

## Table des matières

<b>Résumé</b>	<b>i</b>
<b>Remerciements</b>	<b>iii</b>
<b>Table des matières</b>	<b>iv</b>
<b>Table des figures</b>	<b>viii</b>
<b>Liste des tableaux</b>	<b>xii</b>
<b>Liste des symboles</b>	<b>xii</b>
<b>Liste des acronymes</b>	<b>xiv</b>
<b>Chapitre 1 - Introduction</b>	<b>1</b>
1.1 Demande d'électricité du secteur résidentiel du Québec . . . . .	4
1.1.1 Conditions météorologiques : Température extérieure . . . . .	6
1.1.2 Caractéristiques du bâtiment . . . . .	7
1.1.3 Les habitudes de consommation . . . . .	7
1.1.4 Bénéfices de la prévision de la demande d'électricité des bâtiments . . . . .	8
1.2 Problématique . . . . .	11
1.2.1 Objectifs de la thèse . . . . .	13
1.2.2 Contributions . . . . .	14
1.2.3 Structure de la thèse . . . . .	15
<b>Chapitre 2 - État de l'art</b>	<b>17</b>
2.1 Horizon de prévision . . . . .	18
2.2 Formulation et aspects généraux du problème de prévision . . . . .	19

2.2.1	Modèles boîte blanche . . . . .	22
2.2.2	Modèles boîte grise . . . . .	24
2.2.3	Modèles boîte noire . . . . .	26
2.3	Apprentissage Statistique . . . . .	26
2.3.1	Régression paramétrique . . . . .	28
2.3.1.1	Maximum de vraisemblance et moindres carrées . . . . .	29
2.3.1.2	Modèles de régression linéaire . . . . .	30
2.3.1.3	Modèles paramétriques populaires dans le domaine . . . . .	31
2.3.1.4	Modèles paramétriques non- linéaires . . . . .	33
2.3.2	Modèles non-paramétriques . . . . .	34
2.3.2.1	Cas multivarié . . . . .	37
2.4	Synthèse du chapitre . . . . .	38
<b>Chapitre 3 - Modélisation et prévision de la composante thermosensible</b>		<b>40</b>
3.1	Introduction . . . . .	40
3.2	Description de la base de données . . . . .	41
3.3	Analyse de la relation entre la demande d'électricité et la température . . . . .	46
3.3.1	Méthodologie d'apprentissage adaptatif . . . . .	47
3.3.2	Caractéristiques de la composante thermosensible . . . . .	51
3.3.3	Évaluation de la performance thermique du bâtiment . . . . .	55
3.4	Résultats et discussions . . . . .	60
3.4.1	Méthodes de comparaison . . . . .	60
3.4.1.1	Évaluation de la précision de la prévision . . . . .	62
3.4.1.2	Résultats d'un cas d'étude . . . . .	62
3.4.2	Analyse de sensibilité . . . . .	65
3.4.2.1	Résultats d'un cas d'étude . . . . .	66
3.5	Synthèse du chapitre . . . . .	69
<b>Chapitre 4 - Modélisation et prévision de la composante résiduelle</b>		<b>71</b>
4.1	Introduction . . . . .	71
4.2	Caractéristiques de la composante résiduelle . . . . .	73
4.2.1	Incertitude agissant sur le résiduel . . . . .	73

4.2.1.1	Incertitude liée au modèle et à l'estimation des paramètres .	73
4.2.1.2	Incertitude liée aux mesures . . . . .	73
4.2.1.3	Incertitude liée aux données météo . . . . .	74
4.2.2	Données et méthodes . . . . .	75
4.2.2.1	Cyclicité dans la composante résiduelle . . . . .	75
4.2.2.2	Données de la composante résiduelle . . . . .	75
4.3	Méthodologie de modélisation de la composante résiduelle . . . . .	78
4.3.1	Modèle linéaire . . . . .	84
4.4	Cadre comparatif . . . . .	86
4.4.1	Résultats et Discussion . . . . .	88
4.4.1.1	Résultats de prévision de la composante résiduelle . . . . .	88
4.4.1.2	Résultats de prévision de la demande d'électricité . . . . .	90
4.4.1.3	Explication du second résidu . . . . .	93
4.5	Synthèse du chapitre . . . . .	97
<b>Chapitre 5 - Conclusions et recommandations</b>		<b>99</b>
5.1	Contributions principales . . . . .	99
5.2	Limites et difficultés . . . . .	101
5.2.1	Hypothèse de la température intérieure constante . . . . .	101
5.2.2	Hypothèse d'un estimateur multi-dimensionnel . . . . .	101
5.2.3	Fenêtre de lissage de l'estimateur par noyau . . . . .	101
5.3	Recommandations . . . . .	102
5.3.1	Modélisation multi-dimensionnelle . . . . .	102
5.3.2	Amélioration de l'estimateur par noyau . . . . .	102
5.3.2.1	Choix du noyau . . . . .	102
5.3.2.2	Dimension du noyau . . . . .	103
5.3.3	Explorer les limites du modèle . . . . .	103
5.3.4	Amélioration de l'explication de la composante résiduelle . . . . .	104
5.3.4.1	Sélection de l'ordre des paramètres . . . . .	104
5.3.4.2	Explication de résiduel du résiduel . . . . .	104
<b>Bibliographie</b>		<b>105</b>

<b>Appendices</b>	<b>119</b>
A Techniques de prévision . . . . .	120
A.1 Séries temporelles et régression . . . . .	120
B Les données de cinq maisons . . . . .	123
B.1 Les cinq profils horaires résidentiels réels . . . . .	123
C Notes techniques . . . . .	126
C.1 Coefficient de déperdition d'un bâtiment . . . . .	126
C.2 Fonctions des probabilités jointes et conditionnelles . . . . .	126
C.3 Évaluation de la précision de la régression . . . . .	128
C.3.1 Mesures dépendantes de l'échelle . . . . .	128
C.3.2 Erreurs en pourcentage . . . . .	129
C.3.3 Erreurs d'échelle . . . . .	129

## Table des figures

1-1	Répartition de la consommation finale d'électricité par secteur au Québec. . .	2
1-2	Évolution des maisons et de leurs équipements électriques de 1990 à 2013 au Canada [8]. . . . .	3
1-3	Répartition de la puissance d'électricité par usage au Québec [11]. . . . .	4
1-4	Comparaison du prix de l'électricité entre certaines villes d'Amérique du Nord en 2014 et 2016 [11]. . . . .	5
1-5	Les températures et les puissances d'électricité journalières moyennes [22] . .	6
1-6	Composantes d'un système de prévision de la demande d'électricité auprès du client. . . . .	9
2-1	Classification et application de la prévision de la demande d'électricité [40] .	19
2-2	Modèle régressif pour la prévision directe de la demande d'électricité résidentielle. . . . .	21
2-3	Vue d'un modèle du bâtiment créé avec le logiciel EnergyPlus+SketchUp [50].	23
2-4	Modèle RC simple du bâtiment [56]. . . . .	25
2-5	Structure d'un modèle boîte noire d'un bâtiment. . . . .	26
2-6	Comparaison de la distribution des densités pour différentes fenêtres de lissage [87]. . . . .	36
3-1	Profils de la puissance totale et de la température extérieure pour une résidence unifamiliale. . . . .	41

3-2	Les flux d'énergie dans le bâtiment. . . . .	43
3-3	Structure générale de la décomposition de la puissance d'électricité totale avec l'approche proposée. . . . .	46
3-4	Schéma du modèle Estimation Adaptative de la Densité Conditionnelle (EADC) proposé pour identifier la thermosensibilité de la consommation. . . . .	50
3-5	La densité jointe et conditionnelle de la demande d'électricité estimée : (A) jointe, (B) conditionnelle. . . . .	52
3-6	La densité jointe et conditionnelle de la demande d'électricité estimées après une transformation logarithmique : (A) jointe, (B) conditionnelle. . . . .	53
3-7	Distributions de la densité conditionnelle sachant l'écart de température : (A) données de puissance, (B) données de puissance transformées. . . . .	54
3-8	L'évolution de coefficient de déperdition $UA$ pour la maison étudiée. . . . .	57
3-9	Profils de la composante thermosensible des 5 maisons. . . . .	57
3-10	L'évolution de coefficient de déperdition $UA$ pour l'ensemble des maisons. . . . .	58
3-11	Performance des trois modèles utilisés dans la prévision de $h(x)$ sur un horizon de $\tau = 24$ heures. MLP (2 couches, 7 neurones, et un facteur d'apprentissage $\alpha = 0.012$ ), RLS ( $\mu = 0.92$ ). . . . .	63
3-12	Comparaison des résultats d'analyse de corrélation entre l'écart de température et le résiduel des trois modèles. . . . .	63
3-13	Comparaison de la précision de la prévision des modèles. . . . .	64
3-14	Impact de l'incertitude de la température sur la composante $h(x)$ . . . . .	66
3-15	Comparaison de l'analyse de corrélation des trois modèles en fonction des écarts-types des cinq stations météo à $\tau = 24$ heures. . . . .	67
3-16	Comparaison de la précision de la prévision des trois modèles en fonction des écarts-types des cinq stations météo à $\tau = 24$ heures. . . . .	67
3-17	La puissance mesurée et prédite avec les trois modèles sur un horizon $\tau = 24$ heures (A) zoom d'un mois en hiver. . . . .	68
4-1	Structure générale de l'approche de la décomposition et prévision de la demande d'électricité résidentielle. . . . .	76

4-2	Représentation d'une section de la puissance d'électricité totale, et la composante résiduelle connexe en hiver. . . . .	77
4-3	(a) : Estimateur par noyau centré en point $t_i$ , (b) Déplacement circulaire de l'estimateur par noyau vers l'événement d'utilisation d'électricité. . . . .	80
4-4	Distribution journalière/hebdomadaire de l'espérance conditionnelle au cours de l'année du dimanche au samedi, (a) résiduel sous-estimé, (b) résiduel surestimé en valeur absolue et (c) profil de la dernière semaine de décembre. . . . .	82
4-5	Distribution horaire de l'espérance conditionnelle des résiduels sous-estimé et surestimé : (b) 15 décembre, et (c) 30 juin. . . . .	83
4-6	Schéma bloc de la modélisation et prévision de la composant résiduelle. . . . .	86
4-7	Performance des modèles adaptatifs dans la modélisation et la prévision de la composante résiduelle sur un horizon de $\tau = 24$ heures. AR ( $\rho = 0.98$ ), ARX ( $\lambda = 0.92$ ) . . . . .	89
4-8	Comparaison de la précision de la prévision de la composante résiduelle en fonction d'horizon de prévision. . . . .	90
4-9	Comparer la demande d'électricité mesurée et prédite par les trois modèles sur l'horizon $\tau=24$ heures . . . . .	91
4-10	Comparaison de la précision de la prévision de la demande d'électricité réalisée avec les trois modèles en fonction d'horizon de prévision. . . . .	92
4-11	Comparaison de l'amélioration de la prévision entre les trois modèles . . . . .	93
4-12	Représentation d'une section de la composante résiduelle, et le résiduel $r_{res,k}$ en hiver. . . . .	94
4-13	Autocorrélation de résiduel du résiduel sous-estimé et surestimé . . . . .	95
4-14	Autocorrélation de résiduel sous-estimé et surestimé et leurs interactions avec les charges de chauffe-eau et les autres charges. . . . .	96
4-15	Matrice de corrélation pour évaluer la dépendance entre les charges électriques et les seconds résidus sous-estimé et surestimé. . . . .	96
2-1	Profils de température et de puissance totale pour la résidence unifamiliale 1	123
2-2	Profils de température et de puissance totale pour la résidence unifamiliale 2	124

