibilité du seuil dynamique doit être réglée avec soin, car elle conditionne directement le compromis entre réactivité et robustesse, et son efficacité dépend étroitement des caractéristiques statistiques des résidus observés en phase d'apprentissage.

Ainsi, l'évolution récente de la littérature ne conduit pas seulement à privilégier les approches par prédiction, mais aussi à reconnaître que leur efficacité dépend étroitement de la gestion des résidus et de leur seuillage. Le passage d'un seuil fixe à un seuil dynamique constitue une avancée importante, dans la mesure où il rapproche la décision de détection du comportement réel du signal. Toutefois, cette avancée ouvre à son tour un nouveau champ de réflexion centré sur la calibration du seuil et sur la maîtrise du compromis entre vitesse de détection et taux de faux positifs. En ce sens, la détection d'anomalies fondée sur la prédiction apparaît moins comme un problème purement algorithmique que comme un problème d'équilibre entre modélisation, adaptation et décision.

2.2.4 Outils logiciels et frameworks Open Source

L'analyse des séries temporelles de consommation énergétique, dans le cadre de la détection d'anomalies, ne peut être effectuée de manière isolée, car elle est fortement influencée par un ensemble de variables explicatives contextuelles [81]. En effet, les approches de détection d'anomalies reposent rarement sur l'étude exclusive des séries de consommation brute ; elles intègrent des facteurs externes permettant de mieux interpréter les fluctuations observées et de réduire les fausses alertes [57].

Dans le secteur résidentiel, le comportement des occupants constitue l'un des facteurs les plus influents, car il est directement intégré au profil de consommation électrique en raison de leurs habitudes d'utilisation de l'énergie [82]. D'un point de vue analytique, ce comportement peut être pris en compte dans les algorithmes de détection à travers des indicateurs comportementaux reliant la présence ou l'absence des occupants à la consommation éner-

gétique. Cette information est généralement obtenue grâce à des capteurs de mouvement ou de présence, qui produisent des données binaires (présence/absence) et offrent une base utile pour affiner l'interprétation des variations de charge et éviter la détection erronée d'anomalies [83]. Par exemple, une baisse soudaine de la consommation peut simplement traduire une absence prolongée des occupants plutôt qu'un dysfonctionnement d'un appareil électrique.

Par ailleurs, les facteurs calendaires offrent une dimension complémentaire pour la modélisation du comportement des occupants. Ils reflètent les habitudes de consommation périodiques, influencées par les jours de la semaine, les week-ends, les jours fériés ou encore les heures de la journée, et conditionnent les profils de charge selon les cycles d'occupation du logement [84]. Ainsi, la consommation tend à être plus élevée en soirée ou durant les fins de semaine, lorsque la présence des occupants est accrue, tandis qu'elle diminue la nuit ou lors des absences prolongées. Toutefois, l'intégration de ces informations contextuelles dans les systèmes de surveillance de charge introduit une complexité supplémentaire, elle nécessite à la fois l'ajout de capteurs additionnels et une capacité de stockage accrue pour traiter des volumes de données plus importants. Ce surcroît d'information génère une charge computationnelle plus élevée pour les algorithmes de détection, qui doivent gérer des mégadonnées en temps réel. De plus, la collecte d'informations liées à la présence et aux comportements des occupants soulève des préoccupations relatives à la vie privée, pouvant affecter l'acceptabilité et le confort des utilisateurs dans les environnements résidentiels connectés [4].

En complément de ces variables comportementales et temporelles, les conditions météorologiques jouent également un rôle déterminant dans la précision des systèmes de détection d'anomalies de consommation énergétique [85]. La consommation des ménages est en effet fortement corrélée aux variables climatiques telles que la température extérieure, l'humidité ou encore le rayonnement solaire, qui influencent directement les besoins en chauffage, en

climatisation et en ventilation. Cette dépendance est particulièrement marquée dans les régions à climat rigoureux, comme le Québec, où les longues saisons froides entraînent une hausse significative de la demande énergétique en période hivernale [33]. Dans ce contexte, une augmentation de la consommation ne traduit pas nécessairement une anomalie, mais peut refléter une réponse normale aux conditions environnementales. Ainsi, la contextualisation climatique s'avère essentielle pour garantir une détection fiable, explicable et adaptée aux spécificités des bâtiments résidentiels nordiques.

Dans cette continuité, les comportements humains, les effets calendaires et les conditions climatiques confèrent aux données de consommation énergétique résidentielle une nature hautement dynamique et stochastique. Cette caractéristique résulte de la combinaison de phénomènes déterministes et aléatoires qui façonnent les profils de charge électrique au fil du temps [33]. La dimension dynamique traduit la présence de régularités temporelles observables à différents horizons, des cycles quotidiens (hausse de la demande le matin et en soirée), hebdomadaires (différences entre les jours ouvrables et les week-ends) et saisonniers (augmentation du chauffage en hiver, de la climatisation en été). Ces variations suivent des schémas relativement stables, influencés par la routine des occupants et les conditions environnementales [25]. Dès lors, le choix des outils logiciels devient particulièrement important lorsque l'on souhaite intégrer explicitement ces variables explicatives dans un protocole expérimental rigoureux. Les bibliothèques Python dédiées à la détection d'anomalies offrent une large gamme de frameworks consacrés à cette tâche.

Dans ce cadre, l'usage d'outils open source capables de gérer nativement ces variables contextuelles présente un double avantage. D'abord, il renforce la reproductibilité, car les covariables sont intégrées explicitement dans le pipeline d'apprentissage et d'inférence, au lieu d'être ajoutées de manière spécifique et non systématisée dans des scripts difficiles à ré-

plier³. Ensuite, il améliore la fiabilité scientifique, car le code, les versions logicielles, les transformations et les paramètres peuvent être documentés, partagés et vérifiés. Les travaux sur la reproductibilité en apprentissage automatique montrent d’ailleurs que la disponibilité du code et des pipelines est un facteur essentiel pour reproduire fidèlement les résultats expérimentaux [86]. Ainsi, pour un mémoire de recherche, il est préférable de privilégier une bibliothèque open source qui modélise explicitement le contexte, plutôt qu’un outil nécessitant beaucoup de prétraitements manuels.

Dans cette perspective, on distingue, d’une part, les outils conçus directement pour les séries temporelles, qui proposent une Application Programming Interface (API) cohérente, des objets temporels natifs et des pipelines adaptés à la prévision, au scoring et à la détection [87, 88]. D’autre part, on trouve des bibliothèques de détection d’outliers plus générales, très riches en algorithmes, mais qui ne sont pas pensées nativement pour la dépendance temporelle et nécessitent souvent un fenêtrage ou une transformation préalable des données [89]. En effet, la bibliothèque Darts⁴ ressort comme l’une des options les plus intéressantes pour un travail de recherche, car elle combine la modélisation des séries temporelles, la prévision et la détection d’anomalies dans un cadre unifié [87]. Elle prend en charge les séries multivariées, les covariables passées et futures, ainsi que les covariables statiques, ce qui en fait un choix particulièrement pertinent lorsque l’on souhaite intégrer des variables explicatives comme la météo, les horaires ou des consignes de fonctionnement. De même, Merlion propose une approche très complète, intégrant détection d’anomalies, prévision, post-traitement et support de variables exogènes [88, 90]. Par ailleurs, Automated Time-series Outlier Detection System (TODS) est intéressant sur le plan conceptuel car il vise la construction automatique de pipelines de détection d’anomalies sur séries temporelles multivariées [91]. Toutefois,

3. Salesforce, *Merlion : A Machine Learning Framework for Time Series Intelligence*, <https://github.com/salesforce/Merlion>, consulté en mars 2026.

4. Unit8, *Covariates / Darts Documentation*, <https://unit8co.github.io/darts/userguide/covariates.html>, consulté en mars 2026.

son écosystème semble moins actif et plus fragile du point de vue des dépendances et de la maintenance, ce qui limite son attractivité pour un mémoire qui cherche une base stable et facilement reproductible⁵. À l'inverse, Python Outlier Detection (PyOD) constitue une bibliothèque très mature et très riche en détecteurs d'anomalies, avec une API simple et un large catalogue d'algorithmes classiques et récents. En revanche, elle reste avant tout pensée pour des données tabulaires multivariées indépendantes, et non pour des séries temporelles [89]. Son utilisation dans un contexte temporel est donc possible, mais elle suppose des étapes supplémentaires de préparation des données, comme le fenêtrage ou l'extraction de caractéristiques. Cependant, Python Deep Outlier Anomaly Detection (DeepOD) est davantage orientée vers les approches de deep learning modernes pour la détection d'anomalies. Elle est intéressante pour explorer des modèles récents appliqués aux séries temporelles multivariées, mais elle demande généralement plus de ressources de calcul et un réglage plus fin des hyperparamètres [92].

2.2.5 Analyse des lacunes

L'état de la littérature met en évidence une progression importante des méthodes de détection d'anomalies, tout en révélant des limites majeures qui, loin d'être indépendantes, se renforcent mutuellement et justifient la conduite de ce mémoire.

D'une part, les travaux recensés comparent certes plusieurs familles d'algorithmes, mais le plus souvent sur des jeux de données hétérogènes, issus de contextes différents et évalués selon des variables d'entrée, des horizons temporels, des protocoles expérimentaux et des métriques non unifiés. Dans ces conditions, les résultats demeurent difficilement comparables d'une étude à l'autre, ce qui limite toute conclusion robuste sur la supériorité relative d'une méthode dans le cas spécifique des bâtiments résidentiels. Cette difficulté est d'autant plus

5. Python Package Index, *tods*, <https://pypi.org/project/tods/>, consulté en mars 2026.

marquée que le domaine résidentiel ne peut être réduit à la seule série de consommation électrique. Le comportement énergétique d'une résidence dépend fortement des conditions météorologiques, des habitudes d'occupation, des effets calendaires et, plus largement, du contexte thermique propre au bâtiment. Or, la littérature propose rarement des comparaisons menées dans un cadre où ces différentes sources d'information sont intégrées de manière cohérente et évaluées sous un protocole commun. En outre, une part importante des meilleurs résultats rapportés provient soit de bâtiments non résidentiels, comme les comparaisons de Basu Hela et al. [74] menées sur le jeu LEAD 1.0 regroupant des bâtiments commerciaux, soit de systèmes industriels multivariés, comme les architectures de deep learning évaluées par CHOI et al. [51] sur des données de capteurs industriels, dont les propriétés statistiques, les régimes de fonctionnement et les signatures d'anomalies diffèrent sensiblement de ceux d'un habitat résidentiel nordique. Il en résulte un manque marqué de connaissances comparatives directement transposables au cas résidentiel multi-sources.

Dans le prolongement de ce constat, la question de la paramétrisation apparaît comme un second point de fragilité particulièrement important, notamment dans les approches fondées sur la prédiction et l'analyse des résidus. La littérature récente montre clairement que l'efficacité de ces méthodes ne dépend pas seulement de l'algorithme de prévision utilisé, mais aussi du réglage du mécanisme de décision appliqué aux résidus, en particulier lorsqu'il repose sur un seuil dynamique. Pourtant, peu d'études analysent de manière systématique l'effet de cette paramétrisation sur les performances de détection dans un cadre résidentiel. Certains travaux obtiennent de bonnes performances grâce à un réglage fin, comme ZHANG et al. [75] avec RobustProphet ou ALIGHOLIAN [76] avec Isolation Forest, mais documentent insuffisamment la sensibilité de leurs résultats à ces choix méthodologiques. De même, STJELJA et al. [77] proposent des seuils adaptatifs sans quantifier formellement comment leur réglage affecte le comportement du détecteur selon le type d'anomalie rencontré. En pratique, le réglage de la

sensibilité du seuil, le lissage temporel des résidus ou encore les règles de déclenchement d’alerte sont souvent présentés comme des choix secondaires, alors qu’ils conditionnent directement le comportement opérationnel du détecteur.

Par conséquent, ce manque d’analyse croisée constitue un verrou méthodologique majeur, car il touche au compromis central de l’application visée, assurer une hyper-réactivité de détection, indispensable pour signaler rapidement une anomalie réversible ou un comportement énergivore, tout en minimisant les faux positifs afin d’éviter la fatigue d’alerte et la perte de confiance dans le système. La littérature reconnaît l’existence de ce compromis et montre également, comme le soulignent PAPARRIZOS et al. [78], qu’aucune méthode ne domine systématiquement toutes les autres dans l’ensemble des contextes étudiés. Ce constat renforce l’idée que le paramétrage du seuil ne peut être défini de manière universelle, mais doit être adapté au contexte local. Toutefois, la littérature ne fournit pas encore de cadre expérimental suffisamment structuré pour mesurer, dans le domaine résidentiel multi-sources, comment les choix de paramétrisation influencent concrètement l’équilibre entre rapidité de détection et robustesse des alertes. De plus, en l’absence d’un cadre comparatif unifié, il devient méthodologiquement difficile d’isoler l’effet du paramétrage de celui du contexte ou de la méthode elle-même. Ainsi, le manque de cadre comparatif homogène et l’insuffisante étude de la paramétrisation ne constituent pas deux limites séparées, mais bien deux dimensions étroitement liées d’un même problème scientifique.

C’est précisément à l’intersection de ces limites que se situe le présent mémoire. Il vise, d’une part, à proposer un cadre de comparaison homogène entre plusieurs familles d’algorithmes de détection d’anomalies appliquées à des données résidentielles contextualisées par plusieurs sources d’information. Il cherche, d’autre part, à analyser explicitement l’impact de la paramétrisation sur le compromis entre réactivité et faux positifs. En ce sens, l’objectif

n'est pas uniquement d'identifier l'algorithme le plus performant en moyenne, mais de déterminer dans quelles conditions expérimentales, avec quelles variables contextuelles et avec quel réglage de sensibilité, une méthode devient réellement pertinente pour une application résidentielle orientée vers l'efficacité énergétique.

Chapitre 3 – Méthodologie

3.1 Architecture du système

Le système d'analyse d'anomalies proposé dans le cadre de ce travail s'articule en six étapes successives, allant de l'acquisition des données brutes jusqu'à la prise de décision finale. La première étape consiste en l'acquisition des données de consommation énergétique en temps réel, mesurées au sein d'une résidence unifamiliale située à Trois-Rivières, au Québec. En parallèle, les données météorologiques associées à la même région sont collectées, notamment les températures extérieures, ces variables jouant un rôle déterminant dans la modélisation de la consommation énergétique résidentielle. Ces données brutes font ensuite l'objet d'un prétraitement rigoureux incluant le nettoyage des valeurs manquantes, la synchronisation temporelle des séries et la normalisation des variables, afin de garantir la qualité et la cohérence des entrées du modèle. La deuxième étape consiste à transmettre ces données prétraitées à un modèle prédictif dont le rôle est d'estimer le comportement normal attendu du système de consommation.

Ce modèle apprend les profils habituels de consommation en fonction des conditions météorologiques et des habitudes des occupants, de manière à produire une estimation fiable de la consommation en régime normal. La troisième étape porte sur le calcul des résidus, définis comme l'écart entre la valeur réelle observée et la valeur prédite par le modèle. Ce signal résiduel constitue l'indicateur central de l'analyse, car il isole les déviations par rapport au comportement normal tout en s'affranchissant des effets saisonniers et climatiques déjà capturés par le modèle. Ce signal résiduel alimente ensuite simultanément deux branches pa-

rallèles de traitement. La première est une branche de référence, construite à partir de données enregistrées en régime strictement normal, c'est-à-dire en l'absence de toute anomalie. La seconde est une branche d'analyse, construite à partir de données dans lesquelles une anomalie a été injectée artificiellement, simulant ainsi un comportement anormal de consommation. Dans la branche de référence, un seuil statistique correspondant au quantile de la distribution des résidus normaux est appliqué afin d'identifier les dépassements naturels du signal. Pour chaque dépassement détecté, une surface est calculée selon la géométrie locale du signal résiduel, une approximation triangulaire est utilisée pour les pics isolés composés d'un seul point de dépassement, tandis qu'une approximation trapézoïdale est appliquée pour les séquences de dépassements comportant plusieurs points consécutifs. L'ensemble de ces surfaces forme une distribution empirique propre au régime normal, dont la queue supérieure caractérise les épisodes les plus sévères observables en conditions de fonctionnement nominales. Un seuil de sévérité S_q est alors fixé au quantile haut q_{high} , ce dernier variant entre $P60$ et $P99$ de cette distribution. Ce choix offre une référence statistiquement plus robuste que le simple maximum, en filtrant les valeurs extrêmes potentiellement aberrantes tout en conservant l'information sur le niveau de sévérité tolérable en régime normal. Dans la branche d'analyse, après l'injection de l'anomalie, le même procédé de détection et de calcul de surface est appliqué au signal résiduel anormal. Le critère de décision repose sur le concept de Time-to-Equal-Severity (TES), qui est évalué pour chaque dépassement du signal anormal dont la surface totale excède le seuil S_q . Pour un tel dépassement, le moment de détection est défini comme le premier instant où la surface cumulée, accumulée depuis le début du dépassement, franchit strictement S_q .

Cette évaluation s'appuie sur des métriques classiques de classification binaire, à savoir la précision, le rappel, le F1-score et la spécificité, calculées selon deux approches complémentaires, une évaluation individuelle, fournissant les métriques propres à chaque dépassement

ayant déclenché un TES, ainsi qu’une évaluation globale obtenue en agrégeant l’ensemble des détections par union des zones identifiées. Cette double lecture permet d’apprécier à la fois la qualité locale de chaque détection et la capacité globale du système à détecter les anomalies réelles tout en restant robuste face aux fausses alarmes. La Figure 3.1 présente l’architecture de la méthode de détection des anomalies.

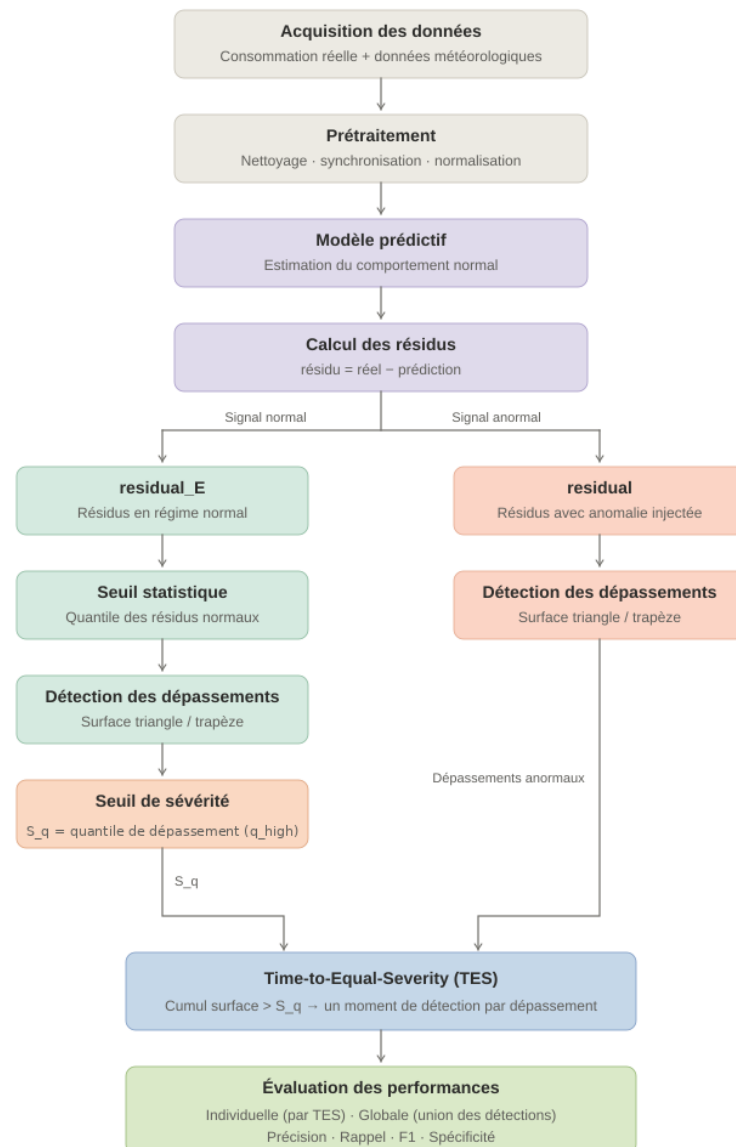


FIGURE 3.1 Architecture de la méthode de détection des anomalies.

3.1.1 Acquisition des données

La base de données utilisée dans cette étude a été construite à partir d'un dispositif d'acquisition en temps réel, combinant deux sources complémentaires. D'une part, les données de consommation électrique résidentielle proviennent d'un système avancé de surveillance de charge, **Emporia Vue Gen 2**, conçu pour enregistrer de manière continue la puissance électrique globale consommée au sein d'une habitation, avec une résolution temporelle d'une minute.

D'autre part, les données météorologiques associées au site étudié sont obtenues à partir de la plateforme en ligne Tomorrow.io¹, qui fournit des informations environnementales détaillées, incluant la température extérieure, l'humidité relative, le rayonnement solaire et la vitesse du vent. Les deux volets, énergétique et météorologique, ont été intégrés dans InfluxDB², sous licence du Massachusetts Institute of Technology (MIT), afin de constituer une base centralisée dédiée au stockage et à la gestion des séries chronologiques. Spécialement conçue pour ce type de données, InfluxDB offre des fonctionnalités adaptées à la structuration, à l'organisation et à l'interrogation efficace des flux collectés. Cette solution se distingue également par sa légèreté et ses performances, ce qui la rend particulièrement appropriée aux environnements disposant de ressources limitées, notamment dans le cadre d'applications de collecte de données en périphérie (*edge computing*)³

Sa rapidité d'exécution, ainsi que sa capacité à traiter en continu de grands volumes de données, assurent une réactivité suffisante pour permettre des traitements analytiques quasi en temps réel.

1. Tomorrow.io, *Weather Intelligence Platform*, 2024. Site consulté en mai 2025, <https://www.tomorrow.io/>.
 2. InfluxData, Capture de données granulaires provenant de capteurs et de données de séries chronologiques, <https://www.influxdata.com/blog/influxdb-3-ideal-solution-real-time-analytics/>, consulté en mai 2025.
 3. Amazon Web Services, *Building an Edge-to-Cloud IIoT Architecture Using InfluxDB and AWS*, <https://aws.amazon.com/blogs/iot/building-an-edge-to-cloud-iiot-architecture-using-influxdb-and-aws/>, consulté en mai 2025.

- **Système de surveillance de charge**

Grâce à son suivi en temps réel, l'Emporia Vue Gen 2⁴ permet de suivre avec précision la consommation électrique du bâtiment. Il délivre des données précises et granulaires qui permettent de mieux comprendre les profils d'usage énergétique en milieu résidentiel. Installé dans une habitation, ce système surveille plusieurs charges électriques, la consommation globale, les appareils électroménagers, le chauffage et d'autres équipements connectés. Il fournit ainsi des informations détaillées pour chaque point de consommation. La Figure 3.2 illustre les différentes étapes du processus de surveillance de la charge.

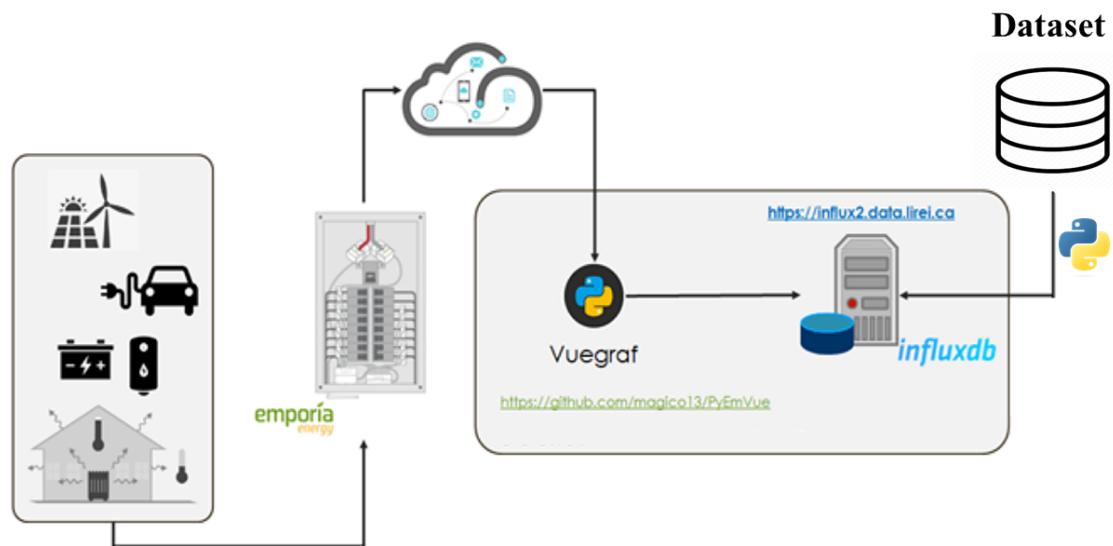


FIGURE 3.2 Système de surveillance de charge en temps réel

Pour acheminer les relevés d'énergie du dispositif de surveillance de charge vers la base de données locale, la bibliothèque Python Vuegraf⁵ joue le rôle d'interface entre le dispositif de mesure d'énergie Emporia Vue Gen 2 et une base de données InfluxDB locale, en assurant l'extraction des mesures via l'API cloud d'Emporia et leur insertion dans la base de données locale.

4. Emporia Energy, *Emporia Cloud API*. Site consulté en 2026, <https://www.emporiaenergy.com/>.

5. Jason Ertel, *Vuegraf*. Site consulté en mai 2025, <https://github.com/jertel/vuegraf>.

La configuration de Vuegraf permet de transférer automatiquement les données collectées par l'Emporia Vue vers la base de données InfluxDB. Elle s'effectue dans un fichier JSON où l'on renseigne les identifiants du compte Emporia (adresse courriel et mot de passe), ainsi que les paramètres de connexion à l'instance InfluxDB cible, l'URL du serveur, l'organisation, le bucket de stockage et le jeton d'accès. Une fois ce fichier créé, l'outil se lance en ligne de commande en précisant son emplacement : `python3 path/vuegraf/src/vuegraf.py vuegraf.json`. Vuegraf s'authentifie alors auprès de l'API Emporia, interroge périodiquement le service cloud pour récupérer les mesures de consommation, puis les insère dans la base InfluxDB locale. Ce flux automatisé, paramétré par le fichier JSON, garantit une synchronisation fiable entre les données de l'Emporia Vue Gen 2 et leur stockage local, prêtes pour des analyses ultérieures.

- **Données météorologiques**

Les données météorologiques relatives à la maison étudiée sont récupérées en temps réel depuis la plateforme en ligne Tomorrow.io, développée par une entreprise technologique fondée en 2016 et spécialisée dans la prévision météorologique. Cette plateforme fournit des données hyperlocales et personnalisées en s'appuyant sur un réseau de capteurs IoT, des données satellites, ainsi que des algorithmes d'intelligence artificielle. Son produit principal, la Weather API, permet d'accéder à des données météorologiques en temps réel, à des prévisions allant jusqu'à 14 jours et à des historiques couvrant plus de 20 ans. Dans le cadre de cette étude, les données ont été récupérées à l'aide d'un script Python, après obtention d'une clé d'accès via le site officiel de la plateforme, afin d'être stockées directement dans la base InfluxDB [93].

La Figure 3.3 illustre le processus de récupération des données météorologiques dédiées à la maison étudiée.

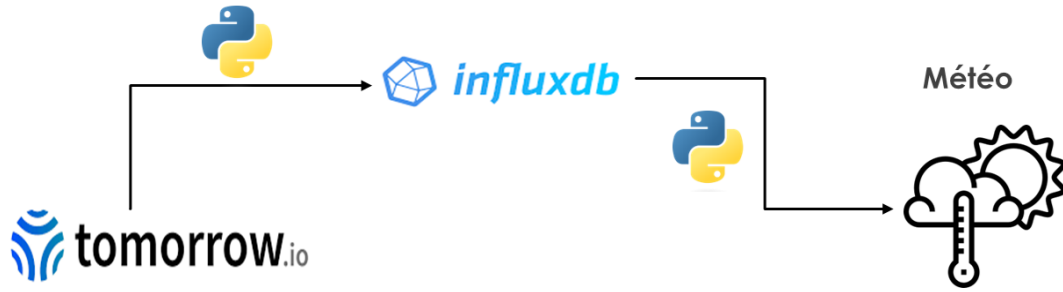


FIGURE 3.3 Système de récupération de données météorologiques

3.2 Acquisition et traitement des données

3.2.1 Intégration et alignement temporel

Pour assurer la cohérence entre les différentes sources de données, une résolution temporelle de référence de 15 minutes a été retenue pour l'intégration et l'analyse. Ce choix repose sur un compromis entre, d'une part, la granularité native fournie par le système de surveillance Emporia Vue, qui mesure la puissance électrique et constitue la variable principale de l'étude, et, d'autre part, la nécessité d'éviter la création artificielle d'une résolution plus fine que celle des autres sources disponibles, notamment les données météorologiques, elles aussi acquises à une cadence de 15 minutes. À partir des mesures de puissance instantanée, les séries ont ensuite été traitées par agrégation temporelle afin de calculer l'énergie consommée sur chaque intervalle de 15 minutes. Cette opération permet de préserver la tendance générale de la consommation tout en atténuant les fluctuations transitoires de courte durée. Soit P_i la puissance mesurée à l'instant i , exprimée en watts (W), et soit Δt le pas d'échantillonnage réel des données, exprimé en minutes. L'énergie élémentaire associée à chaque observation est calculée par :

$$E_i = P_i \times \frac{\Delta t}{60} \times \frac{1}{1000} \quad (3.1)$$

où E_i est exprimée en kilowattheures (kWh). Cette relation découle directement de la formule fondamentale reliant l'énergie à la puissance :

$$E = P \times \Delta t \quad (3.2)$$

avec conversion des unités, soit une division par 60 pour passer des minutes aux heures, puis par 1000 pour convertir les wattheures (Wh) en kilowattheures (kWh).

Pour un intervalle d'agrégation de durée T minutes, l'énergie consommée sur l'intervalle k est obtenue en sommant les énergies élémentaires correspondantes :

$$E_k^{(T)} = \sum_{i \in I_k} E_i \quad (3.3)$$

où I_k désigne l'ensemble des échantillons appartenant au k -ième intervalle de durée T .

Dans le cas particulier d'une agrégation sur 15 minutes, on obtient :

$$E_k^{(15)} = \sum_{i \in I_k} \left(P_i \times \frac{\Delta t}{60} \times \frac{1}{1000} \right) \quad (3.4)$$

Par ailleurs, les données météorologiques étaient déjà disponibles à une résolution temporelle de 15 minutes. Elles n'ont donc nécessité aucune transformation ni rééchantillonnage supplémentaire avant leur intégration au jeu de données final. D'un point de vue informatique, le processus de synchronisation a été mis en œuvre en Python, principalement à l'aide de la bibliothèque pandas. Chaque flux de données a d'abord été normalisé sur le plan temporel, notamment par l'harmonisation du format des horodatages et la prise en compte des changements d'heure saisonniers. Une jointure externe réalisée sur l'index temporel commun

a ensuite permis de construire un tableau de données unifié, dans lequel chaque ligne correspond à un intervalle de 15 minutes et chaque colonne à une variable provenant de l'une des sources intégrées. À l'issue de ce traitement, les valeurs manquantes, les doublons ainsi que les valeurs aberrantes, en particulier aux frontières des plages de disponibilité des différentes sources, ont été identifiés puis traités selon la procédure décrite dans la section suivante.

3.2.2 *Nettoyage des données*

Après l'étape d'intégration et d'alignement temporel, une procédure de nettoyage des données a été mise en œuvre afin d'améliorer la qualité des séries exploitées pour la modélisation. Cette étape avait pour objectif de corriger les défauts liés à l'acquisition, de traiter les discontinuités résiduelles et de limiter l'influence des artefacts de mesure susceptibles de perturber l'apprentissage et l'évaluation des modèles. Dans un premier temps, les observations présentant des incohérences techniques, telles que les doublons, les enregistrements partiellement renseignés ou certaines valeurs aberrantes attribuables au système d'acquisition, ont été identifiées. Ces outliers d'acquisition se distinguent des anomalies de consommation recherchées dans cette étude, dans la mesure où ils résultent d'imperfections instrumentales ou de transmission plutôt que d'un comportement énergétique inhabituel. Leur traitement préalable est donc nécessaire pour éviter toute confusion entre défaut de mesure et anomalie réelle. Dans un second temps, les valeurs manquantes ont été traitées selon la nature et la durée des interruptions observées. Les lacunes de courte durée ont été comblées par interpolation polynomiale de degré trois, méthode retenue pour sa capacité à préserver la continuité et la courbure locale des séries [7]. Lorsqu'une observation isolée précédait une coupure plus longue, une extrapolation linéaire a été utilisée afin de prolonger la tendance immédiate sans générer d'artefacts excessifs. En revanche, les portions trop incomplètes pour permettre une reconstruction fiable n'ont pas été reconstruites abusivement et ont été signalées comme telles

pour les traitements ultérieurs. Cette procédure de nettoyage a ainsi permis de constituer un jeu de données plus homogène, plus fiable et mieux adapté aux analyses menées dans la suite du mémoire. La Figure 3.4 et la Figure 3.5 illustrent respectivement l'évolution conjointe de la puissance consommée et de la température extérieure après prétraitement, ainsi que la relation entre ces deux variables.

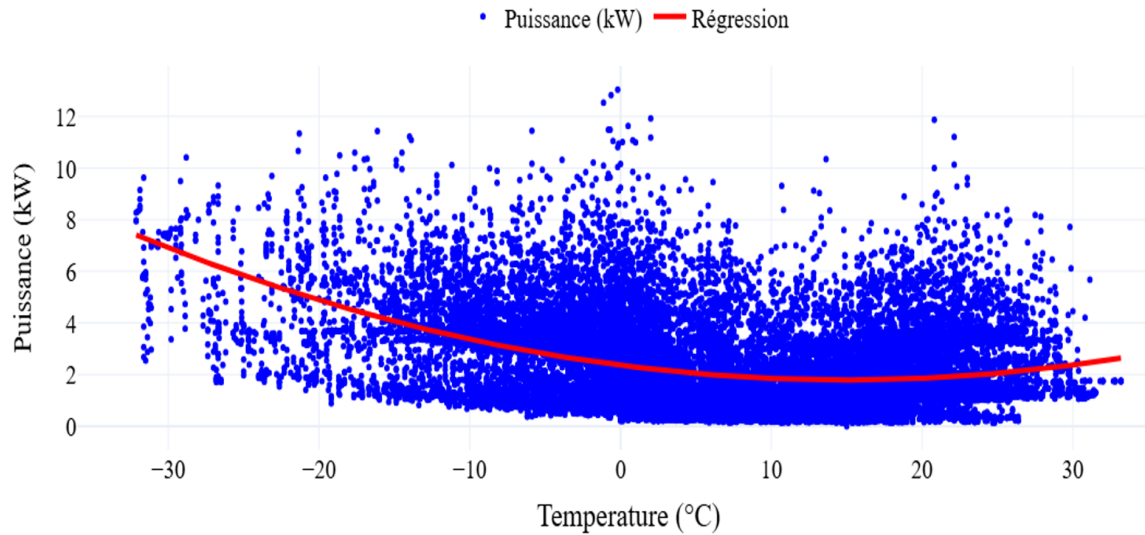


FIGURE 3.4 Relation entre la puissance consommée et la température extérieure avec régression

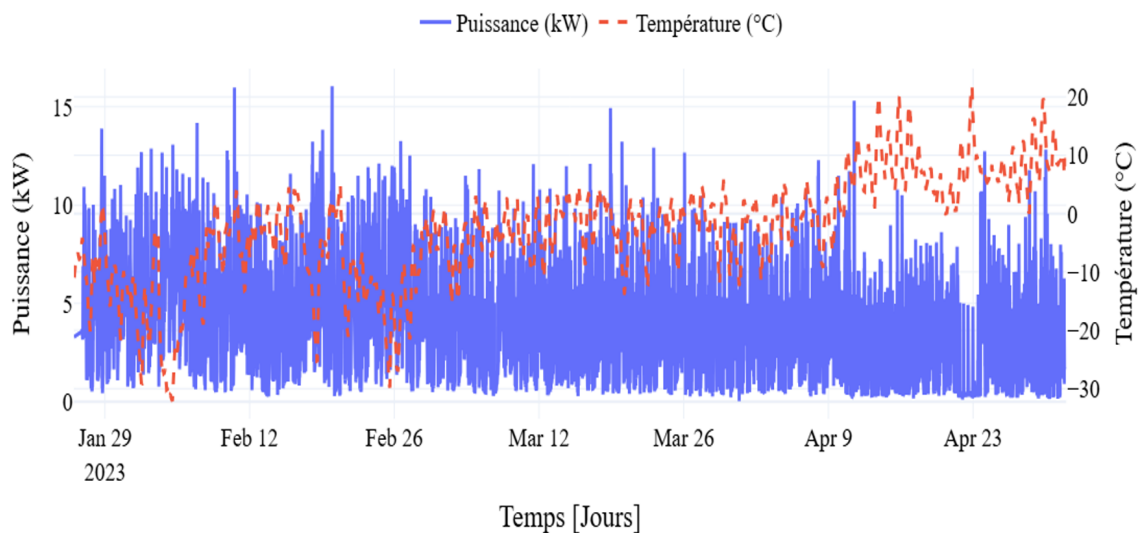


FIGURE 3.5 Évolution de la puissance consommée et de la température extérieure après prétraitement

3.3 Modélisation du comportement normal

3.3.1 Sélection des modèles

Le recours à plusieurs méthodes de modélisation s'explique par la volonté de représenter, dans le cadre comparatif, plusieurs familles d'approches couvrant un spectre large de complexité et de capacités. L'objectif n'est pas de privilégier une seule technique, mais d'évaluer des modèles reposant sur des principes de fonctionnement fondamentalement différents, afin d'identifier ceux qui sont les plus adaptés à la dynamique de la consommation électrique résidentielle. Les familles de modèles qui ont été retenues suivent une progression logique de complexité croissante, les modèles d'apprentissage automatique classique opérant point par point, les modèles séquentiels hybrides, et les modèles à prédiction vectorielle multi-horizon. Les modèles Support Vector Regression (SVR) et eXtreme Gradient Boosting (XGBoost) ont été retenus comme méthodes de référence en raison de leur efficacité éprouvée dans la modélisation de relations non linéaires complexes [33], fréquemment observées dans les profils de charge agrégés. Le SVR, fondé sur la théorie des machines à vecteurs de support, présente l'avantage de fournir une modélisation robuste de la relation entre la consommation et ses variables explicatives, y compris dans des contextes où cette relation est partiellement non linéaire. Son noyau RBF (Radial Basis Function) permet de projeter implicitement les données dans un espace de dimension supérieure, capturant ainsi des structures non accessibles par une régression linéaire [94]. De son côté, XGBoost, fondé sur le principe du gradient boosting, se distingue par sa robustesse, sa capacité de généralisation et son aptitude à capturer efficacement les interactions non linéaires entre variables [95]. Ces deux approches opèrent de manière point par point, à chaque pas de temps, le modèle prédit la consommation à partir des covariables disponibles à cet instant, sans modéliser explicitement les dépendances temporelles entre pas de temps successifs [96]. Elles constituent ainsi des références solides pour

évaluer si l'ajout d'une modélisation séquentielle apporte un gain significatif.

En complément, une architecture neuronale hybride a été introduite afin de combiner extraction de motifs locaux et modélisation temporelle. L'architecture CNN-LSTM a été retenue en raison de son caractère complémentaire, la composante convolutionnelle (Conv1D) permet d'extraire automatiquement les motifs locaux significatifs présents dans la série temporelle (pics de consommation, transitions), tandis que la composante LSTM prend en charge la modélisation des dépendances temporelles plus étendues. Dans cette étude, le modèle CNN-LSTM adopte une stratégie de prédiction vectorielle, à partir d'une fenêtre d'observation d'une journée [97]. Afin de compléter le panorama des approches, le modèle Time-series Dense Encoder (TiDE)⁶ a été retenu comme représentant des architectures à encodeur-décodeur entièrement denses. Contrairement aux approches récurrentes (LSTM) ou convolutionnelles (CNN), TiDE repose sur un empilement de couches denses (MLP) pour l'encodage et le décodage des séries temporelles. Cette architecture présente l'avantage théorique d'une plus grande simplicité structurelle tout en offrant la capacité de prédiction multi-horizon⁷. Son inclusion dans le cadre comparatif permet d'évaluer si une architecture dense pure, sans composante récurrente ni convolutionnelle, peut rivaliser avec les approches plus complexes pour la modélisation du comportement normal de consommation résidentielle. De plus, TiDE intègre nativement la gestion de covariables futures, ce qui permet d'exploiter les prévisions de température comme variables explicatives.

Il est important de préciser que, bien que les 4 modèles retenus soient initialement formulés comme des méthodes de régression supervisée, leur utilisation dans cette étude s'inscrit dans une logique de détection non supervisée fondée sur l'apprentissage du comportement normal. La phase de modélisation repose sur une régression supervisée entre les covariables

6. Unit8, *Time-series Dense Encoder (TiDE)*, https://unit8co.github.io/darts/generated_api/darts.models.forecasting.tide_model.html, consulté en juin 2026.

7. Unit8, *TiDEModel*, Documentation de la bibliothèque Darts, https://unit8co.github.io/darts/generated_api/darts.models.forecasting.tide_model.html, consulté en 2026.

et la consommation, mais la phase de détection est entièrement non supervisée, aucune étiquette d'anomalie n'est utilisée, ni pour l'entraînement, ni pour la calibration des seuils. Les modèles sont entraînés exclusivement sur des données supposées majoritairement saines, représentatives du fonctionnement normal du bâtiment. Le modèle calibré est ensuite utilisé sur des données nouvelles dans un cadre d'inférence, afin de produire des prédictions de la consommation attendue. Les écarts entre les valeurs observées et les valeurs prédites, les résidus sont alors exploités comme indicateurs d'anomalie. Cette stratégie est particulièrement adaptée au contexte résidentiel, où les anomalies réelles sont rares, hétérogènes et peu documentées, ce qui limite fortement la pertinence d'une approche supervisée classique fondée sur des étiquettes d'anomalies. Le choix de ne pas recourir à des méthodes non supervisées pures se justifie par plusieurs considérations. Premièrement, la stratégie basée sur les résidus d'un modèle prédictif fournit un signal d'anomalie physiquement interprétable, l'écart entre la consommation réelle et la consommation prédite a une signification énergétique directe. Deuxièmement, cette approche permet d'intégrer naturellement des covariables exogènes, notamment la température extérieure dans la modélisation, ce que les méthodes non supervisées pures ne permettent pas aussi aisément. Troisièmement, la séparation entre la phase de modélisation (prédiction) et la phase de détection (analyse des résidus) offre une modularité qui facilite le diagnostic et l'optimisation de chaque composante indépendamment. Afin d'évaluer rigoureusement les performances relatives de ces méthodes, un cadre expérimental comparatif commun est mis en place. Toutes les approches sont appliquées au même jeu de données, prétraité selon une procédure identique, avec les mêmes variables d'entrée, les mêmes résolutions temporelles et des fenêtres d'apprentissage et de test harmonisées. Les modèles sont entraînés sur une période représentative du fonctionnement normal, puis évalués sur une journée de test contenant une anomalie injectée de manière contrôlée. L'injection prend la forme d'un level-shift unidirectionnel, d'amplitude proportionnelle, débutant

à un instant fixe et se prolongeant sur une durée fixe. Ce scénario simule une surconsommation non expliquée par les conditions climatiques, telle qu'un comportement d'occupation anormal. Le protocole prévoit en outre une variation systématique de paramètres expérimentaux, notamment le seuil statistique de détection défini par la bande supérieure à différents percentiles. Les performances sont comparées à l'aide d'indicateurs adaptés à la détection d'anomalies, la précision, le rappel, le score F1, la spécificité, le délai de détection, ainsi que le temps de calcul total. Une attention particulière est accordée à la capacité des modèles à minimiser les faux positifs, condition essentielle dans un contexte opérationnel où chaque fausse alarme engendre des coûts d'intervention et érode la confiance des opérateurs envers le système de détection. L'ensemble de ce cadre méthodologique vise à comparer non seulement la capacité prédictive des modèles, mais aussi leur pertinence opérationnelle dans un contexte de détection d'anomalies de consommation électrique résidentielle.

3.3.2 *Extraction de caractéristiques*

L'extraction de caractéristiques constitue une étape déterminante dans la construction d'un système de détection d'anomalies, dans la mesure où elle conditionne directement la richesse informationnelle mise à disposition du modèle lors de l'apprentissage. Plutôt que de soumettre les données brutes telles quelles, cette phase vise à produire des variables transformées, plus expressives et plus exploitables, qui capturent explicitement les structures sous-jacentes de la consommation. Dans le cadre de ce travail, cette démarche est appliquée aux variables temporelles, dont la nature cyclique ne peut être correctement représentée sans un encodage adapté.

La consommation électrique résidentielle est en effet un phénomène intrinsèquement rythmé par le temps. Les occupants se lèvent, cuisinent, chauffent et dorment selon des routines qui se répètent de manière relativement stable d'un jour à l'autre et d'une semaine à

l'autre. Pour qu'un modèle de détection puisse identifier un comportement anormal, il doit d'abord être capable de représenter ces cycles de manière fidèle. À cette fin, un vecteur de position continue au sein de la semaine, noté t_{week} , est construit en couvrant l'ensemble des instants d'une semaine complète de sept jours [98]. Ce vecteur est ensuite répété et tronqué pour couvrir la durée totale de la série, puis décalé d'un offset correspondant à l'heure de début de la série, afin de garantir l'alignement correct entre la position dans la semaine et l'horodatage réel de chaque observation.

Une fois cette position établie, quatre variables sont construites par application de fonctions trigonométriques selon deux fréquences distinctes. Les variables t_1 et t_2 capturent le cycle journalier en réalisant sept oscillations complètes sur l'étendue d'une semaine, tandis que t_3 et t_4 capturent le cycle hebdomadaire en réalisant une oscillation unique sur l'ensemble de la semaine. Formellement, pour une période T et une position t , ces variables s'écrivent :

$$t_{\sin} = \sin\left(\frac{2\pi \cdot t}{T}\right), \quad t_{\cos} = \cos\left(\frac{2\pi \cdot t}{T}\right) \quad (3.5)$$

L'encodage en paires sinus-cosinus est privilégié afin de résoudre la discontinuité inhérente à toute représentation entière du temps. Une variable discrète représentant l'heure de la journée traite l'écart entre 23 h et 0 h comme étant de 23 unités, alors qu'il n'est en réalité que d'une unité. La projection sur le cercle trigonométrique garantit que des instants temporellement proches sont également proches dans l'espace des caractéristiques, quelle que soit la frontière de la journée ou de la semaine. Le Tableau 3-1 résume les quatre variables temporelles construites et leur interprétation.

Ces quatre variables temporelles cycliques, combinées aux variables climatiques synchronisées à l'étape précédente, constituent le socle du vecteur de caractéristiques à partir

TABLEAU 3-1 Variables temporelles cycliques construites par encodage trigonométrique

Variable	Fonction	Fréquence	Cycle capturé
t_1	sin	$7 \times 2\pi / T_{\text{semaine}}$	Journalier
t_2	cos	$7 \times 2\pi / T_{\text{semaine}}$	Journalier
t_3	sin	$2\pi / T_{\text{semaine}}$	Hebdomadaire
t_4	cos	$2\pi / T_{\text{semaine}}$	Hebdomadaire

duquel la ligne de base contextuelle de consommation sera modélisée.