2-3	Profils de température et de puissance totale pour la résidence unifamiliale 3	124
2-4	Profils de température et de puissance totale pour la résidence unifamiliale 4	125
2-5	Profils de température et de puissance totale pour la résidence unifamiliale 5	125
3-1	Densité jointe et conditionnelle. . . . .	127

## Liste des tableaux

1-1	Évolution de la consommation d'énergie résidentielle au Québec (2001-2016), en (TWh) [7]. . . . .	2
2-1	Quelques exemples de noyaux classiques [105] . . . . .	36
3-1	Valeurs moyennes des constantes $UA$ de déperdition thermique de l'ensemble des maisons . . . . .	58
3-2	Les écarts-types moyens de la température extérieure fournis à partir des cinq stations météo [33]. . . . .	65

## Liste des symboles

Terme	Description
$Y_{el}$	Puissance des autres charges électriques (W).
$k$	Indice du temps discret.
$\delta$	Distribution normale.
$Y_{ewh}$	Puissance de chauffe-eau (W).
$Y_{ehs}$	Puissance de chauffage électrique (W).
$w, \Theta$	Vecteurs des coefficients.
$\Sigma$	Matrice de covariance.
$t_k$	Temps calendaire (heure) .
$\mathcal{D}_x$	Domaine de la variable $x$ .
$N$	Nombre d'observations.
$\mathbb{R}_+$	Ensemble des réels positifs.
$e_k$	Erreur de prévision (W).
$x^{int}$	Température intérieure ( $^{\circ}C$ ).
$\alpha$	Facteur d'apprentissage.
$x^{ext}$	Température extérieure ( $^{\circ}C$ ).
$x_b$	Température d'équilibre ( $^{\circ}C$ ).
$x'^{ext}$	Température bruitée ( $^{\circ}C$ ).
$R$	Matrice d'autocorrélation inverse.
$\mathcal{K}$	Fonction de noyau (Kernel).

<b>Terme</b>	<b>Description</b>
$\lambda$	Facteur d'échelle.
$\mu$	Facteur d'oubli.
$\sigma_{x_{ext}}^2$	Variance de la température ( $^{\circ}C$ ).
$r_k^-$	Résiduel surestimé à l'instant $k$ , (W).
$T$	Période temporelle (heure).
$r_k^+$	Résiduel sous-estimé à l'instant $k$ , (W).
$r_{res,k}^+$	Second résidu sous-estimé à l'instant $k$ , (W).
$r_{res,k}^-$	Second résidu surestimé à l'instant $k$ , (W).
$\nu$	Fenêtre de lissage.
$x_k$	Vecteur concaténant les variables exogènes.
$Y$	Puissance d'électricité totale (W).
$y_k$	Mesure ponctuelle de la puissance à l'instant $k$ , (W).

## Liste des acronymes

<b>Terme</b>	<b>Description</b>
AA	Apprentissage Automatique.
AR	Autorégressif.
ARIMA	Autoregressive integrated moving average.
ARIMAX	Autoregressive integrated moving average with eXternal inputs.
ARMA	Autorégressif et moyenne mobile.
ARMAX	Autoregressive–moving-average with eXternal inputs.
ARX	Autoregressive with eXternal inputs.
EADC	Estimation Adaptative de la Densité Conditionnelle.
ECA	Espérance Conditionnelle Adaptative.
GARCH	Generalized AutoRegressive Conditional Heteroskedasticity.
GDD	Gestion De la Demande.
GLDD	Gestion Locale De la Demande.
IA	Intelligence Artificielle.
KDE	Kernel Density Estimation.
LTE	Laboratoire des Technologies de l'Énergie.

<b>Terme</b>	<b>Description</b>
LTLF	Long Term Load Forecasting.
MA	Moyenne mobile.
MAP	Maximum A Posteriori.
ML	Modèle Linéaire.
MLE	Maximum Likelihood Estimation.
MLP	Multi-Layer Perceptron.
MTLF	Medium Term Load Forecasting.
MVS	Machines Vectorielles de Support.
NMAE	Erreur Absolue Moyenne Normalisée.
PRISM	Princeton Scorekeeping Method.
REI	Réseaux Électriques Intelligents.
RNA	Réseaux de Neurones Artificielles.
STLF	Short Term Load Forecasting.
VSTLF	Very Short Term Load Forecasting.

## Chapitre 1 - Introduction

Les compagnies pétrolières et les industries de production d'énergie s'intéressent à la prévision de la demande pour des raisons qui peuvent varier considérablement selon leurs besoins en gestion et en planification. La prévision de la demande d'électricité est considérée comme un processus fondamental dans de nombreuses décisions de production et de gestion d'énergie et de puissance. C'est un outil d'aide à l'anticipation et la détermination de la demande future, facteur clé dans un processus de maîtrise de la disponibilité de l'énergie et la minimisation des coûts d'exploitation [1]–[3]. La prévision de la demande d'électricité est un domaine de recherche actif, en particulier, dans le secteur résidentiel. Ce dernier est caractérisé par une forte demande d'électricité qui nécessite une Gestion Locale De la Demande (GLDD) avec une grande efficacité. À l'échelle canadienne, la consommation d'énergie moyenne annuelle des bâtiments résidentiels, selon les statistiques obtenues par l'enquête sur les ménages et l'environnement pour l'année 2013, se chiffrait à 30 MWh [4]. Cela justifiait le fort intérêt de déployer la prévision de l'énergie pour assurer l'équilibre offre-demande d'électricité et ainsi réduire la consommation et le coût énergétique.

Dans la province de Québec, où s'est déroulée cette étude, l'électricité provenant des barrages hydroélectriques est la source d'énergie la plus utilisée pour chauffer les bâtiments résidentiels. Cette électricité représente environ 71% de la consommation finale totale d'énergie [5].

Selon les statistiques sur la demande en électricité fournies par le Ministère de l'Énergie et des Ressources Naturelles pour l'année 2013, le résidentiel est l'un des secteurs les plus énergivores après le secteur industriel. Il accapare une portion de 38.6% de la consommation finale d'électricité fournie par Hydro-Québec [6]. Le reste se partage entre les secteurs commerciaux et les transports. La Figure. 1-1 expose la répartition de la consommation finale

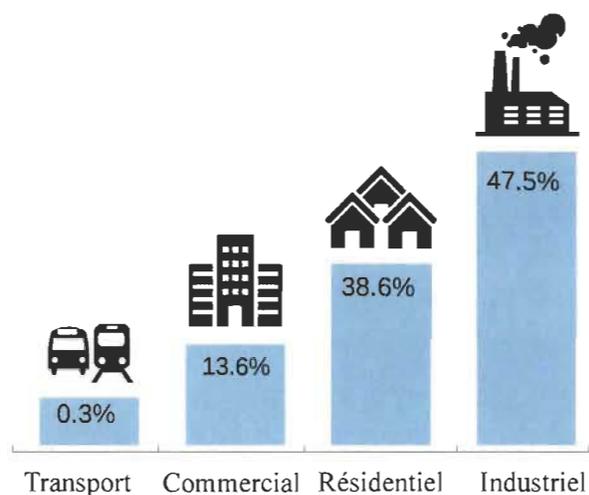


FIGURE 1-1 Répartition de la consommation finale d'électricité par secteur au Québec.

d'électricité, en pourcentage, par secteur en 2013 au Québec. Cette consommation a connu une croissance de 0.7% de 58.5% à 62.2% durant les années 2001 à 2016 [7], comme l'illustre le tableau 1-1 ci-dessous.

TABLEAU 1-1 Évolution de la consommation d'énergie résidentielle au Québec (2001-2016), en (TWh) [7].

	2001	%	2011	%	2016	%	Taux de croissance (%)
Électricité	53.6	58.5	59.0	61.4	60.0	62.2	0.7
Gaz naturel	6.6	7.2	6.0	6.3	5.5	5.7	-1.2
Pétrole	17.4	19.0	15.0	15.7	14.6	15.2	-1.1
Biomasse	14.0	15.3	15.9	16.6	16.3	16.9	1.0
TOTAL	91.7	100.0	96.1	100.0	96.5	100.0	0.3

Cela s'explique en grande partie par l'évolution des surfaces habitables. Par exemple, durant les années 90 la surface moyenne des maisons était de  $122\text{ m}^2$ , alors que durant les années 2000 les surfaces moyennes ont augmenté jusqu'à  $142\text{ m}^2$ . Ainsi que, l'évolution sociale qui a engendré un surnombre d'appareils électriques [8]. Par exemple, le confort social a suscité plus d'appareils ménagers par maison (voir la Figure. 1-2) [9]. De plus, les clients ont substitué graduellement les sources d'énergie traditionnelles à savoir le bois, charbon, mazout et le gaz naturel par l'énergie électrique.

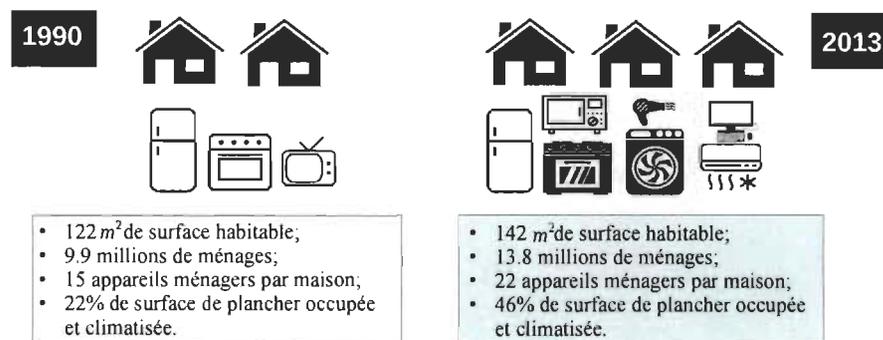


FIGURE 1-2 Évolution des maisons et de leurs équipements électriques de 1990 à 2013 au Canada [8].

Dans le contexte québécois, l'hiver est caractérisé par des changements brusques de la température extérieure. La vague de froid exceptionnelle qui s'est emparée de la région en janvier 2014 a entraîné des pics de puissance records, avec une pointe de puissance de 39 031 MW [10]. Durant ces périodes, le gestionnaire de réseau<sup>1</sup> est fortement sollicité pour faire face aux défis de la hausse de la demande [11].