3.4 Identification des anomalies

3.4.1 Calcul des résidus

En apprentissage automatique, un modèle entraîné sur des données historiques est capable d'estimer le comportement attendu de la variable cible pour de nouvelles observations. Dans le cadre de cette étude, cette estimation correspond à la consommation électrique considérée comme normale pour la résidence étudiée, en fonction du contexte climatique et calendaire associé à chaque instant. Les données réelles, quant à elles, traduisent le comportement effectivement observé du système, tel qu'il est mesuré par le dispositif de surveillance de charge. L'écart entre la valeur prédite par le modèle et la valeur réellement mesurée définit le résidu de prédiction. Ce résidu constitue le signal principal sur lequel repose le mécanisme de détection d'anomalies.

Soit y_t la valeur réelle de la consommation énergétique observée à l'instant t , et soit $\hat{y}_t$ la valeur prédite par le modèle à ce même instant. La prédiction est obtenue à partir d'un ensemble de variables explicatives x_t , selon la relation générale :

$$\hat{y}_t = f(x_t) \quad (3.6)$$

où $f(\cdot)$ désigne la fonction apprise par le modèle de régression. Le résidu de prédiction à l'instant t est alors défini comme la différence entre la mesure réelle et la valeur prédite :

$$r_t = y_t - \hat{y}_t \quad (3.7)$$

Cette quantité traduit l'écart instantané entre le comportement observé et le comportement normal estimé par le modèle. Lorsque r_t est proche de zéro, la prédiction est cohérente avec la mesure réelle, ce qui suggère un fonctionnement normal. En revanche, une valeur élevée en module indique un écart important susceptible de correspondre à une anomalie.

Dans le cadre de la détection, l'intérêt porte généralement non seulement sur le signe du résidu, mais aussi sur son amplitude. Pour cette raison, un score d'écart peut être défini à partir de la valeur absolue du résidu :

$$s_t = |r_t| = |y_t - \hat{y}_t| \quad (3.8)$$

Ce score permet de mesurer l'intensité de l'écart indépendamment de sa direction. Ainsi, une sous-estimation importante et une surestimation importante sont toutes deux interprétées comme des comportements potentiellement anormaux.

Dans notre cas d'étude, la variable cible correspond à l'énergie mesurée sur un intervalle de 15 minutes, y_t représente l'énergie réelle consommée durant cet intervalle, tandis que $\hat{y}_t$ représente l'énergie attendue prédite par le modèle pour ce même intervalle. Le résidu r_t représente donc l'écart de consommation entre le comportement observé et le comportement

normal estimé. L'analyse de la série temporelle des résidus constitue ensuite la base de la procédure de détection d'anomalies.

3.4.2 *Analyse statistique*

Le système de détection proposé dans le cadre de ce travail repose sur l'analyse statistique des résidus de prédiction, définis comme l'écart entre la consommation réellement observée et la consommation prédite par le modèle. L'approche adoptée s'inscrit dans un cadre d'inférence fréquentiste, le modèle produit une prédiction ponctuelle de la consommation à chaque pas de temps, et l'incertitude est quantifiée a posteriori par l'analyse statistique de la distribution des résidus. Cependant, dans la pratique, les systèmes d'apprentissage peuvent disposer de très peu d'observations, notamment lors de la phase initiale. Un modèle entraîné sur un historique insuffisant produira des prédictions de faible qualité, des résidus de grande amplitude et, par conséquent, des intervalles de confiance artificiellement larges qui compromettent toute tentative de détection fiable. C'est pourquoi la méthode proposée impose une condition préalable, la détection ne démarre pas avant que le système ait accumulé un nombre de jours d'historique suffisant pour garantir une prédiction de qualité. Ce nombre est déterminé empiriquement lors de la phase de conception, en évaluant plusieurs longueurs de fenêtre d'entraînement et en retenant celle qui produit des résidus de faible amplitude avec une moyenne proche de zéro, ce qui traduit une prédiction non biaisée et précise.

Lors du déploiement, le système observe le comportement de consommation pendant cette période sans émettre aucune alerte, le temps de constituer l'historique minimal requis. Chaque bâtiment présente un profil de consommation qui lui est propre, dépendant de sa surface, de son isolation, de ses équipements et des habitudes de ses occupants, et cette phase d'initialisation permet au modèle de s'adapter à ces caractéristiques locales avant toute détection. Une fois la phase d'initialisation terminée, le système entre en inférence opérationnelle

selon un mode de prédiction point par point en temps réel. À chaque nouveau pas de temps, le modèle est interrogé avec les covariables mesurées à cet instant, il produit une prédiction de la consommation attendue, et le résidu est calculé immédiatement par comparaison avec la valeur réellement observée.

Dans la phase de détection, le système calcule les intervalles de confiance des résidus dans le but de définir les seuils à partir desquels un écart est considéré comme potentiellement anormal. Par ailleurs, le modèle est réentraîné périodiquement sur les jours les plus récents, avec des hyperparamètres réoptimisés, afin de maintenir une qualité d'inférence stable face à l'évolution des conditions de fonctionnement du bâtiment. À chaque nouveau cycle d'entraînement, la distribution des résidus est recalculée selon l'approche fréquentiste adoptée, fondée sur les fréquences observées des écarts, et les intervalles de confiance sont mis à jour en conséquence. Cet ajustement dynamique permet de capter la variabilité des profils de consommation au fil du temps, ce qui réduit significativement les faux positifs qui surviendraient inévitablement avec un seuil fixe. En effet, le comportement du bâtiment évolue progressivement, changement de saison, modification des habitudes d'occupation, ajout d'équipements, et le modèle intègre naturellement ces changements dans un délai correspondant à la fenêtre optimale. Les résidus reflètent alors la nouvelle normalité, leur moyenne reste proche de zéro, et les seuils se repositionnent en conséquence. Ce mécanisme de fenêtre glissante garantit que les intervalles de confiance ne sont pas des bornes figées, mais des enveloppes adaptatives qui évoluent avec le profil de consommation du bâtiment.

Cependant, les intervalles de confiance sont construits à partir des quantiles empiriques de la distribution des résidus, sans supposer une distribution gaussienne. Ce choix est cohérent avec le cadre fréquentiste adopté, les seuils sont estimés directement à partir des fréquences observées dans les données, sans recourir à un modèle paramétrique. Dans l'optique

de détecter une consommation excessive, seule la bande supérieure est retenue dans l'analyse des résidus. Plusieurs niveaux de cette bande sont testés expérimentalement afin de définir le compromis optimal entre la minimisation des faux positifs et la rapidité de la détection. Un seuil bas favorise une détection rapide mais augmente le risque de fausses alertes, tandis qu'un seuil élevé réduit les fausses alertes mais retarde le moment de détection. L'exploration systématique de ces niveaux permet d'identifier, pour chaque modèle, la configuration qui offre le meilleur équilibre entre ces deux exigences. La Figure 3.6 illustre l'organigramme décrivant les différentes étapes du système de détection des anomalies.

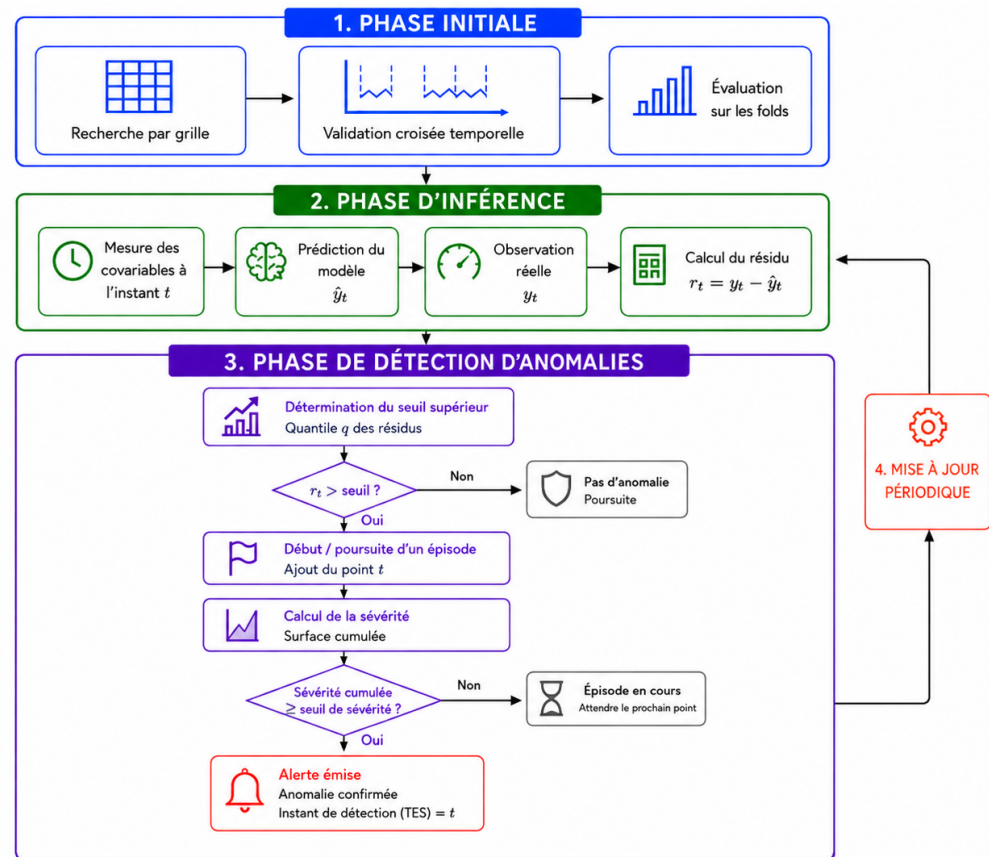


FIGURE 3.6 Organigramme du système de détection des anomalies

3.4.3 Règles de détection et paramétrisation

Le profil de consommation électrique résidentielle résulte principalement de l'interaction entre deux familles de facteurs, les conditions météorologiques, en particulier la température extérieure qui influence directement la demande de chauffage, et les habitudes d'occupation, qui structurent les cycles quotidiens d'utilisation des équipements domestiques. La combinaison de ces facteurs engendre, pour chaque bâtiment, une dynamique de consommation spécifique qu'il convient de modéliser avec précision afin d'établir une représentation fiable du comportement normal. Plus le modèle reproduit fidèlement cette dynamique, plus les résidus de prédiction demeurent faibles et centrés autour de zéro, ce qui améliore la séparation entre fluctuations normales et comportements véritablement anormaux. La qualité des résidus constitue ainsi un élément central du processus de détection. En effet, un modèle dont les erreurs de prédiction sont élevées et dispersées conduit à des bandes de tolérance larges, au sein desquelles les anomalies réelles risquent d'être masquées par le bruit de modélisation. À l'inverse, un modèle produisant des résidus faibles et peu biaisés concentre la distribution des écarts dans une bande plus étroite, de sorte qu'une consommation anormale se traduit par une sortie plus nette hors du domaine de variabilité admissible. Dans cette perspective, la réduction des faux positifs ne repose pas uniquement sur le réglage des seuils de détection, mais d'abord sur la capacité du modèle à représenter avec précision le régime normal du système.

C'est dans cette optique que la stratégie d'optimisation des hyperparamètres vise prioritairement à améliorer la fidélité de la prédiction. Le paramètre le plus structurant de cette optimisation est la longueur de la fenêtre d'apprentissage. Une fenêtre trop courte ne fournit pas suffisamment d'information pour permettre au modèle d'apprendre les régularités du profil de consommation, ce qui dégrade la qualité prédictive et amplifie les résidus. À l'inverse, une fenêtre trop longue peut intégrer des observations devenues partiellement ob-

solètes, notamment en période de transition saisonnière, et introduire ainsi un décalage entre le comportement appris et le comportement courant. En outre, au-delà d'un certain volume d'information, l'ajout de jours supplémentaires n'améliore plus significativement la précision, tout en augmentant le coût computationnel. La fenêtre optimale correspond donc au meilleur compromis entre richesse informative, pertinence temporelle et efficacité de calcul.

En régime normal, un modèle correctement calibré se caractérise par des résidus qui demeurent, sauf dépassements ponctuels rares et de faible ampleur, à l'intérieur des bandes de tolérance définies à partir de leur distribution. Lorsqu'une anomalie survient, les résidus augmentent au contraire de manière plus marquée et plus persistante, produisant un contraste plus net avec les fluctuations habituelles du modèle. Ce contraste constitue le fondement de la règle de détection retenue dans cette étude.

Le processus de détection débute par l'identification des excursions contiguës au-dessus de la bande supérieure. À chaque pas de temps de la phase d'inférence, le résidu courant est comparé au seuil associé à la variabilité normale. Les instants consécutifs pour lesquels ce seuil est dépassé sont ensuite regroupés en segments, chacun constituant une excursion candidate. Cette opération est réalisée au moyen d'un étiquetage des composantes connexes appliqué à un masque binaire défini sur la série des résidus, une valeur de 1 est attribuée lorsqu'un résidu dépasse la bande supérieure, et une valeur de 0 sinon. Les séquences contiguës de dépassement sont alors extraites et identifiées individuellement. Pour chaque excursion, une mesure de sévérité est calculée à partir de la surface comprise entre la courbe des résidus et la bande supérieure. Cette surface est estimée numériquement par la méthode des trapèzes, qui approxime l'intégrale de l'excès de résidu sur toute la durée du dépassement. Dans le cas particulier d'une excursion réduite à un seul pas de temps, la sévérité est approchée par l'aire d'un triangle. Une telle mesure permet de combiner en une seule grandeur l'intensité et la

durée du dépassement. Ainsi, une excursion brève mais de forte amplitude peut présenter une sévérité comparable à celle d'un dépassement plus modéré mais plus prolongé.

La règle de décision repose sur une comparaison entre les excursions observées dans les résidus propres et celles issues des résidus contaminés. Les excursions extraites des résidus propres, c'est-à-dire calculés en l'absence d'anomalie injectée, représentent les fluctuations naturelles produites par le modèle en conditions normales. À partir de la distribution empirique des surfaces de ces excursions normales, un quantile est retenu comme seuil de référence de sévérité. Ce choix permet de capter le niveau de sévérité atteint par les fluctuations naturelles les plus extrêmes, tout en écartant les valeurs aberrantes éventuelles susceptibles de fausser le calibrage. Pour chaque excursion détectée dans les résidus contaminés, une première condition est vérifiée, sa surface totale doit excéder ce seuil de référence ; dans le cas contraire, l'excursion est considérée comme compatible avec le régime normal et ne déclenche aucune alerte. Pour les excursions qui satisfont cette condition, la surface cumulée est ensuite évaluée pas par pas depuis le début de l'excursion, et l'alerte est émise à l'instant précis où cette surface cumulée dépasse le seuil de référence. Cet instant correspond au Time-to-Equal-Severity (TES), qui constitue le moment effectif de détection. Dans le cas particulier d'une excursion réduite à un seul pas de temps, la détection est immédiate dès lors que sa surface excède le seuil. Tant que ces conditions ne sont pas satisfaites, aucune alerte n'est émise, même en présence de dépassements ponctuels de la bande supérieure. Ce mécanisme permet ainsi d'écarter les excursions brèves ou peu significatives et de réserver les alertes aux déviations dont la sévérité cumulée excède celle des fluctuations naturelles les plus extrêmes du régime normal. Un avantage important de cette approche réside dans le fait que le seuil de sévérité n'est pas fixé de manière arbitraire par l'opérateur. Il est au contraire calibré automatiquement à chaque cycle à partir des excursions observées sur les résidus propres, ce qui confère au système un caractère adaptatif. Cette adaptativité est renforcée par l'usage d'une

fenêtre glissante, grâce à laquelle le modèle est réentraîné périodiquement sur les N jours les plus récents. Les bandes de tolérance et les niveaux de sévérité de référence évoluent ainsi avec le profil de consommation du bâtiment, ce qui permet de prendre en compte les changements progressifs liés à la saison, aux habitudes d'occupation ou à l'ajout d'équipements.

L'analyse expérimentale porte sur deux paramètres principaux, la longueur de la fenêtre d'apprentissage, qui conditionne la qualité des résidus, et le niveau de la bande supérieure, qui détermine à la fois les dépassements retenus et l'amplitude des surfaces calculées. Pour chaque combinaison de paramètres, les performances sont évaluées à l'aide de plusieurs indicateurs, à savoir la précision, le rappel, le score F_1 , la spécificité, le délai de détection et le temps de calcul. Le score F_1 est retenu comme critère principal de sélection, dans la mesure où il équilibre les conséquences des faux positifs et des non-détections. Le délai de détection intervient comme critère secondaire, entre deux configurations présentant des performances similaires en termes de F_1 , la préférence est accordée à celle qui permet une détection plus précoce. Cette hiérarchie traduit les exigences du contexte résidentiel étudié, dans lequel la fiabilité des alertes constitue la priorité principale, tout en maintenant un niveau de réactivité compatible avec une intervention corrective utile.

Chapitre 4 – Évaluation et Résultats

4.1 Introduction

Ce chapitre présente l'évaluation expérimentale du système de détection d'anomalies proposé, ainsi que l'analyse comparative des résultats obtenus avec les différents modèles retenus. L'objectif principal n'est pas seulement de mesurer la qualité prédictive des modèles, mais surtout d'examiner leur aptitude à soutenir une détection fiable, rapide et robuste des surconsommations anormales dans un contexte de bâtiment résidentiel. Dans cette perspective, l'évaluation est centrée sur le comportement des résidus de prédiction et sur la capacité des règles statistiques définies au chapitre précédent à transformer ces écarts en alertes pertinentes.

Afin d'assurer une comparaison rigoureuse, l'ensemble des expériences a été conduit dans un cadre harmonisé, où les 4 modèles retenus, SVR, XGBoost, CNN-LSTM et TiDE, sont évalués sur les mêmes données, avec les mêmes variables d'entrée, les mêmes fenêtres d'apprentissage et les mêmes scénarios d'anomalies injectées. Cette standardisation permet d'attribuer les différences de performance observées aux propriétés propres des modèles et non à des variations du protocole expérimental. Les résultats sont ensuite analysés selon plusieurs critères complémentaires, notamment la précision, le rappel, le score F1, la spécificité, le délai de détection et le coût computationnel, de manière à apprécier simultanément la fiabilité statistique et la pertinence opérationnelle des méthodes étudiées. L'évaluation s'inscrit par ailleurs dans la logique de déploiement définie pour ce travail. Le système repose sur une phase initiale d'accumulation d'historique sans émission d'alerte, suivie d'un fonctionnement

en fenêtre glissante avec réentraînement périodique et mise à jour dynamique des seuils issus de la distribution des résidus. Dans ce cadre, l'analyse des résultats vise à identifier les configurations les plus aptes à minimiser les faux positifs tout en conservant une réaction suffisamment rapide face à une hausse anormale de consommation.

4.1.1 Définition des scénarios et vérité terrain

En apprentissage non supervisé, l'évaluation des performances constitue un défi méthodologique majeur, puisque les anomalies réelles sont rarement disponibles sous forme de données explicitement étiquetées. Dans le contexte résidentiel étudié, il n'existe pas de journal d'annotation tenu par les occupants permettant d'indiquer précisément les instants où un comportement anormal, tel qu'une fenêtre laissée ouverte en période de chauffage, s'est effectivement produit. En conséquence, la vérité terrain ne peut pas être construite à partir de labels réels exhaustifs. Dans ce mémoire, la validation des modèles repose donc sur une stratégie d'injection synthétique d'anomalies dans des plages de données considérées comme saines. Cette approche consiste à partir de profils de consommation normaux, puis à y insérer de manière contrôlée une perturbation artificielle représentant une surconsommation anormale. Le scénario retenu dans le cadre expérimental correspond à une anomalie réversible de type décalage positif, appelée *level-shift up* en anglais. Dans notre cas, cette anomalie est représentée par une augmentation temporaire de la consommation par rapport à la consommation attendue en fonctionnement normal. Elle permet ainsi de simuler une hausse prolongée de la demande énergétique, cohérente avec un comportement thermique anormal, par exemple une ouverture de fenêtre en période de chauffage. Contrairement à une panne permanente, ce type d'anomalie est réversible, car la consommation revient à son niveau normal lorsque la cause de la surconsommation disparaît. La Figure 4.1 illustre ce principe en comparant la consommation attendue en régime normal avec la consommation observée lors de l'anomalie

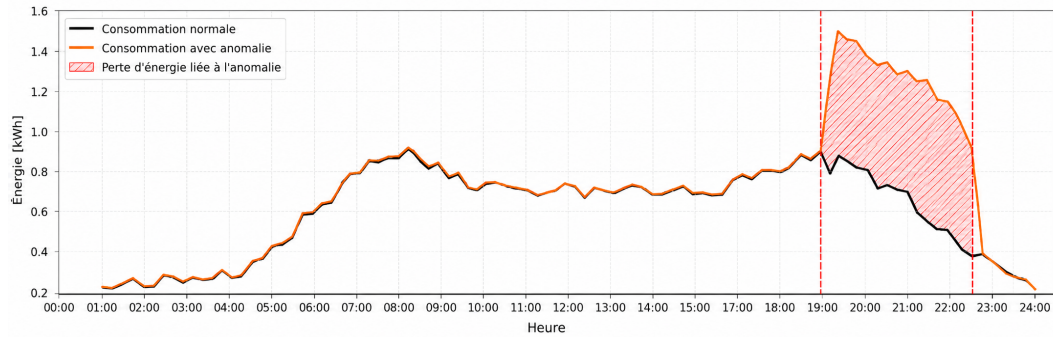


FIGURE 4.1 Exemple d'anomalie injectée de type décalage positif

injectée. Elle met en évidence l'écart introduit artificiellement, utilisé comme vérité terrain pour évaluer la capacité des modèles à détecter la surconsommation anormale.

La vérité terrain est alors définie directement à partir de cette injection, tous les instants appartenant à l'intervalle artificiellement perturbé sont étiquetés comme anomalies, tandis que les autres instants conservent l'étiquette normale. Cette construction fournit une référence binaire claire pour mesurer les performances des modèles en détection, notamment en termes de précision, de rappel, de score F1, de spécificité et de délai de détection. Elle permet aussi d'évaluer explicitement la réactivité du système, c'est-à-dire sa capacité à générer une alerte rapidement après le début de l'événement injecté. Le recours à des anomalies synthétiques se justifie par trois raisons principales. Premièrement, il garantit une maîtrise complète du moment d'apparition, de l'intensité et de la durée de l'anomalie. Deuxièmement, il assure un cadre d'évaluation équitable entre les différents modèles, puisque tous sont testés sur les mêmes signaux et les mêmes perturbations. Troisièmement, il permet d'estimer séparément la sensibilité de détection sur données perturbées et le taux de faux positifs sur données saines. Ainsi, dans ce travail, la vérité terrain n'est pas issue d'une labellisation manuelle par les occupants, mais d'un protocole expérimental contrôlé fondé sur l'injection artificielle d'anomalies dans des données de référence saines.

4.2 Cas d'étude et implémentation expérimentale

4.2.1 Environnement de test et profil du bâtiment

Le cas d'étude porte sur une maison unifamiliale de type bungalow située à Trois-Rivières, au Québec, Canada [22]. Cette résidence individuelle de faible hauteur est représentative d'une partie importante du parc résidentiel québécois. Elle est implantée dans un contexte climatique froid, caractérisé par des hivers longs et rigoureux, durant lesquels l'écart entre la température intérieure de confort et la température extérieure entraîne une hausse importante des besoins de chauffage. Cette réalité rend l'analyse de l'enveloppe thermique particulièrement importante, puisque les pertes de chaleur influencent directement la consommation énergétique en période hivernale. Cependant, Trois-Rivières, située dans la vallée du fleuve Saint-Laurent, appartient à la zone climatique 2 selon la classification utilisée pour les produits de fenestration certifiés ENERGY STAR¹. Les performances énergétiques du bâtiment dépendent donc principalement de la qualité de l'isolation de la toiture, des murs, des fondations, des planchers et des ouvertures. Ainsi, cette maison constitue un cas d'étude pertinent pour évaluer la détection d'anomalies énergétiques en contexte résidentiel en climat froid, notamment lorsqu'il s'agit de distinguer une hausse normale de consommation liée aux conditions climatiques d'une surconsommation anormale causée par une perte thermique temporaire ou par une modification inhabituelle des usages.

4.2.2 Environnement logiciel et ressources Open Source

L'implémentation du cadre comparatif a été réalisée sous Python, en s'appuyant sur un ensemble de bibliothèques Open Source dédiées au traitement des données, à la modélisation et à la visualisation scientifique. Les bibliothèques pandas [99], et NumPy ont été utilisées

1. Hydro-Québec, *Comment choisir une fenêtre certifiée ENERGY STAR*, <https://www.hydroquebec.com/residentiel/mieux-consommer/conseils/fenetres-chauffage-climatisation/choisir-fenetre.html>, consulté en avril 2026.

pour l'intégration [100], l'alignement temporel et la transformation des séries de consommation et des variables météorologiques. scikit-learn a servi à la normalisation des données [101], à l'entraînement du modèle SVR et au calcul des métriques d'évaluation, tandis que xgboost a été utilisé pour le modèle XGBoost. Les modèles d'apprentissage profond ont été implémentés à l'aide de TensorFlow/Keras [102], et de la bibliothèque Darts. Cette dernière joue un rôle central dans le traitement des séries temporelles, car elle permet de manipuler les données sous forme d'objets temporels, d'intégrer des covariables contextuelles telles que la température extérieure et les variables calendaires, et de mettre en œuvre des modèles multi-horizons comme TiDE. L'utilisation de ces outils contribue à rendre les expériences plus reproductibles et à assurer une comparaison cohérente entre les différentes méthodes. Enfin, les bibliothèques Matplotlib et Seaborn ont été utilisées pour produire les figures d'analyse, notamment les courbes de convergence, les prédictions, les résidus et les comparaisons de performance. L'ensemble de cette pile technologique garantit un environnement expérimental transparent, reproductible et extensible.

4.3 Scénarios de test

4.3.1 Anomalies réversibles

Les anomalies réversibles correspondent à des perturbations temporaires du fonctionnement thermique du bâtiment, qui provoquent une surconsommation transitoire sans traduire une modification permanente du profil énergétique. Dans le cadre résidentiel, ce type de situation peut notamment être associé à une perte de chaleur ponctuelle, par exemple lorsqu'une fenêtre est mal fermée ou laissée ouverte pendant une certaine durée en période hivernale. L'intérêt de ces anomalies est qu'elles reproduisent des comportements plausibles d'usage, tout en restant limitées dans le temps, ce qui en fait un cas d'étude particulièrement pertinent pour évaluer la réactivité d'un système de détection. Afin d'analyser ce type d'événement

de manière rigoureuse, les scénarios de test sont définis en comparant isolément un profil de fonctionnement normal à une version perturbée de ce même profil. Autrement dit, l'évaluation repose sur une base saine servant de référence, dans laquelle est introduit un comportement anormal ciblé, de sorte que toute différence observée dans la réponse du système puisse être attribuée à l'anomalie étudiée et non à la variabilité naturelle de la consommation. Cette démarche permet de construire un cadre expérimental contrôlé, particulièrement important dans un contexte non supervisé où les anomalies réelles annotées sont rarement disponibles. Dans ce mémoire, les anomalies réversibles sont principalement reproduites par injection synthétique dans des plages de données saines. Ce choix méthodologique s'explique par l'absence de journaux d'occupation ou d'annotations terrain permettant d'identifier avec précision, dans les données réelles, la survenue d'événements tels qu'une perte de chaleur temporaire. L'injection artificielle permet au contraire de maîtriser explicitement le début, la durée et l'intensité de la perturbation, et donc d'établir une vérité terrain fiable pour l'évaluation. Le scénario retenu correspond à une anomalie de type level-shift up, modélisée par une augmentation de la consommation. Cette perturbation vise à représenter une hausse soutenue de la demande de chauffage, cohérente avec une anomalie thermique réversible.

Ainsi, pour chaque scénario, deux situations sont opposées, d'une part, le fonctionnement normal, utilisé comme référence de comportement sain, et d'autre part, le fonctionnement perturbé par une anomalie réversible ciblée. La vérité terrain est alors définie directement par la plage temporelle d'injection, les instants appartenant à cette plage sont considérés comme anormaux, tandis que les autres instants correspondent au régime normal. Une telle construction permet d'évaluer la capacité des modèles à détecter rapidement une dérive temporaire de consommation, tout en quantifiant leur aptitude à éviter les faux positifs sur les données saines.

4.4 Entraînement des modèles

La performance du système de détection dépend en grande partie d'une étape clé, à savoir le choix des hyperparamètres et de la durée de la fenêtre d'apprentissage lors de l'entraînement. Dans le contexte des séries temporelles, cette étape doit impérativement respecter l'ordre chronologique des observations afin d'éviter toute fuite d'information entre le passé et le futur. À cet effet, une procédure de recherche par grille (grid search) combinée à une validation croisée temporelle par fenêtre glissante a été retenue. Cette approche est cohérente avec les recommandations méthodologiques couramment formulées pour l'évaluation des modèles appliqués aux séries temporelles, selon lesquelles les séparations aléatoires de type k-fold sont inadaptées, car elles peuvent conduire à entraîner le modèle sur des données futures et à l'évaluer sur des données passées². Contrairement à la validation croisée classique, la validation par fenêtre glissante préserve l'intégrité temporelle de la série. Son principe consiste à utiliser une fenêtre d'apprentissage de taille N jours, suivie d'une fenêtre de validation de taille V jours. Après chaque itération, les deux fenêtres sont déplacées de S jours, puis l'opération est répétée jusqu'à épuisement des données disponibles. Chaque déplacement correspond ainsi à un sous-ensemble d'évaluation. Cette procédure permet d'obtenir plusieurs estimations successives de la performance du modèle dans des conditions proches d'un usage réel, où la prédiction est toujours effectuée à partir des seules données passées. Elle permet également de calculer des statistiques agrégées, telles que la moyenne et l'écart-type des performances, afin d'évaluer non seulement la qualité moyenne du modèle, mais aussi sa robustesse dans le temps. Dans cette étude, une période hivernale de 80 jours a été retenue afin de capturer les pics de consommation caractéristiques du contexte québécois, où le chauffage électrique engendre une demande énergétique particulièrement élevée durant la

2. Nahla Davies, *5 Ways to Use Cross-Validation to Improve Time Series Models*, <https://machinelearningmastery.com/5-ways-to-use-cross-validation-to-improve-time-series-models/>, consulté en avril 2026.

saison froide [33]. La recherche de la configuration optimale est réalisée au moyen d'un *grid search* exhaustif. Pour chaque modèle, plusieurs longueurs de fenêtre d'apprentissage sont examinées, soit $N \in \{20, 30, 40, 50, 60, 70, 80\}$ jours, en combinaison avec plusieurs jeux d'hyperparamètres propres au modèle considéré. Les plages explorées sont récapitulées dans le Tableau 4-1, chaque plage étant discrétisée en 20 configurations différentes.

TABLEAU 4-1 Plages d'hyperparamètres considérées lors de l'optimisation des modèles

Modèle	Hyperparamètre	Plage explorée
SVR	C	$\{0,01, \dots, 100\}$
	ϵ	$\{0,001, \dots, 0,2\}$
	γ	$\{\text{scale; auto; } 0,0001, \dots, 0,1\}$
XGBoost	$n_estimators$	$\{50, \dots, 800\}$
	$learning_rate$	$\{0,001, \dots, 0,2\}$
	max_depth	$\{2, \dots, 10\}$
	min_child_weight	$\{1, \dots, 20\}$
	$subsample$	$\{0,5, \dots, 1,0\}$
	$colsample_bytree$	$\{0,5, \dots, 1,0\}$
	reg_lambda	$\{0, \dots, 10\}$
CNN-LSTM	$filters$	$\{16, \dots, 128\}$
	$kernel_size$	$\{2, \dots, 7\}$
	$lstm_units$	$\{16, \dots, 128\}$
	$dropout$	$\{0, \dots, 0,5\}$
	$learning_rate$	$\{0,0001, \dots, 0,01\}$
	$batch_size$	$\{16; 32; 64; 128\}$
TiDE	$temporal_width_past$	$\{2, \dots, 8\}$
	$temporal_width_future$	$\{2, \dots, 8\}$
	$num_encoder_layers$	$\{1, \dots, 4\}$
	$hidden_size$	$\{32, \dots, 256\}$
	$temporal_hidden_size_past$	$\{32, \dots, 256\}$
	$decoder_output_dim$	$\{8, \dots, 64\}$
	$num_decoder_layers$	$\{1, \dots, 4\}$
	$temporal_hidden_size_future$	$\{32, \dots, 256\}$
	$temporal_decoder_hidden$	$\{16, \dots, 128\}$

Modèle	Hyperparamètre	Plage explorée
	<i>dropout</i>	$\{0, \dots, 0,5\}$
	<i>n_epochs</i>	$\{50, \dots, 200\}$
	<i>learning_rate</i>	$\{0,0001, \dots, 0,01\}$

Chaque jeu d’hyperparamètres est évalué sur l’ensemble des fenêtres glissantes définies sur la période d’étude. La taille du jeu de validation correspond à 15 % de celle de la fenêtre d’apprentissage, tandis que le décalage S entre deux sous-ensembles successifs est fixé à trois jours. Cette démarche permet ainsi d’explorer de manière structurée l’influence conjointe de l’historique d’apprentissage et du paramétrage interne du modèle sur la qualité prédictive obtenue. Les résultats issus de cette procédure sont ensuite synthétisés à l’aide d’indicateurs statistiques calculés sur l’ensemble des sous-ensembles. Les configurations reposant sur un nombre trop faible de sous-ensemble sont écartées de l’analyse comparative, car elles ne permettent pas d’évaluer de manière fiable la stabilité temporelle du modèle. Ainsi, l’objectif n’est pas seulement d’identifier la configuration présentant la meilleure performance moyenne, mais également de s’assurer que cette performance est reproductible sur plusieurs segments de la série. La sélection des hyperparamètres de chaque modèle repose sur les performances obtenues au cours des différentes fenêtres de validation et sur plusieurs critères complémentaires. Une fois la configuration optimale déterminée, celle-ci est fixée puis appliquée à une période de test chronologiquement postérieure, réservée à l’évaluation finale des performances. Parmi les critères retenus, le coefficient de détermination moyen R^2 permet d’évaluer la capacité du modèle à expliquer la variabilité de la consommation électrique observée. Il est défini par l’équation (4.1) :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.1)$$

où y_i représente la valeur réelle observée, $\hat{y}_i$ la valeur prédite par le modèle et $\bar{y}$ la moyenne des valeurs observées. Le terme $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ correspond à la somme des carrés des erreurs de prédiction, tandis que $\sum_{i=1}^n (y_i - \bar{y})^2$ représente la variabilité totale des données. Ainsi, le coefficient R^2 exprime la proportion de la variance des observations expliquée par le modèle. Plus sa valeur est proche de 1, plus le modèle reproduit fidèlement les variations de la consommation électrique. Dans le cadre de cette étude, le R^2 moyen calculé sur l'ensemble des sous-ensembles est utilisé pour évaluer la qualité moyenne d'ajustement de chaque configuration. D'autre part, l'écart-type du R^2 est pris en compte comme indicateur de stabilité entre les sous-ensembles. Une faible variabilité est recherchée afin de privilégier les modèles dont les performances demeurent constantes au cours du temps. Cette exigence de stabilité est particulièrement importante dans une perspective de déploiement opérationnel, où un modèle légèrement moins performant en moyenne, mais plus régulier, peut être préférable à un modèle instable. Par ailleurs, la taille de la fenêtre d'apprentissage N constitue également un critère important. À performance comparable, une valeur plus faible de N est privilégiée, car elle réduit le temps d'entraînement, comme le montre la Figure 4.2, qui présente l'augmentation du temps moyen d'exécution par sous-ensemble lorsque N augmente.

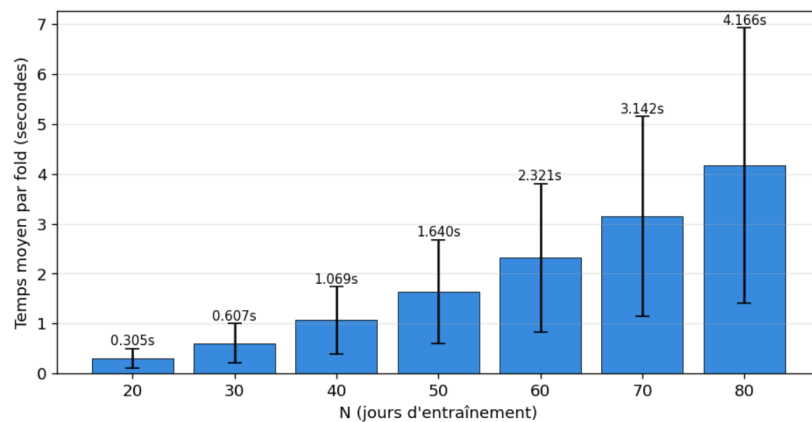


FIGURE 4.2 Temps d'exécution moyen du modèle SVR

À l'inverse, une fenêtre trop longue peut intégrer des données devenues moins représentatives, notamment en raison des variations saisonnières ou des changements dans les habitudes de consommation. Enfin, le nombre de sous-ensembles effectivement disponibles pour l'évaluation est également pris en considération. Ce critère permet de favoriser les configurations reposant sur un nombre suffisant d'itérations d'évaluation, ce qui renforce la fiabilité statistique des estimations obtenues. Ainsi, la procédure d'entraînement retenue ne vise pas uniquement à maximiser la précision prédictive, mais également à concilier la performance moyenne, la stabilité temporelle et l'efficacité opérationnelle. Une telle démarche permet d'identifier, pour chaque modèle, une configuration d'apprentissage à la fois robuste, réaliste et compatible avec les contraintes propres à la prévision de la consommation électrique résidentielle. Les architectures des modèles ainsi que les hyperparamètres retenus sont présentés ci-après.

4.4.1 *Support Vector Regression (SVR)*

Le modèle SVR retenu dans cette étude utilise un noyau gaussien (*Radial Basis Function*, RBF), particulièrement adapté à la modélisation de relations non linéaires entre les variables explicatives et la consommation électrique. Ce choix se justifie par le fait que la demande énergétique résidentielle dépend de manière complexe de plusieurs facteurs, notamment la température extérieure, les effets d'horaire et les habitudes d'occupation. Le noyau gaussien permet ainsi de mieux représenter ces non-linéarités qu'un modèle strictement linéaire, tout en conservant une bonne capacité de généralisation [94]. Les hyperparamètres du modèle ainsi que la procédure d'entraînement retenue sont présentés dans le Tableau 4-2. Contrairement aux modèles d'apprentissage profond, le SVR ne repose pas sur une notion explicite d'*epochs* ou de passes successives visibles sur les données. Son apprentissage résulte de la résolution d'un problème d'optimisation convexe visant à déterminer une fonction de

TABLEAU 4-2 Hyperparamètres du modèle SVR

Paramètre	Valeur
N	30
kernel	rbf
C	0,1
ϵ	0,01
γ	scale

régression qui minimise l'erreur de prédiction tout en contrôlant la complexité du modèle. Par conséquent, l'ajustement du SVR ne s'interprète pas principalement à travers un nombre d'itérations d'entraînement, mais plutôt au regard du choix de ses hyperparamètres.

Dans ce cadre, le paramètre C contrôle le compromis entre la fidélité aux données d'apprentissage et la capacité de généralisation du modèle. Le paramètre ϵ définit la marge d'insensibilité à l'intérieur de laquelle les écarts entre les valeurs observées et prédites ne sont pas pénalisés. Enfin, le paramètre γ , associé au noyau RBF, détermine l'étendue de l'influence locale de chaque observation d'apprentissage.

4.4.2 *eXtreme Gradient Boosting (XGBoost)*

Le modèle XGBoost retenu dans cette étude repose sur une méthode de *gradient boosting* appliquée à des arbres de décision. Cette approche consiste à construire progressivement un ensemble d'arbres faibles, combinés de manière additive afin d'améliorer la qualité de la prédiction finale. Elle est particulièrement pertinente dans le cadre de la prévision de la consommation électrique résidentielle, puisqu'elle permet de représenter efficacement les relations non linéaires ainsi que les interactions complexes entre les variables explicatives [95]. Les hyperparamètres sélectionnés pour ce modèle sont présentés dans le Tableau 4-3. À la différence des architectures d'apprentissage profond, l'entraînement de XGBoost ne s'organise pas autour d'un suivi en *epochs*, c'est-à-dire en passes complètes et explicitement

TABLEAU 4-3 Hyperparamètres du modèle XGBoost

Paramètre	Valeur
N	30
$n_estimators$	100
$learning_rate$	0,05
max_depth	4
min_child_weight	10
$subsample$	0,8
$colsample_bytree$	0,8
reg_lambda	1

observables sur l'ensemble des données. Son ajustement suit plutôt une logique séquentielle, dans laquelle des arbres de décision sont ajoutés les uns après les autres, chaque nouvel arbre ayant pour rôle de corriger les erreurs laissées par la combinaison précédente. L'interprétation du processus d'apprentissage repose ainsi avant tout sur le choix des hyperparamètres qui gouvernent la structure, la vitesse d'apprentissage et le niveau de régularisation du modèle.

Dans cette configuration, le paramètre $n_estimators$ détermine le nombre total d'arbres construits, tandis que $learning_rate$ fixe l'amplitude de la correction apportée par chaque nouvel arbre. Le paramètre max_depth agit sur la profondeur maximale des arbres et contrôle ainsi leur complexité. Le paramètre min_child_weight impose une contrainte minimale sur le poids des observations dans un nœud, ce qui contribue à limiter le surajustement. Les paramètres $subsample$ et $colsample_bytree$ introduisent respectivement un sous-échantillonnage des observations et des variables à chaque itération, ce qui améliore la robustesse du modèle. Enfin, reg_lambda correspond à une pénalisation de type L_2 , destinée à favoriser une meilleure capacité de généralisation.

4.4.3 Convolutional Neural Networks and Long Short-Term Memory (CNN-LSTM)

Le modèle CNN-LSTM repose sur une architecture hybride combinant un réseau convolutionnel unidimensionnel et un réseau récurrent de type LSTM. Cette combinaison vise à exploiter la complémentarité entre deux mécanismes d'apprentissage. D'une part, l'extraction de motifs locaux dans les séries temporelles par convolution, et d'autre part, la modélisation des dépendances temporelles plus longues par mémoire récurrente. Une telle architecture est particulièrement adaptée à la prévision de la consommation électrique résidentielle, où coexistent des variations locales de courte durée et des structures temporelles plus étendues liées aux habitudes d'usage et aux conditions climatiques [97]. L'architecture retenue, présentée à Figure 4.3, s'inscrit dans une logique hybride associant des mécanismes d'extraction locale de caractéristiques et de modélisation séquentielle. Le modèle CNN-LSTM a ainsi été conçu pour tirer parti de la complémentarité entre les couches convolutionnelles unidimensionnelles, aptes à identifier des structures locales dans la série temporelle, et les couches récurrentes de type LSTM, capables de modéliser les dépendances temporelles de plus longue portée. Une telle organisation est particulièrement pertinente dans le cadre de la prévision de la consommation électrique résidentielle, où coexistent des variations de courte durée, des motifs récurrents et des effets de mémoire liés aux usages et aux conditions environnementales.

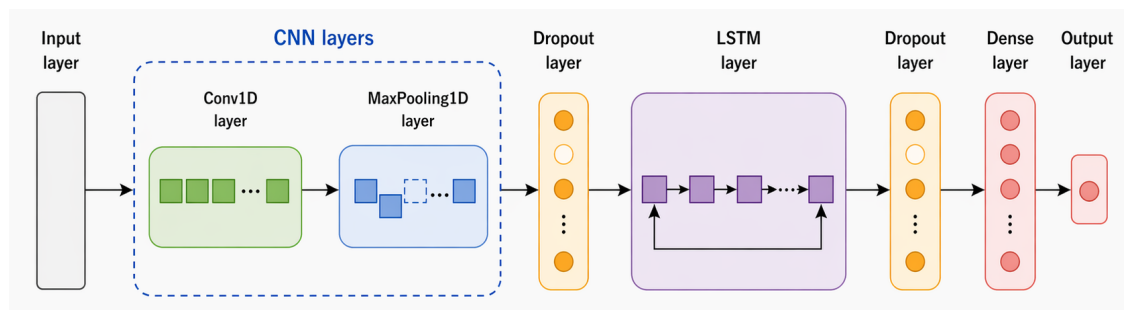


FIGURE 4.3 Architecture du modèle CNN-LSTM

La séquence d'entrée, formée à partir des variables explicatives retenues, est d'abord traitée par une couche Conv1D. Cette dernière applique un ensemble de filtres sur la dimension temporelle afin d'extraire automatiquement des motifs locaux significatifs, tels que des fluctuations brèves, des régularités répétitives ou des variations abruptes du signal. Les hyperparamètres associés à cette couche gouvernent principalement la richesse de la représentation produite et l'étendue temporelle de l'analyse locale. L'activation ReLU est mobilisée afin d'introduire la non-linéarité nécessaire à l'apprentissage de relations complexes entre les variables d'entrée et la consommation prédite [103]. La représentation ainsi obtenue est ensuite transmise à une couche (MaxPooling1D), dont la fonction est de réduire la dimension temporelle des caractéristiques extraites tout en conservant les composantes les plus informatives. Cette opération contribue à atténuer l'effet des fluctuations locales peu pertinentes, à réduire la complexité du modèle et à améliorer sa robustesse. Une première couche de régularisation Dropout est ensuite appliquée afin de limiter précocement le risque de surapprentissage en réduisant la dépendance du réseau à certaines activations particulières [104]. Les caractéristiques ainsi transformées sont ensuite fournies à une couche (LSTM), chargée de capter les dépendances séquentielles de plus long terme. Dans le contexte étudié, cette couche joue un rôle central, dans la mesure où la consommation à un instant donné dépend non seulement de l'état présent du système, mais également de l'évolution récente du signal. La capacité de mémorisation de cette couche est directement conditionnée par ses hyperparamètres, qui déterminent l'aptitude du modèle à représenter la dynamique temporelle sous-jacente. À l'issue de ce traitement récurrent, une seconde couche Dropout est introduite afin de renforcer la régularisation du modèle et de favoriser sa capacité de généralisation. La représentation régularisée est ensuite propagée vers une couche Dense, qui assure une intégration non linéaire de haut niveau des caractéristiques extraites par les blocs précédents. Le nombre de neurones associés à cette couche constitue un hyperparamètre structurant, puisqu'il conditionne la ca-

pacité du modèle à combiner et à transformer l'information avant l'étape finale de prédiction. Enfin, une couche de sortie produit la prédiction finale de la variable cible.

L'apprentissage du réseau est formulé comme un problème de minimisation de la fonction de perte Mean Squared (mse), classiquement utilisée en régression pour pénaliser les écarts entre les valeurs observées et les valeurs prédites. L'optimisation est réalisée à l'aide de l'algorithme (Adam), retenu pour sa stabilité numérique et pour sa capacité à adapter dynamiquement les mises à jour des paramètres au cours de l'entraînement [105]. Dans ce cadre, le taux d'apprentissage constitue un hyperparamètre déterminant, puisqu'il régule l'amplitude des ajustements successifs des poids du réseau. De manière complémentaire, la taille des lots (batch size) influe sur la dynamique d'optimisation, en affectant à la fois la stabilité de l'apprentissage et le coût computationnel. Enfin, le processus d'entraînement est encadré par un mécanisme d'arrêt anticipé (early stopping), destiné à interrompre l'apprentissage dès lors que les performances sur les données de validation cessent de progresser, ce qui permet de prévenir efficacement le surajustement. Les hyperparamètres sélectionnés pour ce modèle sont présentés dans le Tableau 4-4.