Afin de répondre à cette problématique, il devient nécessaire pour le gestionnaire de réseau d'adopter de nouvelles stratégies de GLDD applicables aux bâtiments résidentiels, en évaluant les occupants et leurs habitudes de consommation énergétique [12]. Alors dans le cadre du programme de Réseaux Électriques Intelligents (REI), l'amélioration de l'efficacité énergétique des bâtiments et la mise au point d'un modèle explicatif et prévisionnel permet l'anticipation et la compréhension des phénomènes qui suscitent les variations de la puissance d'électricité.

Dans ce cas, le REI propose de mettre en place les produits technologiques existants dans le marché pour soutenir la gestion de l'énergie électrique dans les résidences et mieux anticiper l'équilibre entre l'offre et la demande pendant les périodes de pointe auprès du client. Cela permet de renforcer l'intégration des énergies renouvelables telles que les panneaux

1. Le gestionnaire du réseau, a un rôle primordial dans la distribution des énergies aux consommateurs. Il a comme mission d'assurer le fonctionnement du réseau face aux pics de consommation et la maintenance du réseau de distribution de l'électricité et du gaz [11].

photovoltaïques pour du consommateur un acteur ou un producteur majeur d'électricité à partir de nombreuses petites sources d'énergie [13]–[15].

Pour comprendre le déploiement du REI dans le contexte québécois, tout en plaçant le client au centre de ce processus, on présente dans la section 1.1 les raisons qui poussent le Québec à instaurer cette technologie dans les bâtiments résidentiels.

### 1.1 Demande d'électricité du secteur résidentiel du Québec

La particularité de la puissance électrique dans les bâtiments au Québec fait appel à la compréhension de la facture annuelle des clients résidentiels. Selon les statistiques sur la répartition de la puissance électrique par usage, données par Hydro-Québec [11], la puissance d'électricité se décompose de la façon suivante : le chauffage des locaux prend la plus grande partie de la facture d'électricité avec plus de 54% de la consommation d'électricité par utilisation. Tandis que, le chauffe-eau peut représenter jusqu'à 20% de la facture d'énergie, particulièrement pour ceux ayant une grande famille qui consomme beaucoup d'eau chaude. Les appareils ménagers consomment 18%, cette puissance se répartit comme suit : le réfrigérateur 33%, la sècheuse 18%, la cuisinière 15%, et autres charges. Enfin, l'éclairage et autres appareils ne consomment globalement que 5% et 3%, respectivement [11], [16]. La Figure. 1-3, illustre l'électricité utilisée dans une résidence, en pourcentage, par usage en 2009 au Québec.

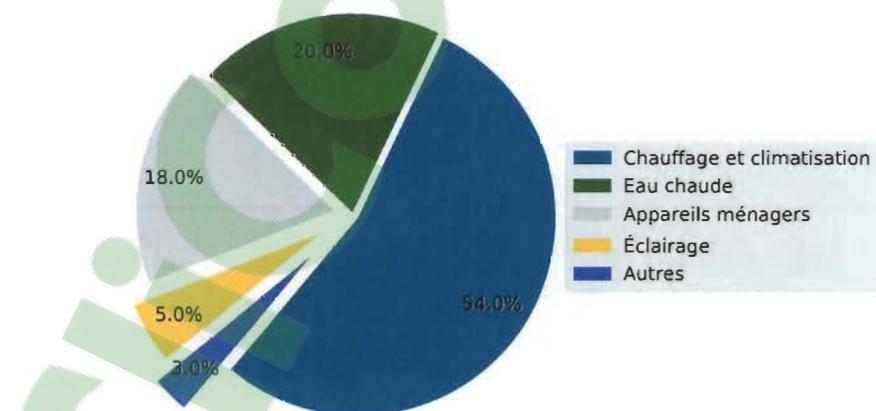


FIGURE 1-3 Répartition de la puissance d'électricité par usage au Québec [11].

Bien que le chauffage de l'espace est la principale charge qui a un impact significatif sur la facture d'électricité pendant les hivers froids, les tarifs d'électricité demeurent concurrentiels selon les données fournies par la société d'état en 2014. C'est d'ailleurs au Québec qu'on trouve le tarif le moins cher par kilowattheure, comparé à la plupart des provinces canadiennes et des États Unis [17]–[19]. La Figure. 1-4, montre une comparaison des prix moyens de l'électricité pour certaines villes d'Amérique du Nord pour les années 2014 et 2016. Sur cette figure, les tarifs d'électricité moyens de l'année 2016 varient de 7.23 cents/kWh à Montréal à 29.52 cents/kWh à New York. En outre, pour la plupart des villes, le prix d'électricité a augmenté entre 2014 et 2016 d'environ 4.01 cents/kWh à Toronto, 3.6 cents/kWh à Chicago et 3.22 cents/kWh à Seattle, alors que Montréal jouit favorablement d'un prix très bas comparativement aux autres villes. Or, selon l'Union des consommateurs du Québec, il est difficile d'établir une comparaison équitable de la facture totale d'électricité. Ceci est dû à la différence des sources d'énergie et aussi aux conditions climatiques des différentes provinces [17].

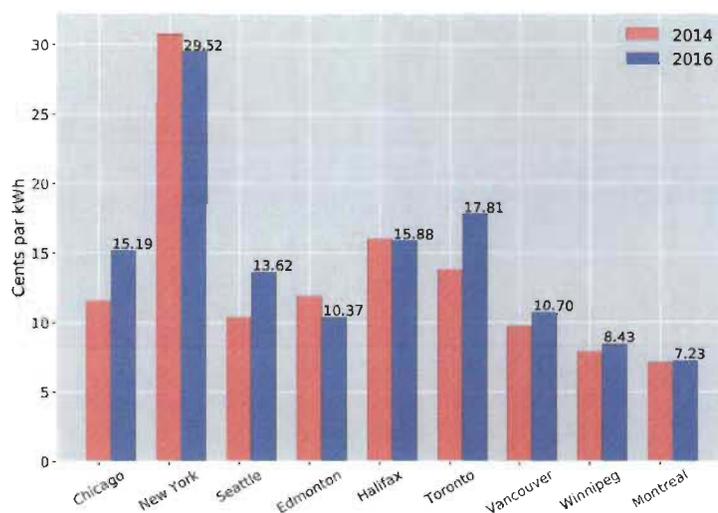


FIGURE 1-4 Comparaison du prix de l'électricité entre certaines villes d'Amérique du Nord en 2014 et 2016 [11].

Ce qu'il faut considérer surtout pour indiquer si un client québécois paie cher ou pas, c'est le coût investi par Hydro-Québec pour la production et le transport jusqu'au point de branchement, notamment pendant la période de forte demande. Or, ceci coûte deux fois moins cher au Québec de produire l'électricité que dans la plupart des états et provinces avec lesquelles la comparaison est effectuée [19], [20].

La problématique de la prévision de la demande d'électricité dans les résidences situées dans les régions nordiques comme le Québec met en jeu des facteurs divers et variés qui rendent la puissance fluctuante. Cette dernière est déterminée principalement par trois facteurs : les conditions météorologiques de la région (principalement la température extérieure), les caractéristiques du bâtiment et les habitudes de consommation. D'autres facteurs, en revanche, interviennent indirectement dans le processus de prévision, par exemple, les incertitudes des prévisions des conditions météorologiques, les incertitudes sur le comportement des habitants et les modèles choisis, les contraintes liées aux données mesurées et à la fréquence d'échantillonnage, etc [21]. La présence de ces incertitudes entraîne une volatilité dans la précision de la prévision de la demande. Dans cette sous-section, nous décrirons en particulier les trois premiers facteurs clés qui font varier la puissance électrique.

### 1.1.1 Conditions météorologiques : Température extérieure

Parce que la position géographique du Québec est soumise à un grand écart de température dans une journée, lors des périodes de grand froid, la température extérieure a un impact très important sur la demande énergétique de chauffage électrique. L'appel de puissance de ces charges s'étale de novembre à mars, avec une demande d'électricité qui occupe environ 60%. La Figure. 1-5 montre un exemple réel des variations journalières moyennes de la température et la puissance totale d'une maison localisée à Montréal.

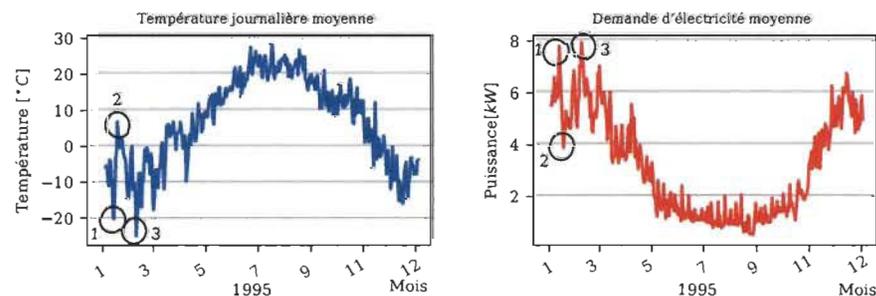


FIGURE 1-5 Les températures et les puissances d'électricité journalières moyennes [22]

Les pics importants de la demande d'électricité (pics encerclés en noir) correspondent aux

pics des températures journalières moyennes. Cette situation particulière influence profondément la demande d'électricité des occupants québécois et augmente à la fois la demande et les variations des pics en puissance.

### *1.1.2 Caractéristiques du bâtiment*

Au Québec, une grande majorité des bâtiments résidentiels sont faits avec une ossature de bois qui est privilégiée par rapport à celle en acier en raison des besoins d'isolation imposés par les baisses de température. Pour cette raison, le bâtiment québécois doit être muni d'un système d'isolation thermique de qualité pour assurer une consommation énergétique efficace.

Dans cette thèse, il a été choisi de considérer principalement l'analyse de niveau d'isolation thermique globale qui nous sera utile dans le chapitre 3. Celle-ci peut entraîner des pertes de chaleur qui peuvent atteindre environ 43% du total des pertes de chaleur. Et les fuites d'air causées par infiltrations peuvent représenter jusqu'à 25% des pertes de chaleur. Tandis que, les portes et les fenêtres peuvent représenter jusqu'à 32% des pertes de chaleur [11], [23].

### *1.1.3 Les habitudes de consommation*

La présence et les habitudes des occupants ont aussi un impact considérable sur la variation de la demande d'électricité des bâtiments résidentiels. Le comportement et les activités aléatoires des occupants figurent parmi les facteurs explicatifs des écarts entre la consommation réelle mesurée et les prévisions. Dans ce sens, le nombre d'occupants a ainsi une incidence sur le degré d'utilisation des appareils électriques énergivores. Les phénomènes comportementaux causés par les activités et le mode de vie des occupants sont directement liés à la variable temporelle, et incluent la tendance saisonnière, les tendances hebdomadaires, journalières et horaires. Ces tendances peuvent caractériser les changements dans la demande entre les week-ends, les jours fériés et les jours ouvrables. Dans certains cas, ces tendances sont utilisées pour identifier le calendrier temporel sur les actions telles que, l'utilisation des appareils électriques, ainsi que la gestion des consignes de chauffage. L'introduction de l'effet temporel ou calendaire dans la prévision de la demande pourrait permettre d'élaborer des scénarios réalistes du temps d'utilisation d'électricité liée à l'activité des occupants. Ces scénarios admettent une meilleure

compréhension de la demande résidentielle d'électricité des bâtiments, mais également une gestion plus précise de cette demande [24].