TABLEAU 4-4 Hyperparamètres du modèle CNN-LSTM

Paramètre	Valeur
N	30
$filters$	32
$kernel_size$	5
$lstm_units$	64
$dropout$	0,1
$learning_rate$	0,001
$batch_size$	32

La Figure 4.4 présente la courbe de convergence du modèle CNN-LSTM. La dynamique d'apprentissage met en évidence une évolution globalement stable, caractérisée par une diminution progressive de la perte d'entraînement, tandis que la perte de validation demeure

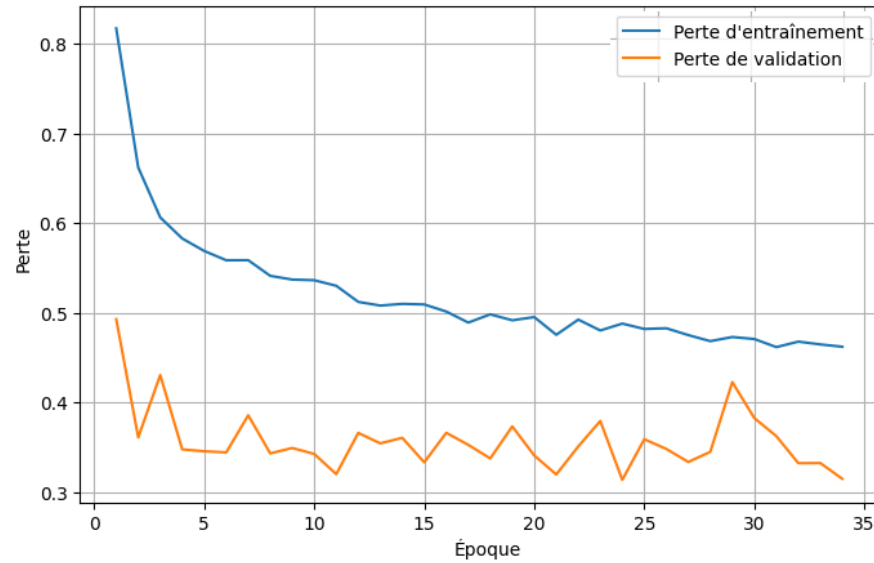


FIGURE 4.4 Courbe de convergence du modèle CNN-LSTM

relativement faible et sans divergence marquée. Un tel comportement suggère que le modèle parvient à extraire des représentations pertinentes tout en conservant une capacité de généralisation satisfaisante, sans signe manifeste de surapprentissage. Cette évolution est observée dans le cadre d'un entraînement autorisé sur un maximum de 100 époques, avec interruption anticipée lorsque l'amélioration de la performance en validation devient insuffisante. Par ailleurs, la longueur de séquence retenue en entrée est fixée à 12 pas de temps, ce qui correspond à un historique de 3 heures et permet au modèle de prendre en compte un contexte temporel suffisamment informatif pour la prédiction, tout en maintenant une complexité d'apprentissage raisonnable.

4.4.4 *Time-series Dense Encoder (TiDE)*

Le modèle Time-series Dense Encoder (TiDE) retenu dans cette étude repose sur une architecture encodeur-décodeur entièrement dense, spécifiquement conçue pour la prévision de séries temporelles multivariées. Contrairement aux approches fondées sur des couches convolutionnelles ou récurrentes, TiDE s'appuie sur des blocs résiduels composés de couches entièrement connectées afin d'extraire et de projeter l'information temporelle contenue dans

l'historique de la série cible ainsi que dans les covariables futures. Cette architecture est particulièrement adaptée lorsque la prévision dépend à la fois de la dynamique récente du signal et d'informations exogènes connues sur l'horizon de prédiction, comme la température ou les variables calendaires [106]. Dans la configuration adoptée, le modèle utilise une fenêtre d'entrée de longueur `input_chunk_length`, correspondant à deux jours d'historique, tandis que l'horizon de sortie est défini par le paramètre (`output_chunk_length`). Cette stratégie permet de formuler le problème sous la forme d'une prévision multi-pas directe, sans recours à une génération séquentielle point par point.

L'architecture interne du modèle, illustrée à la Figure 4.5, s'organise en plusieurs étapes successives. En entrée, une projection dense transforme les covariables passées et futures en représentations de dimension réduite, contrôlées par les paramètres *Temporal width past* (TWP) et *Temporal width future* (TWF). Ces représentations, concaténées avec l'historique de la série cible, alimentent un encodeur composé de blocs résiduels à couches denses, dont le nombre est fixé par (`num_encoder_layers`). L'encodeur produit une représentation latente compacte de dimension (`decoder_output_dim`). Un décodeur, également constitué de blocs résiduels, contrôlé par (`num_decoder_layers`), reconstruit ensuite l'horizon futur à partir de cette représentation.

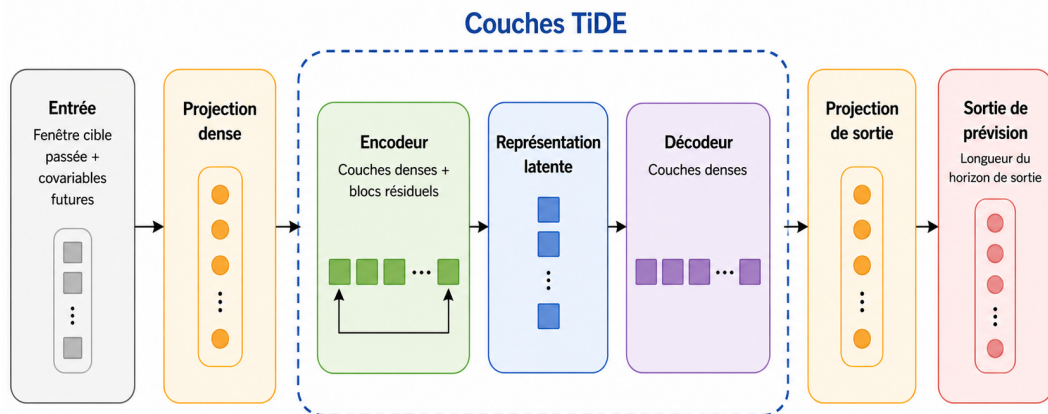


FIGURE 4.5 Architecture du modèle TiDE

Enfin, une projection de sortie, de dimension (`temporal_decoder_hidden`) transforme la sortie du décodeur en prévisions finales sur l'ensemble de l'`output_chunk_length`. La capacité de représentation de l'ensemble du réseau est gouvernée par la dimension de l'espace caché `hidden_size`. Les covariables futures jouent un rôle central dans le processus de prévision. Dans ce travail, elles sont fournies au modèle sur l'ensemble du segment couvrant à la fois la fenêtre d'apprentissage et l'horizon de test. Avant l'entraînement, la série cible et les covariables sont normalisées séparément à l'aide d'un transformateur Scaler, afin d'améliorer la stabilité numérique et de rendre les variables comparables en amplitude.

L'entraînement du modèle est ensuite réalisé sur la série cible normalisée, conjointement avec les covariables futures normalisées, selon les hyperparamètres définis dans la configuration expérimentale. Le Tableau 4-5 résume les hyperparamètres retenus. Ceux-ci portent notamment sur la longueur de la fenêtre d'entrée, l'horizon de sortie, le nombre de couches

TABLEAU 4-5 Hyperparamètres du modèle TiDE

Bloc	Paramètre	Valeur
Input	<i>input_chunk_length</i>	$2 \times \text{points_per_day}$
	<i>output_chunk_length</i>	<i>points_per_day</i>
Dense projection	<i>temporal_width_past</i>	4
	<i>temporal_width_future</i>	4
Encoder	<i>num_encoder_layers</i>	2
	<i>hidden_size</i>	128
	<i>temporal_hidden_size_past</i>	128
Latent representation	<i>decoder_output_dim</i>	16
Decoder	<i>num_decoder_layers</i>	3
	<i>temporal_hidden_size_future</i>	128
Output projection	<i>temporal_decoder_hidden</i>	32
Entraînement	<i>N</i>	30
	<i>dropout</i>	0,3
	<i>batch_size</i>	32
	<i>n_epochs</i>	100
	<i>learning_rate</i>	0,005

de l'encodeur, le nombre de couches du décodeur, la taille de la représentation cachée, la taille des lots ainsi que le nombre maximal d'époques. L'implémentation utilisée repose sur la version originale de TiDEModel de la bibliothèque Darts, avec une gestion automatique de l'accélération matérielle lorsque celle-ci est disponible. Enfin, après l'étape de prédiction, les valeurs estimées sont retransférées dans l'échelle originale de la consommation par inversion de la normalisation appliquée à la série cible. Les prévisions obtenues peuvent alors être comparées directement aux valeurs réelles du jour testé, tant pour l'évaluation des performances prédictives que pour l'analyse ultérieure des résidus en vue de la détection d'anomalies.

La Figure 4.6 présente la courbe de convergence du modèle TiDE. On observe, dans un premier temps, une diminution rapide de la perte d'entraînement ainsi que de la perte de validation, ce qui traduit une capacité d'apprentissage efficace dès les premières époques.

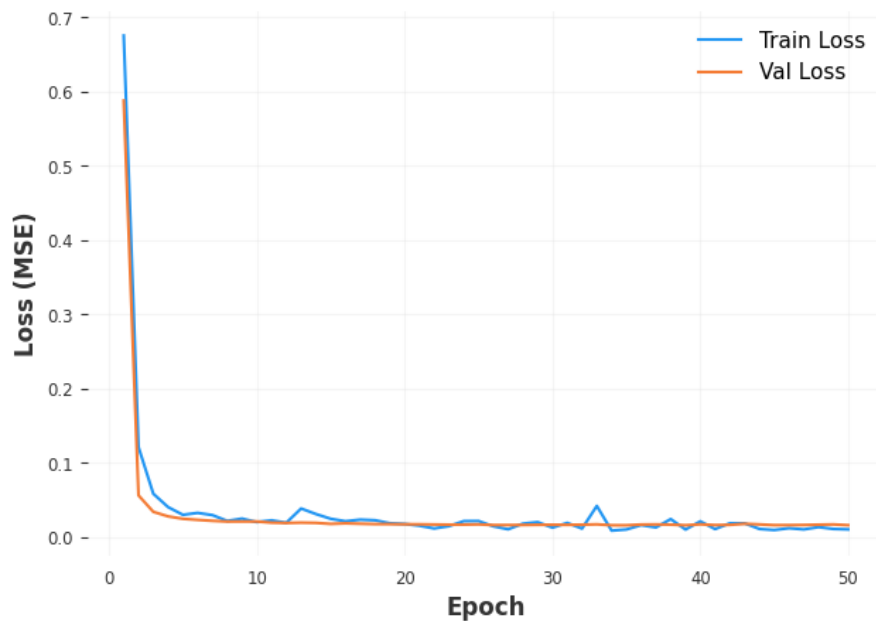


FIGURE 4.6 Courbe de convergence du modèle TiDE

Par la suite, les deux courbes se stabilisent à un niveau faible et demeurent globalement proches l'une de l'autre, sans divergence marquée. Un tel comportement suggère que le modèle parvient à apprendre des représentations pertinentes tout en conservant une capacité de

généralisation satisfaisante. Malgré la présence de légères fluctuations résiduelles, la dynamique observée ne met pas en évidence de surapprentissage prononcé. Dans l'ensemble, cette courbe de convergence indique que le modèle TiDE présente un apprentissage stable et une performance de validation cohérente avec l'objectif de prévision considéré.

4.5 Prédiction du profil de référence

Avant d'examiner les performances du système en matière de détection d'anomalies, il est indispensable d'évaluer la capacité des modèles à reproduire correctement le comportement normal de la consommation. En effet, l'approche retenue dans ce mémoire repose sur l'analyse des résidus de prédiction, définis comme l'écart entre la consommation observée et la consommation estimée par le modèle. La qualité de cette prédiction de référence conditionne donc directement la fiabilité du processus de détection, un modèle peu précis engendre des résidus plus dispersés, ce qui peut élargir artificiellement les seuils statistiques ou, au contraire, accroître le risque de fausses alertes. Dans cette perspective, la performance prédictive des modèles est évaluée à l'aide de quatre métriques classiques de régression, à savoir le RMSE, le MAE, le MAPE et le coefficient de détermination R^2 .

4.5.1 Root Mean Square Error (RMSE)

Le RMSE mesure l'écart quadratique moyen entre les valeurs observées et les valeurs prédites. En donnant un poids plus important aux erreurs de grande amplitude, il met particulièrement en évidence les mauvaises prédictions ponctuelles. Dans le contexte de ce travail, une valeur élevée du RMSE traduirait donc la présence de déviations significatives entre le profil réel et le profil estimé, susceptibles d'élargir les résidus et de perturber le mécanisme de détection [7]. Comme l'indique l'équation (4.2), cette métrique correspond à la racine carrée de la moyenne des écarts quadratiques entre la valeur observée $y(t)$ et la valeur prédite $\hat{y}(t)$, calculée sur un ensemble de n observations :

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y(t) - \hat{y}(t))^2} \quad (4.2)$$

où n désigne le nombre total d'observations de la période d'évaluation, $y(t)$ la consommation réelle à l'instant t , et $\hat{y}(t)$ la consommation estimée par le modèle au même instant.

4.5.2 Mean Absolute Error (MAE)

Le MAE représente l'erreur absolue moyenne entre les valeurs observées et les valeurs prédites. Contrairement au RMSE, il ne surpondère pas les erreurs extrêmes et fournit ainsi une mesure plus directe de l'écart moyen de prédiction. Cette métrique présente l'avantage d'être exprimée dans la même unité que la consommation, ce qui en facilite l'interprétation [7]. Comme le montre l'équation (4.3), elle correspond à la moyenne des valeurs absolues des écarts entre $y(t)$ et $\hat{y}(t)$ sur l'ensemble des n instants considérés :

$$MAE = \frac{1}{n} \sum_{t=1}^n |y(t) - \hat{y}(t)| \quad (4.3)$$

où n représente le nombre total d'observations, $y(t)$ la valeur observée à l'instant t , et $\hat{y}(t)$ la valeur prédite correspondante.

4.5.3 Mean Absolute Percentage Error (MAPE)

Le MAPE exprime l'erreur de prédiction de manière relative, sous forme de pourcentage de la valeur observée. Cette normalisation permet de comparer les performances des modèles indépendamment de l'échelle absolue de consommation et offre une lecture plus intuitive de la précision obtenue. Dans cette étude, il permet ainsi d'apprécier l'importance moyenne des écarts relativement au niveau réel de charge [7]. Sa formulation, donnée par l'équation (4.4), repose sur la moyenne des erreurs absolues rapportées à la valeur observée $y(t)$:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y(t) - \hat{y}(t)}{y(t)} \right| \times 100 \quad (4.4)$$

où n désigne le nombre total d'observations, $y(t)$ la consommation mesurée à l'instant t , et $\hat{y}(t)$ la consommation prédite par le modèle.

4.5.4 Coefficient de détermination (R^2)

Le coefficient de détermination R^2 quantifie la proportion de la variance de la consommation observée qui est expliquée par le modèle de prédiction. Une valeur proche de 1 signifie que le modèle reproduit fidèlement les variations du signal, alors qu'une valeur proche de 0 indique qu'il n'en explique qu'une faible part [7]. Dans le contexte de cette étude, le R^2 permet ainsi d'apprécier la capacité de chaque modèle à représenter la dynamique de consommation du bâtiment en fonction des conditions d'usage et des variables explicatives considérées. Sa définition est donnée par l'équation (4.1), présentée à la section 4.4.

Le Tableau 4-6 présente une comparaison des performances des modèles de prévision selon quatre métriques complémentaires, le RMSE, le MAE, le MAPE et le coefficient de détermination R^2 . Ces indicateurs permettent d'évaluer respectivement l'importance des fortes erreurs de prédiction, l'erreur moyenne absolue, l'erreur relative en pourcentage, ainsi que la capacité du modèle à expliquer la variance des données observées.

TABLEAU 4-6 Comparaison des performances des modèles de prévision

Modèle	RMSE	MAE	MAPE (%)	R^2
SVR	0,445	0,360	83,400	0,424
XGBoost	0,360	0,291	74,360	0,600
CNN-LSTM	0,377	0,310	69,240	0,586
TiDE	0,270	0,238	46,880	0,651

Les résultats montrent que le modèle TiDE présente les meilleures performances de prévision, avec le RMSE, le MAE et le MAPE les plus faibles, ainsi que le coefficient R^2 le plus élevé. Cela indique que ce modèle est capable de reproduire plus fidèlement l'évolution de la consommation, tout en limitant les écarts importants entre les valeurs observées et les valeurs prédites. Le RMSE plus faible montre que TiDE pénalise moins les fortes erreurs de prédiction, tandis que le MAE confirme une erreur moyenne plus réduite. De plus, le MAPE plus faible indique que l'erreur relative reste plus limitée par rapport aux valeurs réelles. Enfin, la valeur plus élevée de R^2 signifie que le modèle explique une plus grande part de la variance des données. Ces performances de prévision sont importantes pour la détection d'anomalies, car celle-ci repose sur l'analyse des résidus. Un modèle plus précis produit généralement des résidus plus stables en fonctionnement normal, ce qui facilite l'identification des écarts anormaux et peut réduire le nombre de fausses alertes.

4.6 Résultats de détection

4.6.1 Analyse du compromis (Réactivité vs Faux positifs)

Cette section présente les résultats obtenus par le système de détection d'anomalies énergétiques fondé sur l'analyse des résidus de prédiction. L'objectif est de quantifier, de manière empirique et reproductible, l'effet du seuil de la bande de confiance sur les performances de détection. Plus précisément, il s'agit d'évaluer dans quelle mesure le choix du seuil influence à la fois la qualité de classification des anomalies et la rapidité de déclenchement de l'alerte. L'analyse ne vise donc pas uniquement à maximiser un score de classification. Dans un contexte résidentiel, un système de détection doit satisfaire deux exigences opérationnelles complémentaires, détecter rapidement une anomalie réelle afin de permettre une réaction précoce de l'utilisateur, tout en limitant les notifications inutiles susceptibles de réduire la confiance dans le système. L'évaluation est ainsi conduite selon deux axes princi-

paux, d'une part, l'analyse du compromis entre réactivité et fausses alertes ; d'autre part, la comparaison finale des algorithmes retenus dans le banc d'essai.

Quatre modèles prédictifs sont comparés, SVR, XGBoost, CNN-LSTM et TiDE. Pour chacun de ces modèles, la décision de détection repose sur l'analyse des résidus de prédiction. À chaque instant t , le modèle produit une prévision $\hat{y}_t$ qui est confrontée à la consommation réellement observée y_t . En régime nominal, lorsque le modèle a correctement appris le comportement énergétique du logement, les résidus oscillent autour de zéro avec une dispersion bornée par le bruit naturel de la consommation. En présence d'une anomalie, les résidus s'écartent significativement de ce comportement attendu, ce qui constitue le signal sur lequel s'appuie la détection. Le signal résiduel alimente simultanément deux branches parallèles de traitement. La première est une branche de référence, construite à partir de données enregistrées en régime strictement normal. La seconde est une branche d'analyse, construite à partir de données dans lesquelles une anomalie a été injectée artificiellement. Les deux branches partagent un même pipeline de calcul, mais produisent des descriptions statistiques distinctes. La branche de référence sert à calibrer la décision, tandis que la branche d'analyse sert à l'évaluer.

Dans la branche de référence, un premier seuil statistique est défini à partir de la distribution des résidus observés en régime normal. Ce seuil, noté $Band_R$ et défini par l'équation 4.5, correspond au percentile associé au niveau de confiance retenu. Il permet ainsi d'identifier, à partir de l'historique d'apprentissage, les dépassements naturels du signal résiduel.

$$Band_R = \text{Percentile}_{Band_R\%}(\{r_t\}_{t \in \text{normal}}) \quad (4.5)$$

où r_t désigne le résidu de prédiction à l'instant t . Tout intervalle temporel durant lequel

le résidu demeure supérieur à cette bande est alors considéré comme un épisode de dépassement. Chaque épisode est caractérisé par un instant de début t_{start} et un instant de fin t_{end} , formant ainsi l'intervalle $[t_{\text{start}}, t_{\text{end}}]$.

Pour chaque épisode détecté, une surface S_i est calculée afin de quantifier l'intensité cumulée du dépassement. Cette surface dépend de la géométrie locale du signal résiduel, une approximation triangulaire est utilisée lorsqu'il s'agit d'un pic isolé composé d'un seul point de dépassement, tandis qu'une approximation trapézoïdale est retenue lorsque l'épisode comprend plusieurs points consécutifs. De manière générale, cette surface, définie par l'équation 4.6, peut être interprétée comme l'intégrale de l'excès du résidu au-dessus du seuil $Band_R$ sur toute la durée de l'épisode.

$$S_i = \int_{t_{\text{start}}}^{t_{\text{end}}} \max(0, r_t - Band_R) dt \quad (4.6)$$

où t_{start} et t_{end} représentent respectivement les instants de début et de fin de l'épisode de dépassement. Le terme $\max(0, r_t - Band_R)$ permet de ne conserver que la partie du résidu située au-dessus du seuil $Band_R$. L'ensemble des surfaces obtenues en régime normal forme une distribution empirique propre au comportement nominal. La queue supérieure de cette distribution caractérise les épisodes les plus sévères pouvant être observés en conditions de fonctionnement habituelles. Un seuil de sévérité S_q est alors défini par l'équation 4.7, à partir du quantile haut de cette distribution.

$$S_q = \text{Percentile}_{S_q\%}(\{S_i\}_{i \in \text{épisodes normaux}}) \quad (4.7)$$

où $\{S_i\}_{i \in \text{épisodes normaux}}$ représente l'ensemble des surfaces calculées sur les épisodes de dépassement observés en régime normal, et $S_q\%$ désigne le niveau de quantile retenu pour

fixer le seuil de sévérité.

Ce choix offre une référence statistiquement plus robuste que le simple maximum, en filtrant les valeurs extrêmes potentiellement aberrantes tout en conservant l'information sur le niveau de sévérité tolérable en régime normal. Dans la branche d'analyse, après l'injection de l'anomalie, le même procédé de détection et de calcul de surface est appliqué au signal résiduel anormal. Le critère de décision repose sur le concept de Time-to-Equal-Severity (TES), évalué pour chaque épisode du signal anormal dont la surface totale excède le seuil S_q . La surface cumulée depuis le début de l'épisode jusqu'à l'instant τ est définie par l'équation 4.8.

$$S(\tau) = \int_{t_{\text{start}}}^{\tau} \max(0, r_t - \text{Band}_R) dt \quad (4.8)$$

où τ représente un instant intermédiaire compris entre le début et la fin de l'épisode, c'est-à-dire $\tau \in [t_{\text{start}}, t_{\text{end}}]$. La fonction $S(\tau)$ mesure donc l'accumulation progressive de l'excès du résidu au-dessus du seuil Band_R . Le moment de détection, noté τ^* , est défini par l'équation 4.9 comme le premier instant où cette surface cumulée franchit strictement le seuil de sévérité S_q .

$$\tau^* = \min \left\{ \tau \in [t_{\text{start}}, t_{\text{end}}] : S(\tau) > S_q \right\} \quad (4.9)$$

Avec τ^* représente l'instant de déclenchement de l'alerte. Un épisode est ainsi déclaré anormal si, et seulement si, il existe un instant $\tau^* \in [t_{\text{start}}, t_{\text{end}}]$ pour lequel la surface cumulée dépasse S_q . Le délai de détection est alors déduit de l'écart entre τ^* et l'instant d'injection de l'anomalie. Cette logique en deux temps répond à deux objectifs distincts mais complémentaires. Le premier filtre, contrôlé par Band_R , fixe la sensibilité globale du système en délimitant ce qui appartient au comportement normal. Le second filtre, contrôlé par S_q , ajuste

l'exigence de sévérité afin d'écarter les épisodes dont la sévérité intégrée dans le temps demeure compatible avec le régime normal. C'est donc la combinaison des deux seuils, plutôt que chacun pris isolément, qui détermine le compromis entre la rapidité de détection et la robustesse face aux fausses alertes.

Les deux seuils sont testés selon cinq niveaux, soit 60 %, 75 %, 90 %, 95 % et 99 %, ce qui conduit à vingt-cinq configurations par modèle et à un total de cent configurations évaluées. Pour chaque configuration, les performances sont mesurées à l'aide de métriques de classification, à savoir la précision, le rappel, la spécificité et le F1-score, ainsi qu'au moyen du décompte des vrais positifs, faux positifs, vrais négatifs et faux négatifs. La vitesse de déclenchement de l'alerte est également mesurée à partir du délai séparant l'instant d'injection d'une anomalie de type décalage de niveau de la première détection produite par le système. Le Tableau 4-7 présente les paramètres du banc d'essai.

TABLEAU 4-7 Paramètres du banc d'essai expérimental et caractéristiques de l'anomalie injectée

Paramètre	Valeur
Granularité temporelle	15 minutes
Fenêtre d'entraînement	30 jours
Type d'anomalie	Décalage positif de niveau (<i>level-shift up</i>)
Heure de début de l'injection	19 h 00
Heure de fin de l'anomalie	23 h 45
Durée de l'anomalie	5 heures
Amplitude du décalage	50 % de la consommation observée
Nombre de points anormaux	20 observations
Nombre de points normaux	76 observations
Seuils $Band_R$ testés sur les résidus	60 %, 75 %, 90 %, 95 % et 99 %
Seuils S_q testés sur les surfaces	60 %, 75 %, 90 %, 95 % et 99 %
Modèles évalués	SVR, XGBoost, CNN-LSTM et TiDE

Le réglage du couple de seuils constitue ainsi le paramètre central de la décision. Les

résultats expérimentaux montrent que la variation de $Band_R$ et de S_q modifie simultanément le nombre de faux positifs, le rappel et le F1-score. L'analyse est donc conduite selon deux axes complémentaires, d'une part, l'évolution du nombre de faux positifs en fonction des seuils et, d'autre part, la contrepartie observée sur le rappel et le F1-score. Dans cette logique, l'analyse débute par l'évolution du nombre de faux positifs, qui constitue l'indicateur central de la fiabilité opérationnelle des notifications.

La Figure 4.7 présente le nombre moyen de faux positifs en fonction de chacun des deux seuils, avec la plage minimale et maximale représentée en zone ombrée.

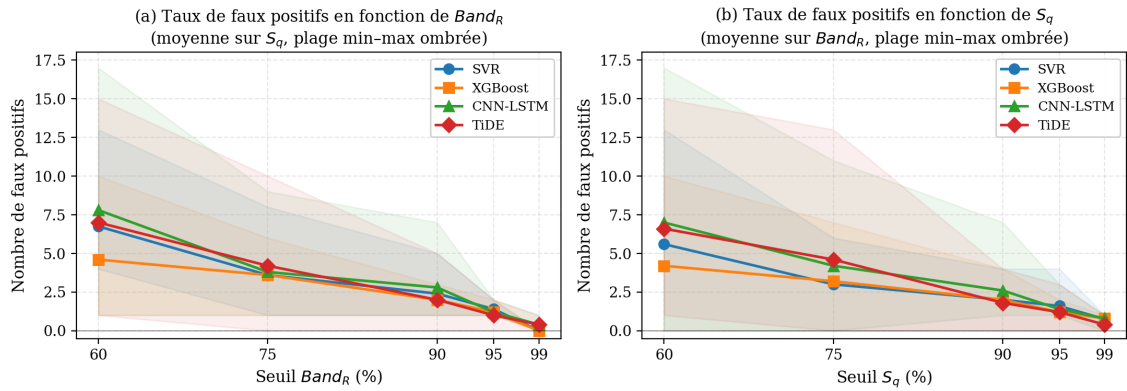


FIGURE 4.7 Évolution du nombre de faux positifs en fonction des seuils $Band_R$ et S_q

Les deux sous-figures révèlent une décroissance monotone très nette du taux de faux positifs lorsque les seuils augmentent, ce qui valide empiriquement le levier principal du système de détection. Sur l'axe $Band_R$, le nombre moyen de faux positifs passe de 4,6 pour XGBoost à 7,8 pour CNN-LSTM à $Band_R = 60\%$, avec un maximum atteignant 17 fausses alertes dans la configuration la plus défavorable. La précision minimale observée, descendant à 52 % dans les configurations les moins favorables, illustre cette difficulté, une part importante des alertes produites peut alors correspondre à des variations normales de la consommation plutôt qu'à de véritables anomalies. À $Band_R = 95\%$, il descend autour de 1,0 à 1,4 selon le modèle, et il devient nul pour la quasi-totalité des configurations situées à

$Band_R = 99\%$. Une décroissance comparable est observée sur l'axe S_q , avec des moyennes passant de 4,2 à 7,0 à $S_q = 60\%$ à 0,4 à 0,8 à $S_q = 99\%$. XGBoost se distingue par la courbe la plus basse à seuils permissifs, ce qui traduit une tendance plus modérée à générer des fausses alertes en régime sensible. L'amplitude de la zone min-max diminue également avec les seuils, ce qui indique que la dispersion entre configurations se réduit à mesure que la décision devient plus stricte. Cette caractérisation confirme que la zone $[90\%, 95\%]$ constitue le domaine où le taux de faux positifs devient acceptable opérationnellement, tout en maintenant une marge suffisante avant la saturation observée à 99% . Cette réduction des faux positifs s'accompagne toutefois d'une contrepartie sur le rappel et le F1-score. La Figure 4.8 présente l'évolution du rappel, en sous-figure (a), et du F1-score, en sous-figure (b), en fonction du seuil $Band_R$ appliqué aux résidus de prédiction.

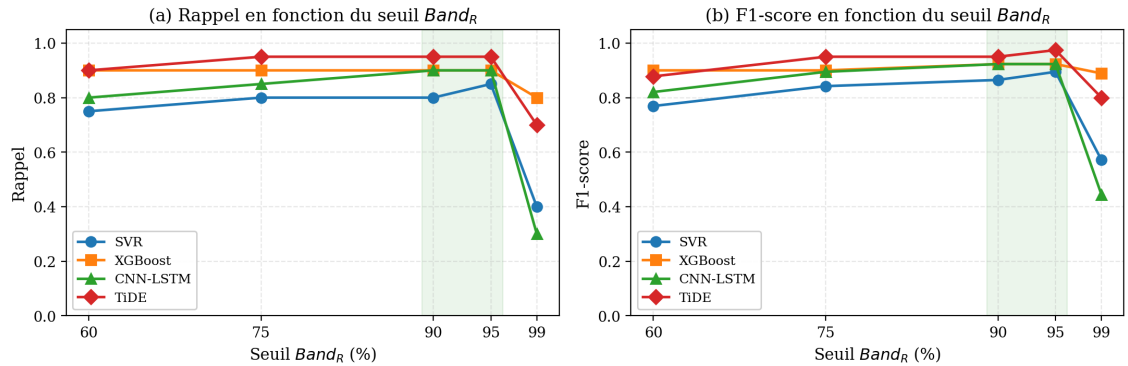


FIGURE 4.8 Évolution du rappel et du F1-score en fonction du seuil $Band_R$

Au seuil $Band_R = 99\%$, seuls les dépassements les plus extrêmes sont considérés comme des anomalies, ce qui entraîne une perte importante de sensibilité, le rappel chute à 40% pour SVR, à environ 30% pour CNN-LSTM, et reste limité à 80% pour XGBoost. TiDE conserve un rappel plus stable, situé entre 75% et 90% selon le niveau de S_q retenu. Entre 90% et 95% , une zone de fonctionnement favorable se dessine où les modèles les plus performants, notamment XGBoost, CNN-LSTM et TiDE, atteignent des niveaux élevés de rappel et de F1-score, compris entre $0,920$ et $0,970$ selon le niveau de S_q . Le modèle SVR

demeure en retrait, avec un rappel généralement plus faible et une progression plus limitée lorsque le seuil augmente. Une exception notable apparaît pour TiDE à $Band_R = 95\%$ et $S_q = 95\%$, où le système atteint simultanément une précision de 1 et un rappel de 0,950 sans aucune fausse alerte. Cette configuration correspond au point optimal observé dans le scénario expérimental considéré et illustre la capacité de TiDE à produire des résidus dont la distribution en régime nominal demeure suffisamment concentrée pour être nettement séparée des résidus issus de la phase anormale. Le choix d'un seuil $Band_R$ compris entre 90 % et 95 % apparaît donc comme le meilleur compromis dans ce scénario expérimental, car il maintient un niveau élevé de détection des anomalies tout en évitant une dégradation excessive du rappel et du F1-score. Pour appuyer cette analyse, la Figure 4.9 isole l'effet du seuil S_q sur les deux indicateurs opérationnels centraux du système, à savoir la précision et le délai de détection. Les valeurs reportées correspondent à la moyenne arithmétique calculée sur l'ensemble des configurations testées, soit les quatre modèles et tous les niveaux de $Band_R$, pour chaque valeur de S_q . Cette agrégation permet de mettre en évidence la tendance globale associée au seuil de surface, indépendamment des spécificités propres à chaque modèle ou à chaque réglage de la bande supérieure.

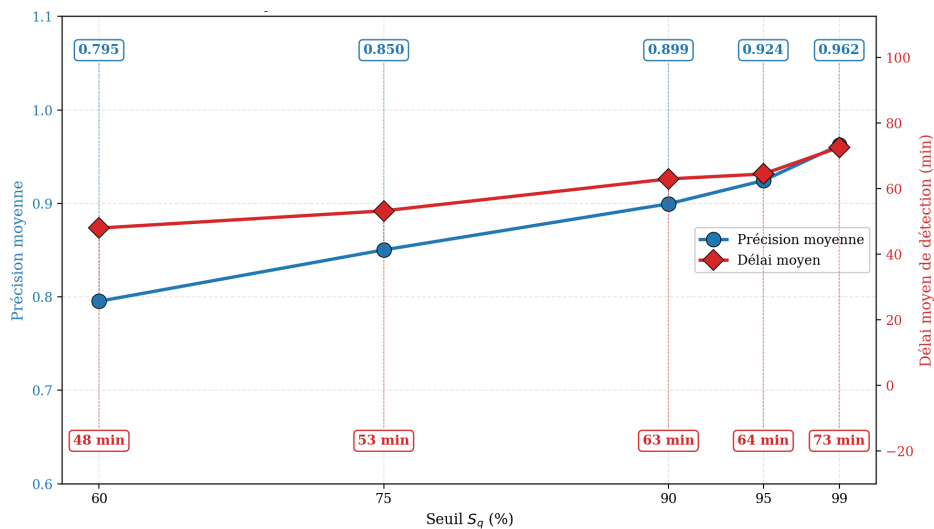


FIGURE 4.9 Compromis précision-délai en fonction du seuil S_q

Cette représentation met en évidence le compromis fondamental entre la fiabilité des alertes et la rapidité de détection résultant du réglage du seuil S_q . Les deux indicateurs présentent une croissance monotone parallèle lorsque S_q augmente de 60 % à 99 %. La précision moyenne passe de 0,795 à 0,962, soit un gain de 16,7 points, ce qui traduit une nette réduction de la proportion de fausses alertes parmi les notifications émises. Simultanément, le délai moyen de détection s’allonge de 48 minutes à 73 minutes, soit une augmentation de plus de 50 %. Cette évolution s’explique par le fait que des dépassements de plus en plus sévères doivent être accumulés avant que la surface cumulée ne franchisse le seuil de décision. Cette progression conjointe démontre qu’aucun choix de S_q ne permet de maximiser simultanément les deux objectifs, un seuil bas privilégie la réactivité au prix d’une précision dégradée, tandis qu’un seuil élevé privilégie la fiabilité des alertes au prix d’une détection plus tardive. Dans un contexte résidentiel, où l’objectif est à la fois d’alerter rapidement l’utilisateur et de limiter les notifications injustifiées, ce résultat impose un arbitrage explicite. Les valeurs intermédiaires de S_q , situées entre 90 % et 95 %, apparaissent comme le meilleur point d’équilibre dans le scénario expérimental considéré, car elles offrent une précision moyenne supérieure à 0,90 tout en maintenant un délai moyen inférieur à 65 minutes.

4.6.2 *Évaluation comparative des algorithmes*

Cette section est consacrée à la comparaison des performances des différents modèles du banc d’essai. Elle vise à analyser de manière synthétique les résultats obtenus par les quatre algorithmes étudiés, soit SVR, XGBoost, CNN-LSTM et TiDE, afin d’identifier le modèle présentant le meilleur comportement global dans les conditions expérimentales et pour le scénario d’anomalie considérés. La comparaison finale repose sur deux dimensions complémentaires. La première concerne la performance intrinsèque de détection, évaluée principalement à l’aide de la précision, du rappel et du F1-score. Le délai de détection est également

pris en compte, mais il ne constitue pas un critère discriminant à la configuration optimale, puisque les quatre modèles déclenchent l’alerte au même moment. Cette dimension permet donc d’évaluer principalement la capacité de chaque modèle à détecter correctement l’anomalie injectée tout en limitant les erreurs de classification. La seconde dimension concerne la pertinence opérationnelle des modèles. Elle tient compte de leur stabilité face aux différents seuils testés, de leur sensibilité au paramétrage ainsi que de leur coût d’entraînement. Cette distinction est importante, car le modèle présentant les meilleures performances statistiques n’est pas nécessairement le plus adapté à un déploiement dans un environnement résidentiel contraint, où les ressources de calcul, la fréquence de réentraînement et la simplicité d’exploitation doivent également être considérées. Le Tableau 4-8 présente la meilleure configuration retenue pour chacun des quatre modèles. Cette synthèse permet d’établir une comparaison expérimentale cohérente et de formuler une analyse comparative des algorithmes selon leurs performances de détection et leur pertinence opérationnelle.

TABLEAU 4-8 Synthèse comparative des modèles à leur configuration optimale (meilleur F1-score)

Modèle	$Band_R$	S_q	Précision	Rappel	F1	Exactitude	Détection	Entr. (s)
SVR	95 %	60 %	0,905	0,950	0,927	0,969	19 h 15	2,4
XGBoost	95 %	75 %	0,950	0,950	0,950	0,979	19 h 15	7,9
CNN-LSTM	95 %	75 %	0,950	0,950	0,950	0,979	19 h 15	67,8
TiDE	95 %	95 %	1,000	0,950	0,974	0,990	19 h 15	319,6

La Figure 4.10 présente une synthèse comparative des performances optimales obtenues par les quatre algorithmes étudiés. Chaque carte regroupe les indicateurs principaux associés au meilleur paramétrage de chaque modèle, notamment le rappel, la précision, l’exactitude, le F1-score, l’heure de détection, le seuil optimal et le temps d’entraînement. Cette représentation permet d’évaluer simultanément la qualité de détection, la réactivité de l’alerte et le coût computationnel de chaque approche.

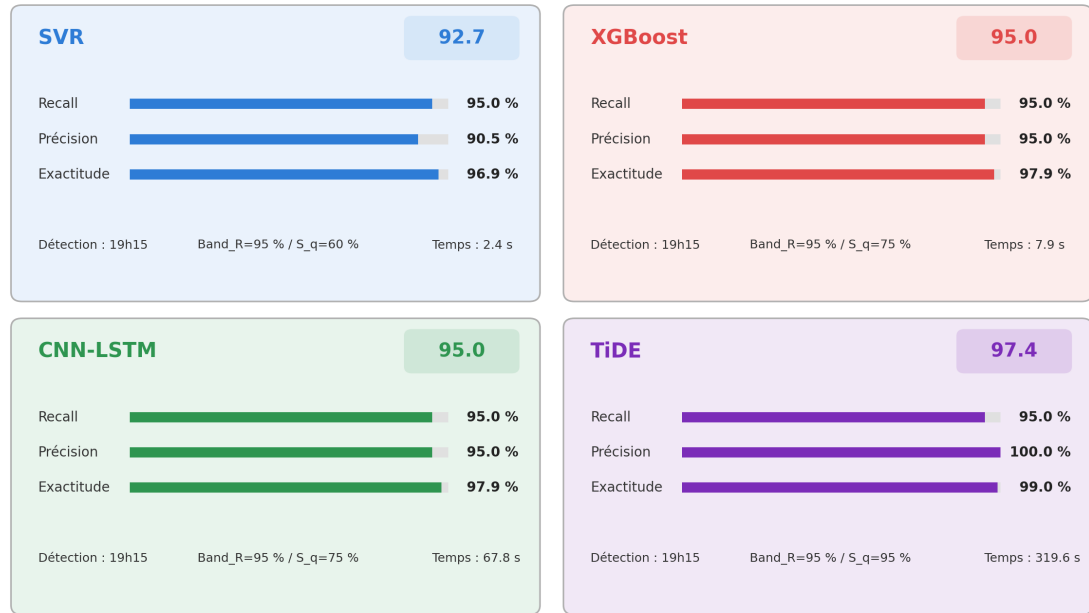


FIGURE 4.10 Synthèse comparative des performances optimales des quatre algorithmes

Les résultats montrent d'abord que les quatre modèles déclenchent leur alerte à 19 h 15 dans leur configuration optimale, ce qui correspond à un délai d'un seul intervalle de quinze minutes après l'injection de l'anomalie à 19 h 00.

La réactivité maximale atteignable par le banc d'essai est donc identique pour les quatre approches dans leur meilleur paramétrage. La différenciation entre les algorithmes se situe alors principalement au niveau du rappel, de la précision, du F1-score et du temps d'entraînement, ainsi qu'au niveau de la robustesse observée à travers les vingt-cinq configurations testées pour chacun. Le modèle SVR présente les performances les plus faibles parmi les quatre méthodes. Dans sa configuration optimale, soit $Band_R = 95\%$ et $S_q = 60\%$, il atteint un rappel de 0,950, une précision de 0,905, une exactitude de 0,969 et un F1-score de 0,927. Toutefois, sa précision moyenne sur l'ensemble des configurations à seuil élevé demeure inférieure à celle des autres modèles, et il ne parvient à atteindre une précision de 100 % qu'au prix d'une perte massive de rappel, qui tombe alors à 40 %. Malgré un temps

d'entraînement très faible de 2,4 secondes, ce modèle reste moins adapté à un usage opérationnel exigeant, car il détecte une proportion plus limitée des points anormaux dans les régions à forte fiabilité du paramétrage et présente un taux de faux positifs plus élevé en moyenne. XGBoost améliore nettement les résultats par rapport au SVR. Dans sa configuration optimale, soit $Band_R = 95\%$ et $S_q = 75\%$, il atteint un rappel de 0,950, une précision de 0,950, une exactitude de 0,979 et un F1-score de 0,950, avec une détection à 19 h 15. Son temps d'entraînement demeure très faible, soit 7,9 secondes. Ce modèle offre donc un compromis très favorable entre performance de détection et coût computationnel. Son délai de détection est identique à celui des autres modèles à leur configuration optimale. Le modèle CNN-LSTM obtient des performances de détection identiques à celles de XGBoost dans sa configuration optimale, avec un rappel de 0,950, une précision de 0,950, une exactitude de 0,979 et un F1-score de 0,950. L'alerte est également déclenchée à 19 h 15. Toutefois, son temps d'entraînement est nettement plus élevé, atteignant 67,8 secondes, et son comportement est plus sensible au choix du seuil. En particulier, lorsque $Band_R$ atteint 99 %, le rappel chute autour de 0,300 et le F1-score s'effondre, ce qui traduit une stabilité plus faible que celle de XGBoost. Dans ce scénario expérimental, la complexité supplémentaire du modèle ne se traduit donc pas par un gain de performance par rapport à XGBoost. Enfin, TiDE présente la meilleure performance intrinsèque de détection parmi les quatre modèles dans le scénario expérimental considéré. Dans sa configuration optimale, définie par $Band_R = 95\%$ et $S_q = 95\%$, il atteint un rappel de 0,950, une précision de 1,000, une exactitude de 0,990 et le F1-score le plus élevé, soit 0,974. Il déclenche l'alerte à 19 h 15, comme les autres modèles à leur configuration optimale. Le délai de détection ne permet donc pas de les départager dans ce scénario. TiDE demeure toutefois le seul modèle à atteindre une précision parfaite tout en maintenant un rappel élevé. Pour les autres algorithmes, l'absence de faux positifs n'est obtenue qu'avec $Band_R = 99\%$, au prix d'une diminution marquée du rappel.

Par ailleurs, TiDE présente six configurations sur vingt-cinq sans aucun faux positif, contre cinq pour SVR et XGBoost et trois pour CNN-LSTM. Ce résultat suggère une meilleure séparation entre les régimes normal et anormal. Son temps d'entraînement demeure toutefois nettement plus élevé, soit 319,6 secondes, ce qui peut limiter son utilisation dans un environnement disposant de ressources de calcul restreintes. L'analyse met ainsi en évidence deux conclusions complémentaires. Sur le plan statistique, TiDE offre les meilleures performances dans le scénario considéré, puisqu'il maximise le F1-score et la précision tout en conservant un rappel élevé. Sur le plan opérationnel, XGBoost constitue une solution particulièrement compétitive, car ses performances sont proches de celles de TiDE, alors que son temps d'entraînement est environ quarante fois plus faible. Le choix du modèle dépend donc du contexte d'utilisation. Dans le scénario expérimental considéré, TiDE apparaît comme la solution la plus pertinente lorsque la priorité est accordée à la qualité de détection et à la fiabilité des alertes. XGBoost constitue toutefois une alternative opérationnelle particulièrement compétitive en raison de son faible coût d'entraînement et de ses performances proches de celles de TiDE. Ces conclusions demeurent limitées aux données, au bâtiment et au type d'anomalie évalués.

Par ailleurs, la Figure 4.11 complète l'analyse précédente en proposant une lecture globale des performances au seuil de 95 %. Ce diagramme radar permet de comparer simultanément les modèles selon plusieurs critères, sans limiter l'analyse à une seule métrique. La forme du polygone associé à chaque algorithme renseigne sur l'équilibre général de ses performances, plus le contour est étendu et régulier, plus le modèle présente un comportement homogène selon les critères évalués. Dans les conditions du banc d'essai, TiDE présente le profil le plus favorable, avec une surface plus étendue que celles des autres modèles. Ce résultat traduit, pour le scénario considéré, une combinaison avantageuse entre la qualité de détection et la fiabilité des alertes.