#### *1.1.4 Bénéfices de la prévision de la demande d'électricité des bâtiments*

Les prévisions de la demande d'électricité faite dans le contexte du REI peuvent se faire au niveau du gestionnaire du réseau et aussi au niveau du consommateur. Toutefois, la plupart des études de modélisation et de prévision ont été réalisées à l'échelle de l'utilité qui est propriétaire de toutes les informations de puissance agrégée de l'ensemble des maisons, de quartier et voir de ville.

Par contre, la réalisation de cette prévision au niveau du client apporte plusieurs avantages, car il est possible d'agir sur la gestion locale de sa demande. À titre d'exemple :

- Permet au client de prendre des décisions en temps-réel sur la gestion de sa puissance [25], et ce, afin d'aider ce dernier à corriger ou changer ses habitudes et optimiser sa demande en électricité. Par exemple, un rapport détaillé sur la consommation thermosensible<sup>2</sup> pourra aider l'utilisateur à détecter une demande de puissance, en proposant une amélioration de l'isolation de sa résidence, ou en changeant la consigne de la température intérieure, etc.
- Permet d'améliorer le pouvoir explicatif qui a été souvent négligé dans la prévision de la demande et ceci par une dissociation de l'ensemble des facteurs qui causent une puissance élevée. Par exemple, la connaissance de la dépendance entre la puissance totale et la température extérieure peut aider le client à expliquer la portion de la puissance due seulement à la température, en excluant les effets du vent et le soleil.
- Permet de déterminer les moments et les activités qui sont à la base de certaines surconsommations afin d'aider le client à mieux planifier sa demande, ou intégrer d'autres sources d'énergies renouvelables [27].

Dans ce cas, l'utilité doit profiter de cette nouvelle situation pour laquelle des millions de clients peuvent passer d'un statut de consommateur à celui de fournisseur et gestionnaire de

---

2. Thermosensibilité d'un bâtiment correspond à l'effet de fluctuation de la puissance en fonction de la température [26]. Ce phénomène est caractérisé par le gradient de température qui s'explique essentiellement par l'usage de charges de puissance électrique, afin de maintenir le bâtiment à une température constante [21].

leur demande d'énergie, et cela permettrait d'alléger la charge sur le réseau et de diminuer la dépendance à l'égard des opérateurs [27].

Toujours auprès du client, voici trois sources de données qui peuvent être utilisées afin de réaliser la prévision de la demande d'électricité en exploitant l'infrastructure de mesurage avancée existante (Figure. 1-6) :

- Compteur électrique intelligent ;
- Informations générées par les charges électriques ;
- Données météorologiques.

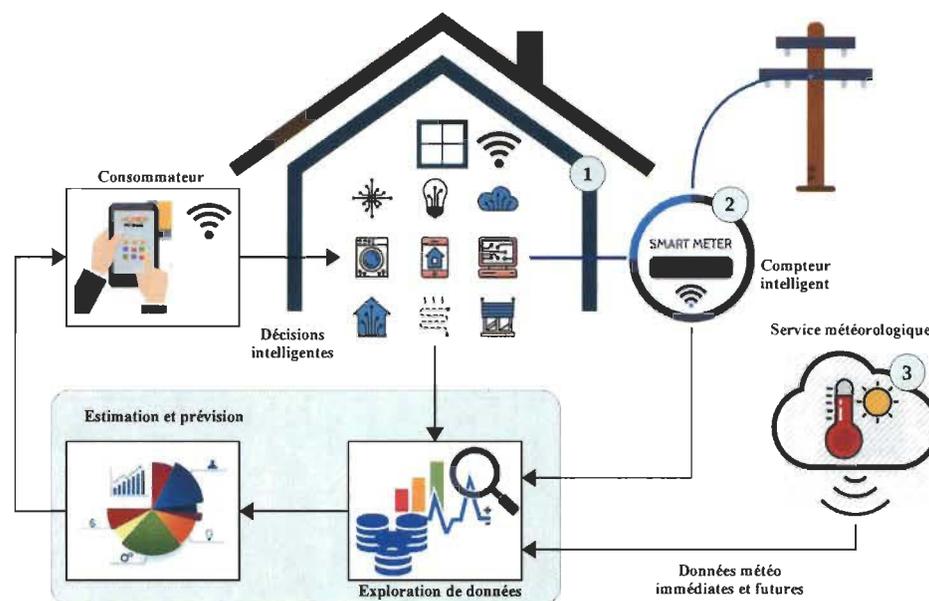


FIGURE 1-6 Composantes d'un système de prévision de la demande d'électricité auprès du client.

*Compteur électrique intelligente* : Avec la venue de ce produit technologique intelligent, plus de 4.8 millions de compteurs intelligents ont été installés dans des maisons en Ontario, en 2014 [28], et 3.75 millions dans toute la province de Québec d'ici 2018 [29]. Avec cette source de données, il devient possible de communiquer dans les deux sens, vers les clients et vers l'utilité. Il peut recevoir des informations qui viennent du gestionnaire de réseau de distribution, comme il peut aussi recevoir de l'information relative aux clients. Avec le compteur intelligent

qui mesure et analyse l'information en temps-réel, un grand nombre de données de l'énergie consommée deviennent disponibles. L'ensemble de ces données contribue à une meilleure compréhension de la demande électrique du bâtiment et également à une gestion plus précise de la demande [30]. De plus, le client peut connaître sa consommation en kilowattheures en temps-réel de jour en jour et même d'heure en heure [31], au lieu d'une consommation cumulative tous les deux mois.

En outre, les compteurs intelligents offrent la possibilité, à l'avenir, de recueillir les données directement auprès du client, plutôt que les demander à l'utilité. Pour, le marché de l'énergie électrique par exemple : OpenEnergyMonitor<sup>3</sup> propose d'utiliser des capteurs moins chers qui mesurent la quantité d'énergie qui passe directement à travers le compteur. Un exemple de ces capteurs Optical Utility Meter LED Pulse Sensor à 34 CAD peut lire exactement ce que le compteur mesure [32].

*Informations générées par les charges électriques :* Il y a une autre direction dans le domaine de REI à partir de laquelle il est possible de générer de l'information pour réaliser la prévision de la puissance d'électricité. Elle utilise des thermostats communicants qui fournissent les données de la température intérieure de la maison sans à installer des capteurs de mesure. Il permet de transmettre l'information de la température directement au système de prévision de la demande. Ainsi, il aide à piloter le chauffage électrique localement ou à distance par une liaison WiFi. En effet, les données des températures intérieures collectées permettent de calculer l'écart de température qui contribue à l'analyse de la thermosensibilité des résidences.

*Les données météorologiques :* Les données météorologiques immédiates et futures, telles température extérieure, la vitesse du vent et ensoleillement, constituent la troisième source d'informations. Ces données sont très utiles pour la prévision de la demande, parce que leur source est fiable et leur précision est plus au moins exacte comparée aux prévisions simulées ou faites localement, ce qui les rend utilisables avec les modèles de prévision de la demande avec des incertitudes quantifiées. Ces données ainsi que leurs incertitudes sur différents horizons

---

3. <https://openenergymonitor.org/>

sont fournies par les stations météorologiques, à savoir : Environment Canada<sup>4</sup>, Weather Underground<sup>5</sup>, Weather Network<sup>6</sup> [33], entre autres.

## 1.2 Problématique

Cette thèse s'articule autour du problème de la modélisation et de la prévision à court terme de la demande électrique dans le cadre du programme de gestion d'énergie dans le secteur résidentiel. Notre intérêt porte sur un modèle explicatif et prédictif de la puissance totale auprès du client sur un horizon de vingt-quatre heures, à partir de prospection de données.

Dans la littérature scientifique, de nombreuses approches basées sur l'exploration de données sont proposées afin de construire des modèles automatiques prédictifs de la puissance électrique résidentielle. Cependant, pour réaliser les prévisions, la plupart de ces approches proposent d'enrichir les modèles par un ensemble de facteurs, par exemple, le vent, l'humidité, l'ensoleillement et la température, plutôt que de faire une analyse de la contribution de chaque facteur sur les variations de la demande par un modèle de pouvoir explicatif. Or, la gestion locale de la demande d'électricité totale dépend de la précision de sa prévision. Cette dernière s'améliore si on considère un modèle très détaillé à partir des données plus descriptives du bâtiment et les mesures météorologiques.

Actuellement, il existe différentes approches physiques basées sur des logiciels de simulation énergétique détaillée permettant de comprendre la modélisation et de prédire la puissance électrique. Cependant, des problèmes liés à la mise en place d'un système de prévision automatisé les rendent moins intéressantes à savoir :

1. Complexité : les outils de simulation énergétique nécessitent des compétences en thermique du bâtiment, et une connaissance accrue de la maison étudiée [34]. En cas d'erreur ou d'oubli dans l'ajout de ces informations, il y aura un impact négatif sur les performances des résultats.
2. Intrusion : ces approches peuvent demander des informations sur la connaissance de la taille, des dimensions, de l'orientation du bâtiment, en plus d'une enquête sur le

---

4. <https://www.weather.gc.ca/>

5. <https://www.wunderground.com/>

6. <https://theweathernetwork.com/ca/>

nombre de résidents, ce qui est difficilement réalisable dans le cas d'un parc résidentiel, voir un quartier.

### 3. Coût d'accès aux informations :

- Obtenir des informations détaillées spécifiques au bâtiment avec ces méthodes requiert un déplacement de techniciens pour la collecte des données [34]. Il faut savoir que cette intervention coûte cher pour le fournisseur d'électricité.
- Les besoins de déploiement et d'installation d'un système de mesurage avancé pour accéder à la puissance et aux données météorologiques de bonne qualité nécessitent des ressources dispendieuses. Par contre, cela coûtera moins cher si le consommateur exploite les infrastructures existantes incluant les compteurs intelligents, les thermostats communicants et certaines stations météo gratuites.

Par ailleurs, la littérature scientifique a proposé une famille d'approches statistiques facilitant l'exploration et l'analyse des données sans le besoin des caractéristiques physiques du bâtiment. Bien que ces modèles soient non-intrusifs, leur application se concentre sur les prévisions pour servir le gestionnaire du réseau en fournissant une valeur future de demande en électricité, plutôt qu'aider le client à comprendre sa consommation afin qu'il puisse décider et gérer sa propre demande. En effet, rares sont les modèles qui se focalisent sur la réalisation des prévisions auprès du consommateur de manière adaptative avec un minimum de mesures. Dans ce cadre, les modèles actuels proposent des prévisions basées sur l'apprentissage hors ligne à partir d'une grande base de données de mesures de puissance et données météorologiques. Le problème avec cet apprentissage est que les estimateurs nécessitent plusieurs années de données mesurées pour former le modèle, réaliser l'apprentissage, tester et valider. Une fois ces étapes réalisées, le modèle commence à fournir les résultats d'analyses. Par contre, les modèles adaptatifs permettent d'introduire les informations préalables sur le profil de la puissance et apprennent au fur et à mesure que les nouvelles observations arrivent. Ceci va aider le modèle à s'adapter aux nouvelles dynamiques de la demande électrique à chaque nouvelle observation.