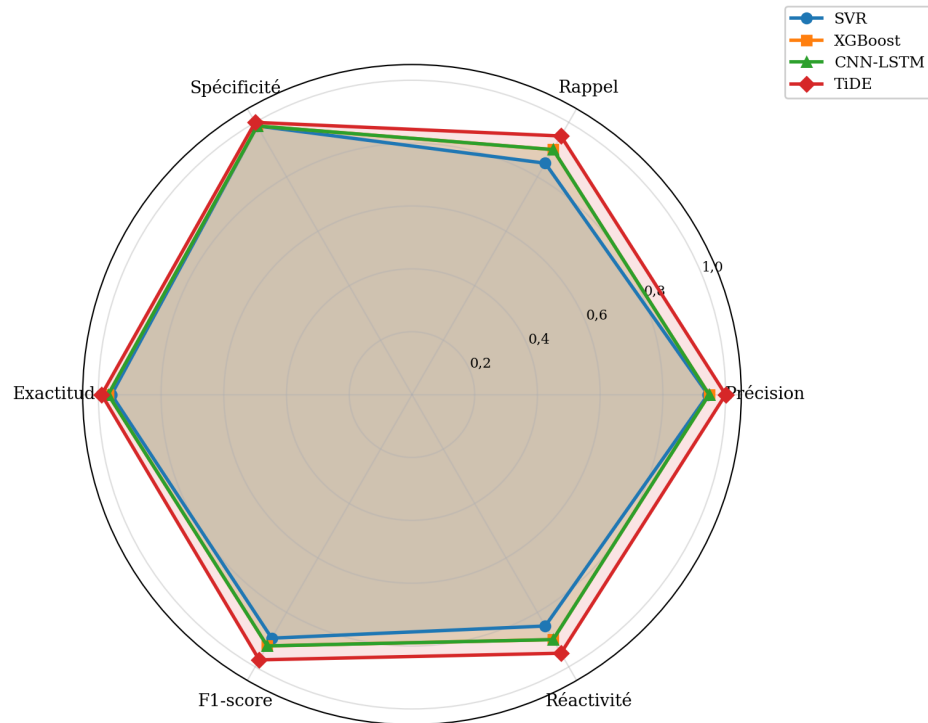


FIGURE 4.11 Profil multi-métriques des quatre algorithmes au seuil optimal global

La rapidité de déclenchement ne permet toutefois pas de différencier les modèles à leur configuration optimale, puisque les quatre algorithmes détectent l'anomalie au même moment. XGBoost et CNN-LSTM affichent des profils proches, ce qui confirme leurs performances comparables au seuil étudié. Leur comportement demeure globalement équilibré, mais légèrement inférieur à celui de TiDE pour les critères liés à la précision et à la fiabilité des alertes. Le modèle SVR présente, quant à lui, un contour plus réduit, ce qui reflète des performances plus limitées en matière de précision moyenne et de stabilité face au paramétrage dans cette expérimentation. Cette représentation confirme ainsi la hiérarchie observée pour le scénario évalué, TiDE obtient la meilleure performance globale au seuil de 95 %, tandis que XGBoost et CNN-LSTM constituent des alternatives compétitives. SVR présente des résultats plus modestes et joue principalement le rôle de modèle de référence dans le cadre expérimental étudié. Ces observations demeurent toutefois propres aux données, au bâtiment

et au type d'anomalie considérés.

4.7 Discussion

Les résultats obtenus dans le scénario expérimental étudié suggèrent que le système de détection proposé pourrait présenter un intérêt pratique pour l'identification d'anomalies énergétiques en contexte résidentiel. L'analyse des résidus de prédiction permet de repérer les écarts significatifs entre la consommation observée et la consommation attendue, en tenant compte du comportement historique du logement et des variables explicatives utilisées par les modèles. Cette approche fournit ainsi une base quantitative pour déclencher une alerte lorsque la consommation s'éloigne du fonctionnement normal appris par le système.

L'utilité opérationnelle du système repose principalement sur deux éléments. Premièrement, les résultats expérimentaux montrent qu'il est possible d'atteindre une absence totale de faux positifs dans certaines configurations bien choisies, en particulier pour TiDE à $Band_R = 95\%$ et $S_q = 95\%$, ce qui constitue un avantage important pour un système destiné à l'utilisateur final. En effet, des notifications trop fréquentes ou injustifiées peuvent réduire la confiance de l'occupant et conduire à l'abandon du dispositif. Dans la zone de fonctionnement favorable identifiée précédemment, soit $Band_R$ compris entre 90 % et 95 %, le nombre de faux positifs reste très limité pour l'ensemble des modèles, avec une moyenne inférieure à deux fausses alertes par configuration sur les soixante-seize points normaux évalués. Deuxièmement, les quatre modèles permettent une détection relativement rapide de l'anomalie dans leur configuration optimale, avec un délai d'un seul intervalle de quinze minutes, ce qui rend possible une intervention précoce avant que la surconsommation ne se prolonge.

Le choix du couple de seuils $(Band_R, S_q)$ joue un rôle déterminant dans le compromis entre fiabilité et réactivité. Les résultats montrent qu'un paramétrage trop permissif, en deçà de $Band_R = 75\%$, expose le système à un nombre élevé de fausses alertes, parfois supé-

rieur à dix sur la fenêtre de test, ce qui diminue la pertinence opérationnelle des notifications. Inversement, un paramétrage trop conservateur, à $Band_R = 99\%$, élimine pratiquement les fausses alertes mais entraîne une chute marquée du rappel, en particulier pour SVR et CNN-LSTM, dont la sensibilité descend respectivement à 40 % et 30 %. Cette dépendance forte des résultats au paramétrage met en évidence l'importance d'une phase de calibration empirique préalable au déploiement, fondée sur l'historique de consommation propre à chaque logement. Cependant, il est nécessaire de distinguer la détection d'une anomalie et l'identification certaine de sa cause. Le système permet de signaler qu'un comportement énergétique inhabituel est observé, mais le diagnostic causal généré doit être interprété comme une hypothèse explicative plutôt que comme une conclusion définitive. Par exemple, une hausse anormale de la consommation peut être compatible avec plusieurs causes possibles, telles qu'un équipement laissé en marche, une déperdition thermique, une modification du comportement d'occupation ou une condition extérieure particulière. Le système peut donc orienter l'utilisateur vers les causes les plus probables, sans garantir à lui seul l'origine exacte de l'anomalie.

Dans cette perspective, les notifications envoyées à l'utilisateur doivent être formulées de manière prudente et actionnable. Une alerte ne devrait pas affirmer directement une cause unique, mais plutôt indiquer qu'une consommation inhabituelle a été détectée à un moment précis et recommander une vérification ciblée. Par exemple, le système pourrait signaler qu'une surconsommation a été observée depuis plusieurs intervalles de temps et suggérer de vérifier les équipements énergivores, le chauffage ou les ouvertures susceptibles d'entraîner une perte thermique. Cette formulation permet d'éviter une surinterprétation du diagnostic tout en fournissant une information utile à l'utilisateur.

La fiabilité du diagnostic causal généré par le système dépend toutefois de plusieurs

conditions qu'il convient d'expliciter, afin de délimiter le cadre dans lequel les recommandations peuvent être considérées comme exploitables. Dans un premier temps, la qualité du diagnostic est conditionnée par la calibration préalable du modèle sur un historique suffisamment représentatif du logement concerné. Les résultats expérimentaux montrent en effet qu'un même modèle peut produire des taux de faux positifs variant de 0 à 17 selon le choix du couple $(Band_R, S_q)$; un déploiement sans calibration adaptée au logement aboutirait donc à des alertes peu fiables. De plus, la fiabilité repose sur un retour d'information de l'utilisateur, qui permet de confirmer ou d'infirmer le caractère anormal des événements signalés. Ce retour peut être recueilli par des mécanismes simples, qui permettent de marquer une alerte comme pertinente ou non pertinente, et utilisé pour ajuster les seuils de manière itérative ou pour affiner les hypothèses causales associées aux profils observés. En complément, la pertinence des hypothèses causales gagne à être recoupée avec d'autres signaux lorsque ceux-ci sont disponibles, par exemple la durée d'absence de l'occupant. Un diagnostic appuyé sur plusieurs sources convergentes est nettement plus fiable qu'un diagnostic fondé uniquement sur l'écart aux résidus. Il convient toutefois de souligner que la fiabilité opérationnelle dépend de la quantification de l'incertitude associée à chaque alerte. Plutôt qu'une étiquette binaire normal/anormal, le système gagnerait à associer à chaque notification un indice de confiance fondé sur l'ampleur du dépassement et sur sa cohérence avec l'historique récent, ce qui permettrait à l'utilisateur de hiérarchiser ses vérifications. Dans l'état actuel des expérimentations, le système peut donc servir d'aide à la décision, mais sa pleine fiabilité supposerait l'intégration progressive de ces quatre conditions, lesquelles constituent autant de pistes pour les travaux futurs.

L'intérêt pratique du système dépend également du choix du modèle utilisé. Dans notre cas d'étude, TiDE présente la meilleure performance intrinsèque de détection, avec un rappel élevé, le F1-score le plus élevé de l'étude et la seule configuration combinant une précision

parfaite avec un délai minimal. Il apparaît donc comme une solution pertinente lorsque la priorité est accordée à la qualité de détection et à la fiabilité des notifications. Toutefois, son coût d'entraînement plus élevé, d'environ cinq minutes, peut limiter son utilisation dans un environnement contraint, notamment lorsqu'un réentraînement fréquent est nécessaire pour suivre l'évolution des habitudes de consommation. XGBoost constitue une alternative opérationnelle particulièrement intéressante, puisqu'il combine une bonne qualité de détection, avec un F1-score de 0,950, une réactivité équivalente à celle de TiDE et un temps d'entraînement d'environ huit secondes, soit près de quarante fois inférieur à celui de TiDE. Dans les conditions étudiées, ce modèle pourrait donc être mieux adapté à un déploiement résidentiel nécessitant des recalibrations fréquentes. Par ailleurs, CNN-LSTM atteint des performances comparables à celles de XGBoost à sa configuration optimale, mais sa plus forte sensibilité aux variations des seuils et son coût d'entraînement intermédiaire le rendent moins avantageux dans le cadre expérimental considéré. Enfin, SVR conserve l'avantage de la simplicité et d'un faible coût computationnel, mais ses performances plus modestes et son nombre moyen de faux positifs plus élevé le positionnent davantage comme un modèle de référence que comme une solution principale.

Chapitre 5 – Conclusion

5.1 Synthèse

Ce mémoire a abordé la question de la détection automatique d'anomalies de consommation énergétique en milieu résidentiel sous l'angle d'un compromis fondamental, déclencher l'alerte le plus tôt possible après l'apparition d'un comportement anormal, tout en limitant les fausses alertes susceptibles d'éroder la confiance de l'utilisateur final. La démarche adoptée a été conduite de manière progressive, depuis l'acquisition des données jusqu'à l'évaluation expérimentale du système de détection.

La première étape a consisté en l'acquisition multi-source des données. Les mesures de puissance instantanée issues du capteur Emporia Vue ont été couplées à des covariables météorologiques externes, notamment la température extérieure, ainsi qu'à des indicateurs temporels. Cette combinaison entre mesures internes au logement et conditions environnementales constitue le socle d'un modèle prédictif capable d'expliquer la consommation énergétique autrement que par sa seule autocorrélation temporelle.

La deuxième étape a porté sur la transformation, la synchronisation et l'agrégation des données. Les différentes sources ont été ramenées à une granularité commune de 15 minutes, afin de garantir une cohérence temporelle entre les mesures de consommation et les variables explicatives. Des variables cycliques liées à l'heure du jour, au jour de la semaine et à la saisonnalité ont également été introduites. Cette préparation unifiée permet de comparer les algorithmes dans des conditions équitables, en s'assurant que les différences observées pro-

viennent principalement des capacités propres des modèles et non de traitements de données divergents.

La troisième étape a consisté à mettre en place un banc d'essai prédictif confrontant quatre familles algorithmiques distinctes, SVR, XGBoost, CNN-LSTM et TiDE. Chaque modèle a été entraîné sur les 30 jours précédant le jour de test. Les métriques classiques de qualité prédictive, calculées en l'absence d'anomalie, montrent une hiérarchie claire entre les modèles. TiDE obtient la meilleure qualité prédictive, avec un coefficient de détermination de 0,651 et une MAPE de 46,8 %, suivi par XGBoost avec $R^2 = 0,600$, CNN-LSTM avec $R^2 = 0,586$, puis SVR avec $R^2 = 0,424$.

La quatrième étape a introduit un critère de détection par sévérité surfacique, fondé sur l'analyse des résidus de prédiction. Le résidu est défini comme l'écart entre la consommation observée et la consommation prédite. La détection ne repose pas uniquement sur un dépassement ponctuel de la bande supérieure $Band_R$, mais sur la surface cumulée du dépassement anormal. Cette surface est ensuite comparée à un seuil de sévérité S_q estimé au quantile haut de la distribution des surfaces observées en régime nominal. Ainsi, la décision d'alerte tient compte simultanément de l'amplitude et de la persistance de l'écart, ce qui permet de filtrer les dépassements isolés ou peu significatifs.

La cinquième étape a consisté à évaluer le système à l'aide d'un protocole d'injection contrôlée d'anomalies. Une anomalie de type level-shift a été injectée à un instant précis sur le jour de test. Deux seuils complémentaires ont été balayés selon une grille à cinq niveaux chacun, soit 60 %, 75 %, 90 %, 95 % et 99 % pour $Band_R$ et S_q , afin d'analyser empiriquement le compromis entre réactivité et fiabilité. Au total, cent configurations ont été évaluées, correspondant aux quatre algorithmes testés sur les vingt-cinq couples de seuils.

Les résultats de détection mettent en évidence deux constats principaux. Premièrement,

le compromis entre réactivité et fiabilité des alertes est nettement matérialisé par les données. Le nombre moyen de faux positifs varie de 4,6 à 7,8 selon le modèle lorsque $Band_R$ est fixé à 60 %, avec un maximum atteignant 17 fausses alertes dans la configuration la plus défavorable, où la précision descend à 0,520. À l'inverse, à $Band_R = 95\%$, le nombre moyen de faux positifs se stabilise autour de 1,0 à 1,4 selon le modèle, et il devient nul pour la quasi-totalité des configurations à $Band_R = 99\%$. La précision moyenne agrégée sur l'ensemble des modèles et des niveaux de $Band_R$ progresse de 0,795 à $S_q = 60\%$ jusqu'à 0,962 à $S_q = 99\%$, tandis que le délai moyen de détection passe corrélativement de 48 à 73 minutes. Cette double progression confirme empiriquement qu'aucun choix de seuil ne permet d'optimiser simultanément les deux objectifs, ce qui légitime l'arbitrage opérationnel proposé.

Deuxièmement, les différences entre les modèles apparaissent principalement au niveau de la précision, du rappel, du F1-score, de la robustesse aux seuils et du coût computationnel. Le délai de détection ne constitue pas un critère discriminant à la configuration optimale, puisque les quatre modèles détectent l'anomalie 15 minutes après son injection. TiDE, avec le couple $(Band_R = 95\%, S_q = 95\%)$, présente la meilleure performance globale dans le scénario étudié, avec un F1-score de 0,974, une précision de 1,000 et un rappel de 0,950. XGBoost et CNN-LSTM, tous deux évalués avec le couple $(95\%, 75\%)$, atteignent un F1-score de 0,950. SVR, avec le couple $(95\%, 60\%)$, obtient un F1-score plus faible, soit 0,927. Ces résultats suggèrent qu'une meilleure qualité prédictive, associée à des résidus plus resserrés en régime normal, favorise une séparation plus nette entre les comportements normaux et anormaux. Le choix final du modèle dépend toutefois du contexte d'utilisation. Dans le cadre expérimental analysé, TiDE constitue la solution la plus performante sur le plan statistique, tandis que XGBoost représente une alternative opérationnelle particulièrement compétitive en raison de son faible coût d'entraînement et de ses performances proches de celles de TiDE.

Bien que la réactivité ait été prise en compte dans l'analyse du compromis entre rapidité de détection et réduction des fausses alertes, elle ne constitue pas un critère discriminant entre les modèles à leur configuration optimale dans le scénario étudié. Les quatre méthodes déclenchent en effet l'alerte à 19 h 15, soit quinze minutes après l'injection de l'anomalie. La comparaison finale repose donc principalement sur les métriques de détection, la robustesse au paramétrage et le coût computationnel.

5.2 Contributions principales

La présente recherche apporte trois contributions complémentaires. La première est d'ordre architectural et concerne l'intégration de données hétérogènes issues du compteur intelligent et de sources météorologiques externes dans un pipeline prédictif unifié. Cette fusion permet de contextualiser la consommation électrique résidentielle en tenant compte des conditions environnementales, plutôt que de s'appuyer uniquement sur la régularité temporelle du signal. La deuxième contribution est d'ordre méthodologique. Elle repose sur la formalisation d'un critère de détection non supervisé fondé sur l'analyse des résidus de prédiction et sur la sévérité surfacique cumulée des dépassements. Ce critère, calibré à partir des seules données nominales, permet de détecter des écarts anormaux sans nécessiter de base annotée d'anomalies, tout en offrant un cadre explicite et reproductible pour ajuster le compromis entre fiabilité des alertes et rapidité de détection. La troisième contribution est d'ordre comparatif et applicatif. Elle consiste à mettre en place un banc d'essai reproductible permettant d'évaluer plusieurs familles algorithmiques dans un même protocole expérimental. Cette comparaison fournit un cadre d'aide au choix des modèles selon les contraintes de déploiement, notamment en matière de qualité de détection, de réactivité, de robustesse et de coût computationnel. L'ensemble de ces contributions vise à rapprocher la détection d'anomalies d'un usage opérationnel dans le contexte de l'efficacité énergétique résidentielle.

5.3 Limites de l'étude

Le présent travail comporte plusieurs limites qui encadrent la portée des résultats obtenus. La première concerne la dépendance à la qualité des données d'entrée. Les performances du système reposent sur la précision des mesures de puissance, sur la fiabilité des covariables météorologiques et sur leur bonne synchronisation temporelle. Toute erreur de mesure, dérive de calibration ou discordance entre les sources peut affecter la qualité des prédictions et, par conséquent, la fiabilité de la détection. Une deuxième limite tient à l'absence d'information directe sur l'occupation du logement. Les habitudes des occupants, les changements ponctuels d'usage ou les événements exceptionnels ne sont pas explicitement observés par le modèle. Ces facteurs peuvent modifier la consommation de manière importante et compliquer la distinction entre une variation normale du comportement résidentiel et une anomalie réelle. Une troisième limite concerne le coût computationnel des modèles les plus complexes. Les approches les plus performantes peuvent nécessiter davantage de ressources pour l'entraînement, ce qui peut limiter leur utilisation dans des environnements contraints ou dans des contextes nécessitant des recalibrations fréquentes. Le choix du modèle doit donc tenir compte à la fois de la qualité de détection et des contraintes pratiques de déploiement.

Les résultats de cette étude doivent être interprétés dans les limites du cadre expérimental retenu. Les performances observées et la hiérarchie établie entre les modèles demeurent liées aux données, aux conditions d'évaluation et aux choix méthodologiques adoptés. Elles ne peuvent donc pas être généralisées à l'ensemble des contextes résidentiels ou des situations de détection d'anomalies. Des validations complémentaires dans des conditions expérimentales plus diversifiées sont nécessaires afin de confirmer la portée et la robustesse de ces résultats.

5.4 Travaux futurs

Plusieurs perspectives peuvent être dégagées à partir des limites identifiées. Un premier axe concerne le déploiement du système en environnement embarqué ou à la périphérie du réseau. Une telle approche permettrait de réduire la dépendance à un serveur central, de limiter la latence et de renforcer la protection des données personnelles en conservant localement les mesures fines de consommation. Un deuxième axe porte sur le passage d'un système de détection passive vers un système d'aide au contrôle actif. Le détecteur pourrait être couplé à des dispositifs domestiques, notamment des thermostats intelligents, afin de proposer des ajustements adaptés lorsque des consommations anormales sont identifiées. Cette évolution nécessiterait toutefois un module de diagnostic plus avancé, capable de distinguer les causes possibles de l'anomalie tout en conservant une validation finale par l'utilisateur.

Un troisième axe concerne l'élargissement du protocole de validation à des conditions expérimentales plus diversifiées. Une telle démarche permettrait de vérifier la stabilité des performances observées, de confirmer la hiérarchie établie entre les modèles et de mieux évaluer leur capacité de généralisation. Il serait ainsi pertinent d'étudier une plus grande diversité d'anomalies, de logements, de profils de consommation et de conditions climatiques. Une validation longitudinale permettrait également de mieux prendre en compte les transitions saisonnières, les changements d'usage et la nécessité de recalibrer périodiquement les modèles et les seuils. Cet élargissement contribuerait à renforcer la portée et la robustesse des conclusions obtenues.

Un quatrième axe vise l'adaptation temporelle des seuils de décision. Plutôt que d'utiliser un seuil global unique, des bandes de confiance adaptées aux différentes périodes de la journée pourraient mieux représenter la variabilité naturelle de la consommation résidentielle. Cette approche permettrait d'améliorer la fiabilité des alertes durant les périodes naturelle-

ment variables et d'accroître la sensibilité pendant les périodes plus stables. Un cinquième axe concerne l'hybridation des modes de détection. Le critère de sévérité surfacique pourrait être complété par des mécanismes dédiés aux dérives lentes et aux changements progressifs, afin de couvrir à la fois les anomalies ponctuelles, les surconsommations soutenues et les évolutions à long terme liées aux équipements ou au bâtiment.

Enfin, un dernier axe concerne l'enrichissement des sources de données. L'intégration de capteurs domestiques supplémentaires, tels que des détecteurs de présence, des prises intelligentes ou des capteurs d'ouverture, permettrait d'améliorer l'interprétation causale des anomalies. Le système pourrait ainsi produire des recommandations plus compréhensibles et plus directement applicables par l'utilisateur final.

Références

- [1] Manuel Baude et al. « Chiffres clés du climat : France, Europe et Monde ». In : *Ministère de la Transition Écologique, France* (). Édition 2023. URL : <https://www.statistiques.developpement-durable.gouv.fr/edition-numerique/chiffres-cles-du-climat-2024/fr/pdf/chiffres-cles-du-climat-2024.pdf>.
- [2] Alexandru Andrei et al. « Chiffres clés de l'énergie ». In : *Ministère de la Transition Écologique, France* (). Édition 2024. URL : <https://www.statistiques.developpement-durable.gouv.fr/edition-numerique/chiffres-cles-energie-2024/pdf/Chiffres-cles-energie-2024.pdf>.
- [3] Hydro-Québec. *Faits saillants de l'État d'avancement 2021 du Plan d'approvisionnement 2020 2029*. 2021. URL : https://nouvelles.hydroquebec.com/media/filer_private/2021/11/01/faits_saillant_etat_avancement_2021_approvisionnement_final.pdf.
- [4] Nilson Henao. « Désagrégation non-intrusive de charges électriques résidentielles ». Thèse de doct. Université du Québec à Trois-Rivières, 2018.
- [5] Whitmore Johanne et Pierre-Olivier Pineau. *État de l'énergie au Québec*. Chaire de gestion du secteur de l'énergie, Tech. Rep, Édition 2023.
- [6] Farshad Etedadi Aliabadi. « Stratégie de Gestion Optimale de l'Énergie et de la Puissance pour un Groupe de Bâtiments Résidentiels Basé sur des Multi-Agents Coopératifs ». Thèse de doctorat. Université du Québec à Trois-Rivières, mai 2023.
- [7] Aicha Hayaoui. « Optimisation de la consommation énergétique des serres dans le contexte du québec ». Mémoire de Maîtrise. Université du Québec à Trois-Rivières, Avril 2024.
- [8] Whitmore Johanne et Pierre-Olivier Pineau. *État de l'énergie au Québec*. Chaire de gestion du secteur de l'énergie, Tech. Rep, Édition 2022.
- [9] Thomas Storrington. *Daily Stats Article*. Residential heating fuel and equipment, 2019. URL : <https://novascotia.ca/finance/statistics/news.asp?id=17217>.
- [10] WILLIAM DEVIA. « Gestion intelligente de la recharge des véhicules électriques dans le secteur résidentiel ». Mémoire de maîtrise. Université du Québec à Trois-Rivières, oct. 2020.
- [11] Hydro-Québec. *Une énergie renouvelable pour un Québec durable*. Rapport sur le développement durable 2023.
- [12] HYDRO-QUÉBEC. *Rapports annuels 2015 à 2023*. Rapports annuels. 2015–2023.