Or, la mise en action de ces scénarios prévisionnels auprès du client peut rendre la problématique de la modélisation complètement différente. Ceci nécessite de disposer d'un modèle de pouvoir explicatif basé sur une compréhension optimale de la consommation

thermosensible du bâtiment et une connaissance moindre des données intrusives qui permettent une meilleure anticipation de la consommation d'électricité totale. Toutefois, dans le contexte québécois où la demande en électricité est très sensible à la température extérieure, il faut donc trouver un moyen pour compléter le modèle prévisionnel avec un modèle explicatif en introduisant :

1. Les aléas de la température extérieure pour expliquer la consommation thermosensible due principalement à l'utilisation du chauffage d'espace affectées directement par les changements des températures, ainsi que les autres usages tels que la climatisation et la piscine en été ;
2. La variable temporelle sur une base de l'effet calendaire pour expliquer la puissance causée par d'autres phénomènes tels que les activités et le mode de vie des occupants ;
3. L'aspect adaptatif pour suivre à chaque instant les changements en besoin d'énergie produite par les variations climatiques et comportementales des occupants et passer d'une version figée de l'explication de la demande à une version évolutive permettant une compréhension et un suivi de manière adaptative.

### *1.2.1 Objectifs de la thèse*

Être de plus en plus explicatif et précis dans les prévisions de la demande d'électricité résidentielle, en estimant les valeurs futures de la consommation, constitue la motivation de cette thèse. L'objectif primordial de ce travail est de proposer une approche explicative et prédictive capable de décrire de manière adaptative la demande en électricité. C'est dans ce contexte que ce projet de recherche s'est défini, il vise à construire un cadre de modélisation et de prévision de la demande totale à court terme, qui repose sur l'exploration des relations entre les données mesurées. Étant donné que, ces relations contiennent de l'information cachée sur la consommation, on essaye d'extraire la consommation thermosensible et celle qui se base sur l'effet calendaire afin d'améliorer les connaissances du client sur sa demande.

Divers sous objectifs ont été fixés :

- Conception d'un modèle d'exploration des données de type boîte noire appliqué aux maisons réelles, qui peut identifier et reproduire le comportement thermique du bâtiment face aux excitations météorologiques ;

- Développement d'une approche non-paramétrique explicative et prédictive, capable de décomposer la puissance totale en deux composantes, et identifier les fluctuations de la demande électrique dues aux aléas de la température extérieure ;
- Proposition d'une approche combinée (non-paramétrique et paramétrique) qui prend en compte l'aspect comportemental de l'occupant à partir de la variable calendaire et identifier comment le client consomme en fonction de l'heure de la journée, et du jour de la semaine.

### 1.2.2 Contributions

#### *A : La composante thermosensible de la demande électrique*

1. Développement d'une approche non-paramétrique adaptative qui permet l'apprentissage adaptatif du comportement thermique du bâtiment et le suivi des changements de la demande d'électricité.
2. Extraction de la thermosensibilité de la demande d'électricité québécoise produite par l'aléa de la température. Le comportement thermique et énergétique de la résidence permet de calculer le coefficient de déperdition du bâtiment qui contribue à l'évaluation du niveau d'isolation global, et cela, sans aucune information préalable des caractéristiques thermiques des maisons étudiées.

*B : La composante résiduelle de la demande électrique :* La composante de la demande non explicable par la température extérieure représente la composante non thermosensible appelée aussi "résiduelle" qui définit l'écart entre la demande totale mesurée et la part thermosensible. La modélisation et la prévision de cette composante ont été élaborées pour expliquer la partie non déterministe de la demande. Cette partie est caractérisée par la périodicité de quelques phénomènes qui sont fortement influencés par l'effet calendaire, relié au temps d'utilisation de l'électricité.

La démarche adoptée dans cette thèse consiste à élaborer une modélisation de la composante résiduelle de la puissance d'électricité totale qui vise à :

1. Diviser la composante résiduelle en deux séries temporelles, une positive et l'autre négative, et proposer un nouveau mécanisme de modélisation de ces deux résidus, en

tenant en compte de la cyclicité des phénomènes non déterministes environnementaux et comportementaux.

2. Proposer une approche non-paramétrique basée sur l'estimation de l'espérance conditionnelle des résiduels afin de repérer en ligne le calendrier d'utilisation périodique d'électricité, sans connaissance préalable de la relation entre la puissance électrique connexe et le temps d'utilisation de l'énergie.
3. Développer une procédure adaptative capable de réaliser un apprentissage en-ligne en suivant les changements non-linéaires du profil des résidus sous-estimés et surestimés qui varient selon les habitudes de vie et les technologies nouvelles installées dans les résidences.

### *1.2.3 Structure de la thèse*

Le contenu du chapitre 1 constitue la démarche du développement de la méthodologie de la modélisation et de prévision à court terme de la demande d'électricité.

Le chapitre 2 présente un état de l'art sur les modèles de prévision à court terme de la demande d'électricité. On y trouve d'abord une section qui décrit brièvement les quatre catégories de l'horizon de prévision, suivie par une section sur les aspects généraux du problème de la prévision. Nous abordons dans cette section la façon dont l'apprentissage statistique est appliqué au problème de la modélisation énergétique d'un bâtiment. Une autre section a été adressée aux trois approches de modélisation qui regroupe les modèles physiques, boîtes noires et grises. Nous justifions aussi le choix de l'approche appliquée dans cette étude et nous présentons la structure de l'ensemble des modèles proposés pour la prévision de la demande. Le chapitre 3 propose l'étude de la themosensibilité de la demande québécoise en se basant sur un minimum de données pour analyser la relation entre la demande des résidences réelles, localisées à Montréal, et la température extérieure. Cette analyse a été réalisée à l'aide d'un modèle explicatif non-paramétrique basé sur l'estimateur par noyau gaussien mono-variable et multi-variables, qui ne requiert aucune connaissance préalable des caractéristiques thermiques des résidences.

Une évaluation de la performance thermique de la résidence a été effectuée en considérant une transformation logarithmique sur les données de demande électrique. Une autre section a

été consacrée à l'analyse de la sensibilité du modèle afin de prendre en compte les incertitudes sur les données d'entrées. Il est en effet nécessaire de tester la robustesse du modèle aux incertitudes sur la prévision de la température, en propageant les incertitudes fournies par cinq stations météo sur la température réelle et voir leurs impacts sur la thermosensibilité. Ensuite une analyse comparative de la performance de la prévision de la composante thermosensible vis-à-vis celles obtenues avec les modèles de réseaux de neurones et les modèles linéaires a été réalisée.

Au chapitre 4, une approche combinée de modélisation et de prévision de la composante résiduelle a été présentée de manière à assurer son couplage avec l'approche de prévision de la composante thermosensible présentée au troisième chapitre. L'idée originale de ce chapitre est d'extraire les périodicités de la composante résiduelle relatives à l'utilisation des appareils électriques et aux comportements stochastiques des occupants. Ensuite la méthodologie qui permet d'établir l'approche proposée a été présentée, en introduisant un modèle non-paramétrique adaptatif basé sur l'estimateur par noyau gaussien cyclique, en rapport avec l'effet calendaire. Par la suite un modèle de régression paramétrique a été employé. Ce modèle utilise en entrée la tendance-cyclique identifiée du résiduel pour prédire une composante résiduelle simulée. Le chapitre se termine par des comparaisons des modèles proposés avec d'autres méthodes connues dans l'analyse des résiduels. Ces comparaisons ont été réalisées et discutées afin d'évaluer la précision de la prévision du résiduel et de la demande d'électricité totale. Dans le chapitre 5, les principales contributions, les limites de notre recherche et les perspectives envisagées pour la poursuite de ce travail ont été synthétisées au sein d'une conclusion générale.

## Chapitre 2 - État de l'art

Depuis l'évolution industrielle successive et suite aux demandes accrues en matière d'énergie électrique, les prévisions sont devenues une nécessité impérative afin de répondre aux besoins futurs en électricité et maintenir l'équilibre avec la production [35]. Soumises en permanence à de nombreux facteurs, la production et la puissance électrique évoluent en tout temps [36]. Pour mieux anticiper l'équilibre entre l'offre et la demande et acheminer l'électricité nécessaire à toutes les résidences à chaque instant, le fournisseur d'électricité utilise essentiellement l'historique de la puissance et les données météo relatives à la température, la vitesse du vent et l'ensoleillement. À la fin, c'est le comportement de chaque occupant qui détermine le niveau de puissance et la demande d'énergie réellement atteints [24], et ceci à travers ses activités et ses utilisations des appareils électriques installés dans la résidence.

Bien que la capacité de stockage de l'énergie électrique est limitée dans les bâtiments, de nombreuses incertitudes sur la gestion de la demande et l'économie de la puissance d'électricité résidentielle sont mises en question. La plus importante incertitude concerne l'utilisation des appareils énergivores et l'horizon auquel nous pouvons envisager une gestion locale efficace d'énergie pendant les heures de pointe journalière. Pour couvrir ces incertitudes, plusieurs approches et méthodes ont été élaborées selon les besoins en prévision à court, moyen et à long terme, et ainsi en fonction de l'application.

Parmi ces méthodes figurent celles de l'apprentissage statistique qui sont appliquées dans les projets de prévision de la demande d'électricité à échelle de l'utilité [15]. Pour le côté client, ces modèles peuvent être employés aussi en utilisant l'information de la puissance générée et consommée, avec une fréquence des mesures faible. Dans ce contexte local, les fluctuations de la puissance électrique réagissent rapidement aux conditions climatiques et aux comportements non déterministes des occupants. Puisque notre analyse portera principalement

sur la prévision à court terme, nous présentons dans ce chapitre un panorama général sur les méthodes de modélisation et prévision de la demande d'électricité, identifiées dans la littérature comme les techniques les plus importantes pour la prévision à court terme de la demande énergétique.

## 2.1 Horizon de prévision

Avant de présenter les modèles de prévision de la demande d'électricité, il est souligné que la prévision peut être divisée en quatre catégories en fonction de l'horizon temporel :

- prévision à très court terme (Very Short Term Load Forecasting (VSTLF), en anglais) pour les analyses avec un horizon de moins d'une heure (p.ex. aux 15 minutes) à l'avance [1].
- prévision à court terme Short Term Load Forecasting (STLF) qui est de quelques heures à un jour à l'avance. Cet horizon est nécessaire à l'anticipation des charges électriques énergivores pendant les heures de pointe et à la gestion locale de la demande journalière [1], [30]. Notre étude se concentre sur la prévision à court terme, auxquels cas nous pouvons envisager et anticiper une GLDD dans le bâtiment résidentiel.
- prévision à moyen terme Medium Term Load Forecasting (MTLF) qui est généralement de quelques semaines. Cet horizon est utilisé pour planifier la maintenance et les équipements de l'industrie de production d'électricité, ainsi pour envisager les ventes prévues, et les budgets à dépenser pour acheter les machines aidant à augmenter la production [1], [3].
- prévision à long terme Long Term Load Forecasting (LTLF) qui est de plus d'un an. L'horizon est appliqué par les industries d'électricité pour gérer les réserves d'énergie en planifiant d'augmenter leurs capacités de production, transmission et distribution d'énergie électrique [1], [3], [37]–[39].