- [13] ABDOUL WAHAB DANTE. « Gestion intelligente de la recharge des véhicules électriques dans le secteur résidentiel ». Thèse de doctorat. Université du Québec à Trois-Rivières, Août 2023.
- [14] Yassine Himeur et al. « Efficient Multi-Descriptor Fusion for Non-Intrusive Appliance Recognition ». In : *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. 2020, p. 1-5. DOI : 10.1109/ISCAS45731.2020.9180493.
- [15] K. Carrie Armel et al. « Is disaggregation the holy grail of energy efficiency ? The case of electricity ». In : *Energy Policy* 52 (2013). Special Section : Transition Pathways to a Low Carbon Economy, p. 213-234. ISSN : 0301-4215. DOI : <https://doi.org/10.1016/j.enpol.2012.08.062>. URL : <https://www.sciencedirect.com/science/article/pii/S0301421512007446>.
- [16] Shulong Wang et al. « A novel IoT access architecture for vehicle monitoring system ». In : *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*. 2016, p. 639-642. DOI : 10.1109/WF-IoT.2016.7845396.
- [17] Cyrine Zid. « Analyse et détection des anomalies dans un système IOT distribué hybride ». Mémoire de maîtrise. Polytechnique Montréal, oct. 2020.
- [18] Lorenzo Mascali et al. « A machine learning-based Anomaly Detection Framework for building electricity consumption data ». In : *Sustainable Energy, Grids and Networks* 36 (2023), p. 101194. DOI : 10.1016/j.segan.2023.101194.
- [19] L. Canale et al. « A novel model to detect abnormal thermal energy consumption in existing buildings with limited metering infrastructure ». In : *Energy and Buildings* 323 (2024), p. 114803. DOI : 10.1016/j.enbuild.2024.114803.
- [20] Mohammed Alosaimi, Omer Rana et Charith Perera. « Testbeds and Evaluation Frameworks for Anomaly Detection within Built Environments : A Systematic Review ». In : *ACM Computing Surveys* 57.9 (2025), p. 227. DOI : 10.1145/3722213.
- [21] Ken Takahashi, Ryoza Ooka et Atsuyoshi Kurosaki. « Seasonal threshold to reduce false positives for prediction-based outlier detection in building energy data ». In : *Journal of Building Engineering* 84 (2024), p. 108539. DOI : 10.1016/j.jobeb.2024.108539.
- [22] Moussa Ibrahim Moumouni et al. « Shrunked-Space Search Method for LVCTs' Parameters Identification ». In : *IEEE Transactions on Industrial Electronics* 71.10 (2024), p. 12911-12921. DOI : 10.1109/TIE.2024.3355517.
- [23] Y. Lin et al. « A Review on the Impact of Outdoor Environment on Indoor Thermal Environment ». In : *Buildings* 13.10 (2023), p. 2600.
- [24] ASHRAE. *ASHRAE Handbook – Fundamentals*. Atlanta, GA : American Society of Heating, Refrigerating et Air-Conditioning Engineers, 2021.
- [25] Fatima Amara. « Modélisation et prévision de la demande d'électricité résidentielle ». Université du Québec, Doctorat en génie électrique. Thèse de doctorat. Université du Québec à Trois-Rivières, 2018.
- [26] Canadian Wood Council. *Thermal Performance of Light-Frame Assemblies*. Documentation technique. 2015.

- [27] ISO. *ISO 6946 : Building components and building elements – Thermal resistance and thermal transmittance – Calculation methods*. International Organization for Standardization, 2017.
- [28] ISO. *ISO 13786 : Thermal performance of building components – Dynamic thermal characteristics – Calculation methods*. International Organization for Standardization, 2017.
- [29] R. Rugani et al. « Impact of Building Thermal Inertia in Different Climates ». In : *Proceedings of the IBPSA Conference*. 2021. DOI : 10.26868/25222708.2021.30515.
- [30] National Research Council Canada. *National Energy Code of Canada for Buildings*. 2020.
- [31] Cristina Guzman. « Contrôle et gestion intelligents d'énergie et de puissance dans les systèmes électriques résidentiels ». Thèse de doct. Université du Québec à Trois-Rivières, 2019.
- [32] A. Maboudi Reveshti et al. « Energy consumption prediction in an office building by examining occupancy rates and weather parameters using the moving average method and artificial neural network ». In : *Heliyon* 10 (2024), e25307. DOI : 10.1016/j.heliyon.2024.e25307.
- [33] Khensa Dab. « Caractérisation électrique, thermique et comportementale hiérarchisée d'un agrégat de résidences en vue de la prévision énergétique à court terme ». Thèse de doct. Université du Québec à Trois-Rivières, 2024.
- [34] M. Hosseini, K. Javanroodi et V. M. Nik. « High-resolution impact assessment of climate change on building energy performance considering extreme weather events and microclimate ». In : *Sustainable Cities and Society* 78 (2022), p. 103634. DOI : 10.1016/j.scs.2021.103634.
- [35] Da Yan et al. « Occupant behavior modeling for building performance simulation : Current state and future challenges ». In : *Energy and Buildings* 107 (2015), p. 264-278. DOI : 10.1016/j.enbuild.2015.08.032.
- [36] H. Kaddouri et al. « Impact of Insulation using Bio-sourced Materials on the Thermal and Energy Performance of a Typical Residential Building in Morocco ». In : *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 117.1 (2024), p. 43-59. DOI : 10.37934/arfmts.117.1.4359.
- [37] Luis Fernando Rueda Vasquez. « Modélisation Du Comportement Des Occupants Pour Des Applications De Gestion De La Demande D'électricité Dans Le Secteur Résidentiel ». Thèse de doctorat. Université du Québec à Trois-Rivières, 2021.
- [38] Filipe Rodrigues et al. « Short-Term Load Forecasting of Electricity Demand for the Residential Sector Based on Modelling Techniques : A Systematic Review ». In : *Energies* 16.10 (2023), p. 4098. DOI : 10.3390/en16104098. URL : <https://repositorio.ipl.pt/bitstreams/7d35ede7-347a-48b8-b339-4dc84514d38e/download>.

- [39] Pamela Corriveau-Peev, Louis Gosselin et André Potvin. « Identifying and classifying thermal anomalies in Arctic houses using two infrared cameras and different metrics ». In : *Building and Environment* 280 (2025), p. 113101. DOI : 10.1016/j.buildenv.2025.113101.
- [40] Mathieu Bonte, Françoise Thellier et Bérangère Lartigue. « Impact of occupant's actions on energy building performance and thermal sensation ». In : *Energy and Buildings* 76 (2014), p. 219-227. ISSN : 0378-7788. DOI : <https://doi.org/10.1016/j.enbuild.2014.02.068>. URL : <https://www.sciencedirect.com/science/article/pii/S0378778814002047>.
- [41] Marta A. R. Lopes et al. « Estimating energy savings from behaviours using building performance simulations ». In : *Building Research & Information* 45.3 (2017), p. 303-319. DOI : 10.1080/09613218.2016.1140000.
- [42] Wei Guo et al. « Classification of Anomaly Patterns in Integrated Energy Systems Based on Conditional Variational Autoencoder and Attention Mechanism ». In : *Energies* 16.11 (2023), p. 4367. DOI : 10.3390/en16114367. URL : <https://doi.org/10.3390/en16114367>.
- [43] S. Mirzabeigi, R. Razkenari et P. Crovella. « Détection des anomalies thermiques de l'enveloppe des bâtiments par une technique intégrée de vision et de segmentation sémantique ». In : *Buildings* 15.15 (2025), p. 2672.
- [44] B. Kari, B. Perrin et J. C. Foures. « Modélisation du comportement hygrothermique des parois multicouches : prévision des risques de condensation dans les parois à isolation intérieure et extérieure ». In : *Materials and Structures* 25.152 (1992), p. 558-565. DOI : 10.1007/BF02472452.
- [45] Arthur L. Samuel. « Some Studies in Machine Learning Using the Game of Checkers ». In : *IBM Journal of Research and Development* 3.3 (1959), p. 210-229. DOI : 10.1147/rd.33.0210.
- [46] Varun Chandola, Arindam Banerjee et Vipin Kumar. « Anomaly Detection : A Survey ». In : *ACM Computing Surveys* 41.3 (2009), p. 1-58. DOI : 10.1145/1541880.1541882.
- [47] Victoria J. Hodge et Jim Austin. « A Survey of Outlier Detection Methodologies ». In : *Artificial Intelligence Review* 22.2 (2004), p. 85-126. DOI : 10.1023/B:AIRE.0000045502.10941.a9.
- [48] Yassine Himeur et al. « Artificial intelligence based anomaly detection of energy consumption in buildings : A review, current trends and new perspectives ». In : *Applied Energy* 287 (2021), p. 116601. DOI : 10.1016/j.apenergy.2021.116601.
- [49] David Kaimenyi Marangu, Stephen Thiiru Njenga et Rachael Njeri Ndung'u. « A Comparative Analysis of Class Imbalance Handling Techniques for Deep Models in the Detection of Anomalies in Energy Consumption ». In : *Int. J. of Artificial Intelligence and Applications* 15.6 (2024), p. 1-19.

- [50] Aitor Sánchez-Ferrer, Borja Calvo et Jose A. Lozano. « A Review on Self-Supervised Learning for Time Series Anomaly Detection : Recent Advances and Open Challenges ». In : *arXiv preprint arXiv :2501.15196* (2025). arXiv :2501.15196v2 [stat.ML].
- [51] Kukjin Choi et al. « Deep Learning for Anomaly Detection in Time-Series Data : Review, Analysis, and Guidelines ». In : *IEEE Access* 9 (2021), p. 120043-120065. DOI : 10.1109/ACCESS.2021.3107975.
- [52] Charu C. Aggarwal. *Outlier Analysis*. 2nd. Springer, 2017.
- [53] Jie Gui et al. « A Survey on Self-supervised Learning : Algorithms, Applications, and Future Trends ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46.12 (2024), p. 9052-9071. DOI : 10.1109/TPAMI.2024.3415112.
- [54] Hadi Hojjati, Thi Ho et Narges Armanfard. *Self-Supervised Anomaly Detection : A Survey and Outlook*. Mai 2022. DOI : 10.48550/arXiv.2205.05173.
- [55] Massoud Amin et Bruce Wollenberg. « Toward a Smart Grid : Power Delivery for the 21st Century ». In : *IEEE Power and Energy Magazine* 3.5 (2005), p. 34-41. DOI : 10.1109/MPAE.2005.1507024.
- [56] A. Gabbai. « Kevin Ashton describes "the internet of things" : The innovator weighs in on what human life will be like a century from now ». In : (jan. 2015).
- [57] Xiaohui Wang et al. « Power Consumption Predicting and Anomaly Detection Based on Long Short-Term Memory Neural Network ». In : *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. 2019, p. 487-491. DOI : 10.1109/ICCCBDA.2019.8725704.
- [58] Yi Song et al. « Unsupervised anomaly detection of industrial building energy consumption ». In : *Energy and Built Environment* (2024). DOI : 10.1016/j.enbenv.2024.12.005.
- [59] Danhua Liu et al. « An Unsupervised Abnormal Power Consumption Detection Method Combining Multi-Cluster Feature Selection and the Gaussian Mixture Model ». In : *Electronics* 13.17 (2024). ISSN : 2079-9292. URL : <https://www.mdpi.com/2079-9292/13/17/3446>.
- [60] Julien Lesouple et al. « How to introduce expert feedback in one-class support vector machines for anomaly detection ? » In : *Signal Processing* 188 (2021), p. 108197. ISSN : 0165-1684. DOI : <https://doi.org/10.1016/j.sigpro.2021.108197>.
- [61] Yassine Himeur et al. « Smart power consumption abnormality detection in buildings using micromoments and improved K-nearest neighbors ». In : *International Journal of Intelligent Systems* 36.7 (2021), p. 3573-3595. DOI : 10.1002/int.22459.
- [62] A. Kumar et al. « Revolutionising anomaly detection : a hybrid framework for anomaly detection integrating isolation forest, autoencoder, and Conv. LSTM ». In : *Knowledge and Information Systems* 67 (2025), p. 11903-11953. DOI : 10.1007/s10115-025-02580-6.

- [63] S. Jamshidi et al. *Lightweight Autoencoder-Isolation Forest Anomaly Detection for Green IoT Edge Gateways*. arXiv preprint arXiv :2511.18235. 2025. arXiv : 2511 . 18235 [cs.LG].
- [64] David Holtz et al. « A data-efficient active learning architecture for anomaly detection in industrial time series data ». In : *Flexible Services and Manufacturing Journal* (2025). Published online : 12 February 2025. DOI : 10 . 1007 / s10696 - 024 - 09588 - 0.
- [65] Enguerrand de Rautlin de la Roy. « Détection d'anomalies dans les données mesurées dans les bâtiments par association de connaissances expertes et de techniques d'apprentissage automatique ». Theses. Université de Bordeaux, nov. 2023. URL : <https://theses.hal.science/tel-04438194>.
- [66] Emran Altamimi et al. « Distribution-based residual analysis of load forecasting for enhanced anomaly detection ». In : *International Journal of Electrical Power Energy Systems* 173 (2025), p. 111435. ISSN : 0142-0615. DOI : <https://doi.org/10.1016/j.ijepes.2025.111435>. URL : <https://www.sciencedirect.com/science/article/pii/S0142061525009834>.
- [67] Varun Badrinath Krishna, Ravishankar K. Iyer et William H. Sanders. « ARIMA-based Modeling and Validation of Consumption Readings in Power Grids ». In : *Critical Information Infrastructures Security. 10th International Conference, CRI-TIS 2015, Revised Selected Papers*. Sous la dir. de Stephen D. Wolthusen, Marianthi Theocharidou et Erich Rome. T. 9578. Lecture Notes in Computer Science. Cham : Springer, 2016, p. 199-210. DOI : 10 . 1007 / 978 - 3 - 319 - 33331 - 1_16.
- [68] B. A. Abhale et al. « Electricity Theft Detection in Smart Grids Using Sarimax & OCR ». In : *International Journal of Advanced Research in Computer and Communication Engineering* 13.5 (2024), p. 547-550. DOI : 10 . 17148 / IJARCCCE . 2024 . 13582.
- [69] Muhammad Fahim et Alberto Sillitti. « An Anomaly Detection Model for Enhancing Energy Management in Smart Buildings ». In : *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (Smart-GridComm)*. 2018, p. 1-6. DOI : 10 . 1109 / SmartGridComm . 2018 . 8587597.
- [70] S. Shahid et al. « Anomaly detection in building energy consumption using LSTM-based prediction and statistical thresholding ». Anglais. In : *Energy and Buildings* (2025). Détection d'anomalies dans la consommation énergétique des bâtiments basée sur la prévision par LSTM et le seuillage statistique. DOI : 10 . 1016 / j . enbuild . 2025 . 116129.
- [71] Gcinizwe Dlamini et Swati Megha. « Predicting Household Power Consumption : Using Gradient Boosting and Deep Quantile Regression Model ». In : *Journal of Physics : Conference Series* 1694 (2020), p. 012031. DOI : 10 . 1088 / 1742 - 6596 / 1694 / 1 / 012031.

- [72] Rituka Jaiswal et al. « Anomaly Detection in Smart Meter Data for Preventing Potential Smart Grid Imbalance ». In : *Proceedings of the 2021 4th Artificial Intelligence and Cloud Computing Conference*. New York, NY, USA : Association for Computing Machinery, 2022, p. 150-159. ISBN : 9781450384162. DOI : 10.1145/3508259.3508281.
- [73] Cheng Fan et al. « Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data ». In : *Applied Energy* 211 (2018), p. 1123-1135. ISSN : 0306-2619. DOI : <https://doi.org/10.1016/j.apenergy.2017.12.005>. URL : <https://www.sciencedirect.com/science/article/pii/S0306261917317166>.
- [74] Basu Hela, Praveen Prasad Handigol et Pandarasamy Arjunan. « Are Time Series Foundation models good for Energy Anomaly Detection ? » In : *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*. E-ENERGY '25. Rotterdam, Netherlands : ACM, 2025, p. 656-665. ISBN : 979-8-4007-1125-1. DOI : 10.1145/3679240.3734633.
- [75] Wenping Zhang et al. « RobustProphet : Time Series Anomaly Detection Method ». In : *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*. IEEE, 2021, p. 157-161. DOI : 10.1109/CEI52496.2021.9574464.
- [76] et al. Aligholian. « Unsupervised online anomaly detection in smart building energy data : a comparative study ». In : *Proceedings of the IEEE International Conference on Smart Energy Systems and Technologies (SEST)*. 2019.
- [77] Davor Stjelja et al. « Building consumption anomaly detection : A comparative study of two probabilistic approaches ». In : *Energy and Buildings* 313 (2024), p. 114249. DOI : 10.1016/j.enbuild.2024.114249.
- [78] John Paparrizos et al. « Advances in Time-Series Anomaly Detection : Algorithms, Benchmarks, and Evaluation Measures ». In : *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. KDD '25. 2025. DOI : 10.1145/3711896.3736565.
- [79] C. C. Howard et F. N. Otobo. « Real-Time Anomaly Detection for Nigerian Power Grid Stability : An Integrated Time Series and Machine Learning Approach ». In : *American Journal of Applied Statistics and Economics* 4.1 (2025), p. 154-165. DOI : 10.54536/ajase.v4i1.5421.
- [80] Pierre Boniol et al. *VUS : Effective and Efficient Accuracy Measures for Time-Series Anomaly Detection*. Version v1. 2025. arXiv : 2502.13318. URL : <https://arxiv.org/abs/2502.13318> (visité le 04/04/2026).
- [81] « Artificial intelligence based anomaly detection of energy consumption in buildings : A review, current trends and new perspectives ». In : *Applied Energy* 287 (2021), p. 116601. ISSN : 0306-2619. DOI : <https://doi.org/10.1016/j.apenergy.2021.116601>. URL : <https://www.sciencedirect.com/science/article/pii/S0306261921001409>.

- [82] « A survey of recommender systems for energy efficiency in buildings : Principles, challenges and prospects ». In : *Information Fusion* 72 (2021), p. 1-21. ISSN : 1566-2535. DOI : <https://doi.org/10.1016/j.inffus.2021.02.002>. URL : <https://www.sciencedirect.com/science/article/pii/S1566253521000178>.
- [83] Zhun Yu, Benjamin C. M. Fung et Fariborz Haghighat. « Extracting knowledge from building-related data — A data mining framework ». In : *Building Simulation* 6.3 (2013), p. 207-222. DOI : 10.1007/s12273-013-0117-8. URL : <https://doi.org/10.1007/s12273-013-0117-8>.
- [84] Megha Gaur et al. « Performance Evaluation of Techniques for Identifying Abnormal Energy Consumption in Buildings ». In : *IEEE Access* 7 (2019), p. 62721-62733. DOI : 10.1109/ACCESS.2019.2915641.
- [85] Raffi Sevlian et Ram Rajagopal. « A scaling law for short term load forecasting on varying levels of aggregation ». In : *International Journal of Electrical Power Energy Systems* 98 (2018), p. 350-361. ISSN : 0142-0615. DOI : <https://doi.org/10.1016/j.ijepes.2017.10.032>. URL : <https://www.sciencedirect.com/science/article/pii/S0142061517306956>.
- [86] Joelle Pineau et al. « Improving Reproducibility in Machine Learning Research ». In : *Journal of Machine Learning Research* 22.164 (2021), p. 1-20. URL : <https://jmlr.org/papers/v22/20-303.html>.
- [87] Julien Herzen et al. « Darts : User-Friendly Modern Machine Learning for Time Series ». In : *Journal of Machine Learning Research* 23.124 (2022), p. 1-6. URL : <https://jmlr.org/papers/v23/21-1177.html>.
- [88] Aadyot Bhatnagar et al. « Merlion : End-to-End Machine Learning for Time Series ». In : *Journal of Machine Learning Research* 24.226 (2023), p. 1-6. URL : <https://www.jmlr.org/papers/v24/22-0809.html>.
- [89] Yue Zhao, Zain Nasrullah et Zheng Li. « PyOD : A Python Toolbox for Scalable Outlier Detection ». In : *Journal of Machine Learning Research* 20.96 (2019), p. 1-7. URL : <https://jmlr.org/papers/v20/19-011.html>.
- [90] Salesforce. *Merlion : A Machine Learning Framework for Time Series Intelligence*. 2026. URL : <https://github.com/salesforce/Merlion> (visité le 30/03/2026).
- [91] Kwei-Herng Lai et al. « TODS : An Automated Time Series Outlier Detection System ». In : *Proceedings of the AAAI Conference on Artificial Intelligence*. T. 35. 18. 2021, p. 16060-16062. URL : <https://cdn.aaai.org/ojs/18012/18012-13-21506-1-2-20210518.pdf>.
- [92] Hongzuo Xu et al. « Deep Isolation Forest for Anomaly Detection ». In : *IEEE Transactions on Knowledge and Data Engineering* (2023), p. 1-14. DOI : 10.1109/TKDE.2023.3270293.

- [93] Juan Antonio Domínguez Jiménez. « Modélisation des communautés énergétiques basée sur la cosimulation et méthodes pour leur gestion optimale distribuée avec énergie renouvelable et stockage thermique sous incertitude ». Thèse de doctorat. Université du Québec à Trois-Rivières, nov. 2025.
- [94] Alex J. Smola et Bernhard Schölkopf. « A Tutorial on Support Vector Regression ». In : *Statistics and Computing* 14.3 (2004), p. 199-222. DOI : 10.1023/B:STCO.0000035301.49549.88.
- [95] Tianqi Chen et Carlos Guestrin. « XGBoost : A Scalable Tree Boosting System ». In : *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, p. 785-794. DOI : 10.1145/2939672.2939785.
- [96] Gianluca Bontempi, Souhaib Ben Taieb et Yann-Aël Le Borgne. « Machine Learning Strategies for Time Series Forecasting ». In : *Business Intelligence*. T. 138. Lecture Notes in Business Information Processing. Berlin, Heidelberg : Springer, 2013, p. 62-77. DOI : 10.1007/978-3-642-36318-4_3.
- [97] Bharat Bohara et al. « Short-Term Aggregated Residential Load Forecasting using BiLSTM and CNN-BiLSTM ». In : *arXiv preprint arXiv :2302.05033* (2023). URL : <https://arxiv.org/abs/2302.05033>.
- [98] « A compositional kernel based gaussian process approach to day-ahead residential load forecasting ». In : *Energy and Buildings* 254 (2022), p. 111459. ISSN : 0378-7788. DOI : <https://doi.org/10.1016/j.enbuild.2021.111459>.
- [99] Wes McKinney. « Data Structures for Statistical Computing in Python ». In : *Proceedings of the 9th Python in Science Conference*. 2010, p. 56-61. DOI : 10.25080/Majora-92bf1922-00a.
- [100] Charles R. Harris et al. « Array Programming with NumPy ». In : *Nature* 585.7825 (2020), p. 357-362. DOI : 10.1038/s41586-020-2649-2.
- [101] Fabian Pedregosa et al. « Scikit-learn : Machine Learning in Python ». In : *Journal of Machine Learning Research* 12 (2011), p. 2825-2830.
- [102] Martín Abadi et al. « TensorFlow : Large-Scale Machine Learning on Heterogeneous Distributed Systems ». In : *arXiv preprint arXiv :1603.04467* (2016). arXiv : 1603.04467. URL : <https://arxiv.org/abs/1603.04467>.
- [103] Vinod Nair et Geoffrey E. Hinton. « Rectified Linear Units Improve Restricted Boltzmann Machines ». In : *Proceedings of the 27th International Conference on Machine Learning*. Omnipress, 2010, p. 807-814.
- [104] Nitish Srivastava et al. « Dropout : A Simple Way to Prevent Neural Networks from Overfitting ». In : *Journal of Machine Learning Research* 15.56 (2014), p. 1929-1958.
- [105] Diederik P. Kingma et Jimmy Ba. « Adam : A Method for Stochastic Optimization ». In : *3rd International Conference on Learning Representations (ICLR)*. San Diego, CA, USA, 2015. DOI : 10.48550/arXiv.1412.6980. arXiv : 1412.6980 [cs.LG].
- [106] Abhimanyu Das et al. « Long-Term Forecasting with TiDE : Time-series Dense Encoder ». In : *Transactions on Machine Learning Research* (2023).