La Figure. 2-1 montre les quatre classes de prévision de la demande électrique. Il est à noter que toutes les catégories d'horizons nécessitent essentiellement l'historique de la demande électrique, les informations météorologiques (selon la disponibilité des informations), mais aussi les données économiques et démographiques pour le cas des prévisions à long terme.

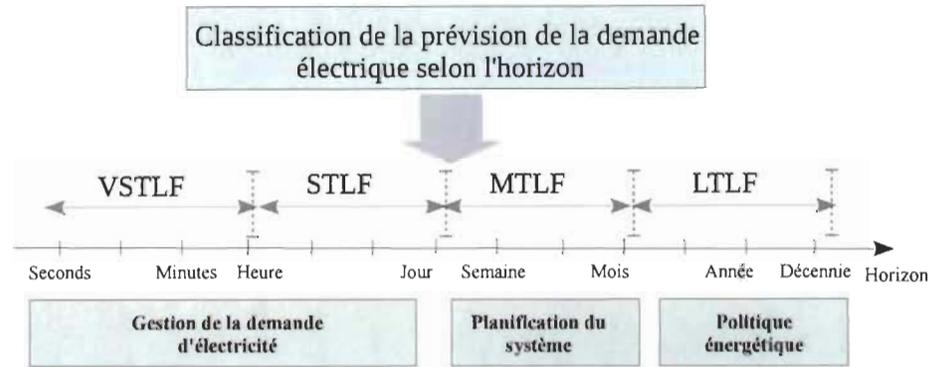


FIGURE 2-1 Classification et application de la prévision de la demande d'électricité [40]

## 2.2 Formulation et aspects généraux du problème de prévision

De façon générale, un modèle de prévision a pour objectif de simuler l'évolution temporelle de la variable à prédire sur un horizon donné. Au départ, il y a les équations qui décrivent la prévision comme un problème de régression. Cette dernière est très importante en apprentissage statistique lorsque l'objectif est d'identifier la fonction qui permet l'estimation de la variable concernée à partir de l'historique des observations ou des mesures.

Afin d'accomplir la prévision de la demande d'électricité, il faut modéliser d'abord la puissance qui est une variable quantitative à partir de l'historique des données disponibles et les variables explicatives qui sont aussi quantitatives. Pour cela, il existe deux familles de modèles : le premier utilise seulement le passé des observations de la variable à expliquer elle-même et le deuxième utilise l'ensemble des variables explicatives.

1. modèles auto-régressifs ou uni-variés Autorégressif (AR);
2. modèles avec variables exogènes Autoregressive with eXternal inputs (ARX).

Avant de définir les deux modèles, nous introduisons la notation mathématique générale pour le problème de modélisation, ainsi que pour la description des méthodes AR et ARX de l'état de l'art. Pour cela, la nomenclature suivante est adoptée :

- $k$  correspond à l'index du temps discret ;
- $\hat{y}_k$  est une estimation de la mesure  $y_k$  ;
- $\tau \in \mathbb{Z}$  est un horizon de prévision ;

- $\hat{y}_{k+\tau}$  est une estimation à l'horizon  $\tau$  ;
- $y_k = \{y_{k-i}\}_{i=0,\dots,N}$  est un vecteur contenant les  $N + 1$  dernières mesures à partir du temps présent  $k$  ;
- $x_k^j = \{x_{k-M}^j\}_{j=0,\dots,M}$  : dénote les valeurs de la variable explicative à l'instant  $k$  ;
- $x_k$  est un vecteur concaténant toutes les entrées du modèle ciblant une estimation  $\hat{y}_k$  ou  $\hat{y}_{k+\tau}$ . Ce vecteur est construit de la façon suivante :

$$x_k = \begin{bmatrix} x_k^1 \\ \vdots \\ x_{k-N}^1 \\ \vdots \\ x_k^M \\ \vdots \\ x_{k-N}^M \end{bmatrix} \quad (2-1)$$

Les modèles autorégressifs exploitent les dépendances temporelles entre les données d'une série temporelle sans étudier la dépendance avec les variables explicatives. C'est à dire, la variable exogène est liée aux observations de la puissance électrique  $y_k$  elle-même et son historique. Mathématiquement, il faut chercher une fonction de prévision de la forme  $\hat{y}_{k+\tau} = f(y_k)$ . Par contre, les modèles avec des variables exogènes expliquent la relation entre les facteurs quantitatifs et la puissance elle-même. Parmi les facteurs les plus utilisés dans la prévision de la demande sont :

- Le temps calendaire ;
- Les données météorologiques :
  - température extérieure
  - ensoleillement
  - vent
- Les données provenant des systèmes de mesurage et des charges communicantes (par exemple : le compteur intelligent et le thermostat communicant).

En effet, en fonction de la disponibilité des variables, deux types de modèles avec des variables exogènes peuvent être considérés dans un problème de prévision :

1. Modèle régressif pour une prévision directe : scénario de la relation entre le passé des variables explicatives et le présent ou le futur de la variable à prédire, comme illustré dans la Figure. 2-2. Dans cette dernière, les facteurs exogènes sont représentés par l'ensemble des données  $x_k$ . Dans ce cas, il faut chercher à estimer la fonction de la prévision de la forme  $\hat{y}_{k+\tau} = f(x_k, y_k)$ .

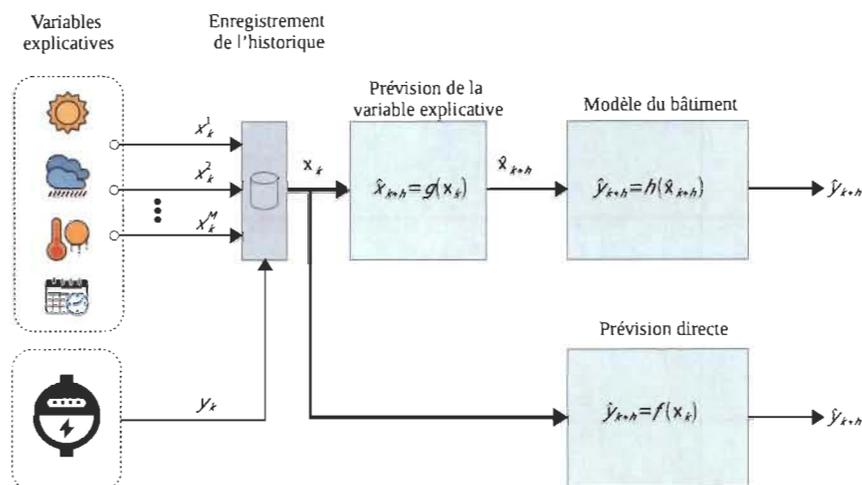


FIGURE 2-2 Modèle régressif pour la prévision directe de la demande d'électricité résidentielle.

Le modèle direct considère une forte corrélation entre le présent des facteurs explicatifs et le futur de la variable d'intérêt. Ce qui est évidemment difficile à assurer, surtout avec des facteurs météorologiques hautement stochastiques comme la température extérieure, le soleil et la vitesse du vent. Cependant, lorsque le lien entre la puissance et un facteur mesurable dans le présent est connu, l'incertitude de ce facteur dans le futur va affecter directement la qualité de la prévision de la puissance.

2. Modèle régressif pour une prévision indirecte : pour ce scénario, les prévisions des variables exogènes fournies par les services météorologiques sont utilisées au lieu des mesures locales. Pour ce cas, la disponibilité de données exogènes futures est fondamentale. Étant donné l'élan des technologies de l'information et de la communication, il est de plus en plus facile d'accéder aux prévisions météorologiques de haute précision.

En termes d'information, le modèle direct doit être équivalent au modèle indirect si une prévision locale des facteurs exogènes est employée. C'est-à-dire, dans le cas où les prévisions météorologiques ne sont pas disponibles, il est possible de construire un modèle de prévision autorégressif pour les variables exogènes de la forme  $\hat{x}_{k+h} = g(x_k)$ . Par la suite les prévisions peuvent être utilisées pour estimer le futur de la puissance en suivant la chaîne  $\hat{y}_{k+\tau} = h(\hat{x}_{k+\tau}) = h(g(x_k))$ . Ceci peut fonctionner aussi par la construction d'un modèle de prévision direct  $\hat{y}_{k+\tau} = f(x_k)$ .

Généralement, le travail de modélisation de la puissance consiste à trouver une expression mathématique cohérente avec la nature des données disponibles et le comportement du système représenté par le bâtiment résidentiel. En effet, un modèle adapté à la problématique de prévision de la demande d'électricité doit considérer les aspects suivants :

- l'incertitude des variables exogènes ;
- le comportement peu déterministe des occupants ;
- la disponibilité des données de prévision météorologique.

Parmi un éventail assez large des méthodes de régression, nous présentons les modèles de prévision à court terme qui ont été développés et testés au cours des dernières décennies. Les méthodes de prévision les plus utilisées peuvent être classées en (i) Modèles physiques de type *boîte blanche*, (ii) Modèle de type *boîte noire* et (iii) modèles de type *boîte grise*. Ces modèles sont présentés aux sections 2.2.1, 2.2.2 et 2.2.3. L'étude de la littérature des modèles prédictifs populaires couvrant les trois groupes sont présentés dans [41]–[43]. En effet, ces trois types de modèles ont été classifiés, et comparés afin de déterminer les caractéristiques et les limites liées à la modélisation de la demande d'électricité résidentielle.

### 2.2.1 Modèles boîte blanche

La modélisation physique de type *boîte blanche* est basée sur les lois physiques pour estimer la demande en électricité résidentielle [44]. Elle est principalement utilisée lors de la phase de conception, par exemple pour le dimensionnement des équipements (CVCA<sup>1</sup>). Les modèles physiques utilisent un ensemble complet d'équations basées sur la géométrie détaillée des zones thermiques à modéliser, les propriétés des matériaux, et les caractéristiques

1. Chauffage, ventilation et conditionnement de l'air

énergétiques des bâtiments afin de modéliser les charges CVCA et décrire avec précision la dynamique du bâtiment [30].

Les modèles physiques ont été implémentés dans plusieurs logiciels de simulation de bâtiment, certains sont présentés dans [45], [46]. Les outils de simulation les plus utilisés sont EnergyPlus, ESP-R, et TRNSYS [30], [47], [48]. Ces logiciels se basent sur les propriétés physiques des bâtiments pour simuler la consommation d'électricité et estimer la température intérieure dans une maison sur différentes échelles du temps (année, mois, jour ou heure) [49]. De plus, la simulation peut être effectuée pour différents endroits dans le monde avec des fichiers météo spécifiques à la ville choisie. Un exemple de vue d'un bâtiment virtuel réalisé avec le simulateur EnergyPlus+SketchUp est illustré dans la Figure. 2-3.

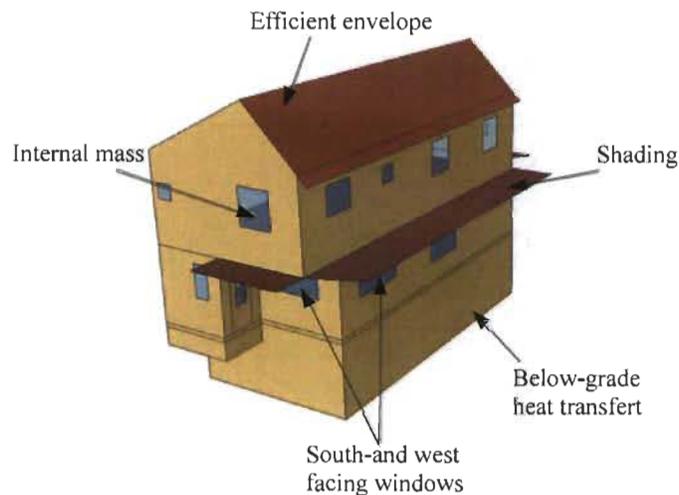


FIGURE 2-3 Vue d'un modèle du bâtiment créé avec le logiciel EnergyPlus+SketchUp [50].

Ces modèles sont appliqués pour prédire le comportement thermique d'un bâtiment, y compris la demande d'électricité. Ils sont considérés comme les seules méthodes capables de modéliser la puissance d'électricité sans la nécessité de l'information historique de la demande [21]. Toutefois, leur efficacité est limitée en raison de l'énorme temps et coût qu'il faut pour que l'utilisateur puisse obtenir et définir tous les paramètres et les caractéristiques du bâtiment [51]. En outre, des informations spécifiques et détaillées sont requises sur la structure du bâtiment, l'équipement CVCA installé, les conditions météorologiques, l'emplacement et l'orientation de la résidence [30], [51].

Pour une étude sur la prévision de la demande d'électricité d'un bâtiment résidentiel existant, les modèles physiques sont intrusifs dans le sens où leurs calibrations requièrent beaucoup d'informations et ceci ne correspond pas à l'objectif de la thèse qui repose sur une prévision avec un minimum de capteurs et de mesures. Par conséquent, les modèles physiques sont écartés pour ce travail, ainsi, ils sont inadéquats pour décrire les comportements des occupants. Ces derniers sont modélisés dans les simulateurs boîte blanche par des valeurs représentatives, ou par des scénarios déterministes [24]. Or, dans le cas réel, l'occupation ne peut pas être prédéfinie par un profil identique. Pour cette raison, les modèles physiques ne peuvent pas être retenus comme une solution à la problématique de caractérisation de l'occupation qui varie largement et de manière imprévisible.

La modélisation inverse qui englobe les modèles boîtes noires et grises est une solution alternative pour l'analyse de la performance énergétique des bâtiments existants. Elle est adaptative, et plus facile à paramétrer que les modèles physiques [48].

### 2.2.2 Modèles boîte grise

Les modèles *boîte grise* ou semi-physiques ont un concept intermédiaire qui consiste à combiner la *modélisation physique* pour simuler le comportement thermique du bâtiment et les *modèles boîte noire* que nous abordons dans 2.2.3. Les modèles boîte grise sont physiquement compréhensibles et s'adaptent à la réalité des mesures. Ils ont été introduits au début des années 90 pour des applications spécifiques de contrôle automatique des systèmes et l'étude d'optimisation [47]. Cette catégorie de modèles consiste à modéliser le bâtiment par un circuit électrique équivalent [52]. Dans ce cas, ces modèles impliquent la connaissance des paramètres thermiques, tels que la capacité et la résistance thermique de différents ordres (RC, R2C2, R3C2, ...). Les modèles boîte grise reposent sur la modélisation de l'enveloppe du bâtiment par analogie électrique, qui permet de comprendre la structure du modèle par une observation rapide du circuit équivalent. Dans la Figure. 2-4 nous présentons un modèle RC simple d'un bâtiment, montrant la résistance et la capacitance thermique équivalente. Les modèles boîte grise ont été appliqués également dans l'implémentation des systèmes de contrôle, de la prévision et pour l'amélioration de la demande en électricité dans le bâtiment en utilisant des stratégies d'économie d'énergie [47], [53], [54]. En effet, ils peuvent être

appliqués dans les cas d'estimation de la demande électrique due au chauffage et au système de refroidissement ainsi que dans l'évaluation de la température intérieure [52], [55]. Un

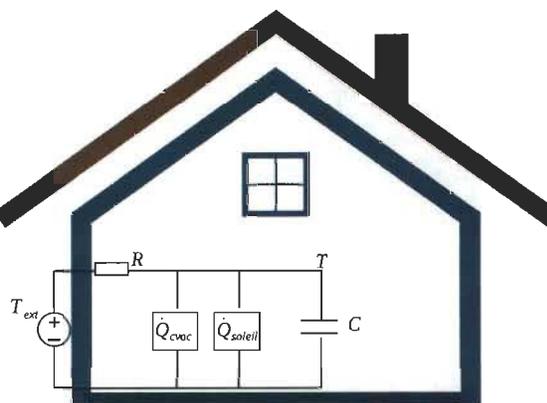


FIGURE 2-4 Modèle RC simple du bâtiment [56].

modèle par analogie thermique-électrique a été développé en [57] pour prédire la charge de climatisation et la température intérieure, dont le comportement thermique du bâtiment a été modélisé avec treize résistances thermiques et huit capacités (13R8C). Différentes combinaisons de résistance thermique et de capacité sont considérées pour les murs internes et externes, les plafonds et les planchers et les fenêtres. Il a été combiné avec un modèle de régression non-linéaire pour ajuster les paramètres RC. Un autre travail de Swan et al [58], qui ont utilisé les modèles physiques pour modéliser le chauffage et la climatisation mixés avec les modèles de réseaux de neurones pour modéliser la demande de l'eau chaude sanitaire [21].

Les modèles boîte grise peuvent être considérés dans le cas où un modèle physique du bâtiment est disponible, notamment lorsqu'il s'agit de la modélisation des bâtiments existants où généralement il est difficile de reconstruire un modèle physique détaillé [59], dans telle situation les modèles boîte grise pourraient s'appliquer. Ils permettent aussi de décrire les paramètres thermiques et de garder une signification physique des paramètres du modèle de bâtiment. En effet, ils nécessitent des moyens informatiques moins lourds que ceux requis par les modèles de boîte blanche, car ils présentent une meilleure flexibilité [52]. Cependant, certains inconvénients sont propres à ces modèles, comme la nécessité d'une double compétence pour l'utilisateur en thermique du bâtiment et des connaissances des modèles boîtes noires.

### 2.2.3 Modèles boîte noire

Les modèles *boîte noire* permettent d'estimer la demande en électricité des équipements CVCA et la puissance totale du bâtiment à partir des mesures uniquement. Leur précision dépend de la quantité et de la qualité des données disponibles. Ils peuvent aussi modéliser les phénomènes non-linéaires et s'adapter aux changements de la demande d'électricité par un apprentissage adaptatif aux nouvelles données d'observations [52]. La Figure. 2-5 montre un exemple représentatif de ces modèles qui utilise la base de données qui contient l'historique des observations des variables mesurées, en appliquant l'intelligence artificielle pour produire la sortie d'intérêt [30]. Bien que les modèles de boîte noire ne donnent aucune interprétation

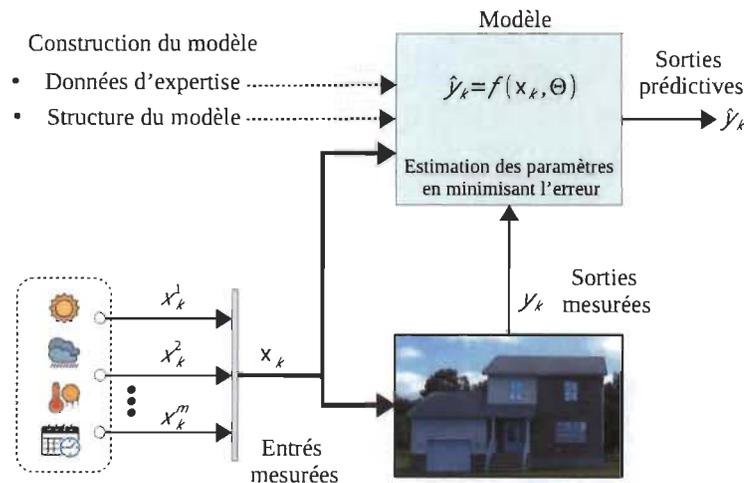


FIGURE 2-5 Structure d'un modèle boîte noire d'un bâtiment.

physique des résultats du bâtiment étudié [59], [60], ils sont très présents dans la modélisation et la prévision de la demande d'électricité. Ils utilisent l'apprentissage statistique qui représente un champ d'études de l'intelligence artificielle.

## 2.3 Apprentissage Statistique

Dans le contexte de la modélisation boîte noire, les modèles statistiques sont largement utilisés pour la prévision de la puissance électrique et les besoins en chauffage/refroidissement dans un bâtiment. Ils analysent la relation entre la demande d'électricité et les facteurs

explicatifs [61]. Une définition formelle du modèle est donnée par :

$$y_k = m(x_k, \theta) + e_k \quad (2-2)$$

Où  $\theta$  est un ensemble de paramètres et  $e_k$  dénote l'erreur avec des hypothèses quelle est distribuée normalement avec une moyenne nulle et une variance  $\sigma$  et qu'elle soit non-corrélée avec les variables explicatives.

Quel que soit le domaine d'application, le problème de modélisation doit faire face à deux tâches majeures :

1. Sélection du modèle : plusieurs modèles pour  $m(x_k, \theta)$  peuvent être utilisés pour expliquer les données observées [62];
2. Apprentissage : problème d'estimation des paramètres  $\theta$  permettant d'expliquer au mieux une ou plusieurs réalisations avec des hypothèses sur l'erreur [62].

Le problème de prévision de la demande d'électricité est complexe et influencé par la sélection du modèle. La proposition du bon modèle va aider à traduire le comportement thermique et énergétique du bâtiment par des équations mathématiques des phénomènes internes et externes qui font appel à la puissance électrique.

Dans la littérature, autant de modèles de prévision prennent la forme d'une régression conduisent à des estimations bien différentes de  $m(x_k, \theta)$  et qui génèrent des évolutions distinctes de la puissance électrique en fonction des hypothèses supposées sur les données et sur la fonction  $m()$ . En outre, pour l'apprentissage, l'estimation des valeurs des paramètres  $\theta$  qui expliquent le mieux les données peut être une tâche plus ou moins difficile en fonction du type de modèle et des données disponibles. Pour cela, il faut chercher la structure du modèle qui colle le mieux possible avec la variable d'intérêt  $y_k$  et ceci avec des paramètres bien ajustés permettant de minimiser le résiduel ou l'erreur  $e_k$ . Cependant, en pratique, il est quasiment impossible de trouver exactement les  $y_k$  à cause de présence d'une combinaison d'incertitudes liées implicitement au choix du modèle, à l'ajustement des paramètres ou bien incertitudes liées aux données utilisées pour l'estimation. Donc en termes de résiduel, il s'agit d'estimer les paramètres de façon à minimiser l'erreur.

Dans l'ensemble, l'intérêt récent pour employer les techniques d'apprentissage automatique permettant une meilleure gestion de l'incertitude a débouché sur l'adoption de façon presque systématique des méthodes probabilistes, plus spécifiquement de l'inférence statistique. En effet les travaux rapportés dans [63]–[66] ont prouvé l'applicabilité de ces approches dans la modélisation de la demande d'électricité résidentielle. Justement, l'apprentissage automatique pour la modélisation paramétrique consiste à formuler un problème inverse dont les paramètres des modèles seront estimés de façon automatique à partir des techniques d'exploration des données. Pour se servir de ces concepts, les méthodes paramétriques et non-paramétriques qui sont deux types de modèles de régression utilisés dans l'apprentissage statistique sont décrites.

### 2.3.1 Régression paramétrique

Dans le cadre d'apprentissage statistique, les méthodes paramétriques supposent des hypothèses et des connaissances préalables sur la forme de la liaison entre la demande électrique  $y_k$  et les facteurs  $x_k$  [67]. Dans ce contexte, la régression vise à déterminer les valeurs pour  $\theta$  qui expliquent le mieux la variable  $y_k$  à partir de la relation  $\mathbb{E}(y_k|x_k, \theta) = m(x_k, \theta)$ . C'est-à-dire, la valeur espérée de la demande électrique  $y_k$  pourrait être identifiée en estimant la moyenne en fonction de la distribution conditionnelle  $p(y_k|x_k, \theta)$  [68].

En outre, la démarche de la régression paramétrique se base sur deux estimateurs de modélisation de la demande résidentielle d'électricité. Elle consiste à utiliser le Maximum Likelihood Estimation (MLE) et l'estimateur Maximum A Posteriori (MAP) [69], [70]. À partir de l'ensemble de mesures  $\{\mathbf{x}, \mathbf{y}\} = \{(y_k, x_k) | k = 1 \dots N\}$  [71], l'estimateur MLE est à la base de l'analyse statistique classique qui repose sur la loi des observations [64]. Ceci propose d'estimer les paramètres à partir de la maximisation directe de la vraisemblance  $P(\mathbf{y}|\mathbf{x}, \theta)$ . Cet estimateur permet de cerner dans le cadre statistique quelques techniques de régression bien connues dans la littérature telles que les moindres carrés ordinaires. Le lien entre MLE et les moindres carrés sera présenté en 2.3.1.1.

Par ailleurs, l'estimateur MAP, issue de l'inférence bayésienne formule le problème de régression en considérant les paramètres en tant que des variables aléatoires. Sur ce point, les paramètres du modèle sont estimés à partir de la maximisation de la loi postérieure  $P(\theta|\mathbf{x}, \mathbf{y})$

[64]. Pour ce faire, l'estimateur MAP est défini comme :

$$\hat{\theta}_s = \arg \max_{\theta} (p(\mathbf{y}|\mathbf{x}, \theta)p(\theta)) \quad (2-3)$$

Le type de solution du problème de l'équation (2-3) dépend en grande partie de la complexité du modèle  $m(\mathbf{x}_k, \theta)$  et des hypothèses faites sur les distributions de probabilité. Avant d'approfondir l'apprentissage bayésien, une brève description des autres estimateurs très utilisés dans le domaine est présentée.

### 2.3.1.1 Maximum de vraisemblance et moindres carrées

Les modèles de régression utilisent souvent une loi gaussienne pour  $p(y_k|\mathbf{x}_k, \theta)$  et une loi uniforme pour  $p(\theta)$ . Sous ces hypothèses l'estimateur MAP généralise l'estimateur MLE (priori uniforme) et l'estimateur à moindres à carrées (vraisemblance gaussienne) [72], [73]. En effet, si  $p(y_k|\mathbf{x}_k, \theta) = \mathcal{N}(y_k|m(\mathbf{x}_k, \theta), \sigma)$ , la vraisemblance du modèle est donnée par :

$$p(\mathbf{y}|\mathbf{x}, \theta) = \prod_{k=1}^N \mathcal{N}(y_k|m(\mathbf{x}_k, \theta), \sigma) \quad (2-4)$$

$$= \frac{1}{\sigma \sqrt{2\pi}} \prod_{k=1}^N \exp\left(-\frac{(y_k - m(\mathbf{x}_k, \theta))^2}{2\sigma^2}\right) \quad (2-5)$$

Toutefois, la maximisation de ce produit est compliquée du point de vue analytique, il est donc préférable de transformer le produit en une somme qui est plus pratique pour les calculs et ensuite maximiser le logarithme de  $p(\mathbf{y}|\mathbf{x}, \theta)$  [74]. Dans ce cas, la valeur de  $\theta$  est recherchée via la formulation suivante :

$$\hat{\theta} = \arg \max_{\theta} (\mathcal{L}(\theta; \mathbf{y}, \mathbf{x})) \quad (2-6)$$

Où  $\mathcal{L}(\theta; \mathbf{y}, \mathbf{x}) = \ln(p(\mathbf{y}|\mathbf{x}, \theta))$ . Pour le cas de la vraisemblance gaussienne proposée par l'équation (2-4) [75], [76], ceci est égale à :

$$\mathcal{L}(\theta; \mathbf{y}, \mathbf{x}) = -\frac{1}{2\sigma^2} \sum_{k=1}^N (y_k - m(\mathbf{x}_k, \theta))^2 - \ln(\sigma \sqrt{2\pi}) \quad (2-7)$$

Dans ce cas, la méthode de vraisemblance dispose d'une hypothèse où l'erreur suit une distribution normale, et que la valeur de  $\sigma$  est constante. La solution du problème (2-6) consiste alors à minimiser exactement la somme de l'erreur quadratique du critère des moindres carrée (MCO, ou OLS en anglais) [75], [77] qui s'écrit :

$$\hat{\theta} = \arg \min_{\theta} \left( \sum_{k=1}^N (y_k - m(x_k, \theta))^2 \right) \quad (2-8)$$

Maintenant à partir des échantillons de la puissance électrique et des variables explicatives, le cas de la régression paramétrique linéaire et non-linéaire utilisée dans la modélisation de la demande est considéré.

### 2.3.1.2 Modèles de régression linéaire

Le modèle de régression simple considère une relation fonctionnelle simple entre les entrée  $x_k$  et la sortie  $y_k$  à partir d'un vecteur de coefficients  $w$  :

$$y_k = w\phi(x_k) + e_k \quad (2-9)$$

Où  $\phi(x_k)$  est une fonction sur les variables explicatives. La relation entre les variables est linéaire et s'applique aux paramètres  $w$  du modèle et non aux variables d'entrée. Dans le cas le plus simple,  $e_k$  est considéré comme un bruit gaussien de distribution normale et une solution est recherchée pour  $w$  afin de maximiser la fonction de l'équation suivante [75] :

$$\frac{\partial \mathcal{L}(w; \mathbf{y}, \mathbf{x})}{\partial w} = \sum_{k=1}^N (y_k - w\phi(x_k))\phi(x_k) = 0 \quad (2-10)$$

Maximiser la fonction définie par (2-7) débouchant dans l'équation (2-10) configure la méthode de moindres carrés ordinaires pour la régression linéaires [78].

$$\hat{w}_{OLS} = (\phi(\mathbf{x})^T \phi(\mathbf{x}))^{-1} \phi(\mathbf{x})^T \mathbf{y} \quad (2-11)$$

### 2.3.1.3 Modèles paramétriques populaires dans le domaine

Plusieurs modèles de régression paramétriques ont été appliqués dans le domaine de la prévision de la demande en électricité du bâtiment de manière adaptative. Nous citons par exemple le filtre de Kalman qui utilise l'algorithme RLS pour la mise à jour des paramètres. Il existe aussi la méthode de Princeton Scorekeeping Method (PRISM) qui est une technique de régression linéaire qui analyse la corrélation entre la demande d'électricité et la température extérieure [30]. Elle a montré une bonne performance en estimation du coefficient de déperdition du bâtiment [79]. D'autres modèles de régression existants dans [80]–[83] utilisent la régression linéaire par morceaux avec deux variables climatiques : les degrés jours de chauffage (HDD) et les degrés-jours de climatisation (CDD) lorsqu'elles sont disponibles [21].

En outre, nous présentons brièvement l'analyse classique de séries temporelles sur lesquelles se fonde une partie de cette étude. Les différentes techniques de série temporelle qui décrivent la valeur future de la demande d'électricité par une combinaison linéaire des observations précédentes de la puissance. Deux processus linéaires sont à la base des séries temporelles, le processus de la Moyenne mobile (MA) sur les erreurs et le processus AR qui est utilisé lorsque la valeur à prévoir est corrélée avec les valeurs précédentes [84].

Ces deux modèles peuvent être fusionnés en un processus linéaire de modèles Autorégressif et moyenne mobile (ARMA) également connu sous le nom de modèle Box-Jenkins. Les modèles ARMA sont développés dans [85] pour prévoir le prix de l'électricité sur une semaine à venir en se basant uniquement sur leurs valeurs horaires passées. Les modèles de série temporelle ARMA et Autoregressive integrated moving average (ARIMA) peuvent être étendus pour prendre en compte plusieurs variables exogènes. En effet, en présence des variables exogènes ces deux modèles sont nommés respectivement Autoregressive–moving-average with eXternal inputs (ARMAX) et Autoregressive integrated moving average with eXternal inputs (ARIMAX) [30]. Le modèle ARMAX est développé dans [86] pour prévoir la charge électrique quotidienne de la République Tchèque jusqu'aux deux prochains jours, le modèle prend en compte la température extérieure. Nous avons mis dans l'annexe A l'accent plus en détail sur les techniques de prévisions qui comprennent, la méthode de MA, le modèle

AR, modèle ARMA, ARX, et ARMAX.

Les modèles de séries temporelles montrent en général une bonne performance dans l'évaluation des tendances stationnaires de la puissance électrique par rapport à un horizon court de prévision [84]. Cependant, ils dévoilent certaines limitations s'il existe des dépendances non-linéaires entre la puissance électrique et les variables explicatives, et ceci en raison de la nature stochastique inhérente de la demande. Ainsi, plus le nombre de régresseurs dans les séries temporelles est grand, plus le risque de déterminer les bons paramètres est élevé, et donc l'incertitude sur les paramètres et le modèle augmente [30].

*Avantages et limites :* En général, les méthodes de régression paramétriques sont faciles à interpréter et rapides à mettre en œuvre. Leur grande puissance est le fait qu'ils ne nécessitent pas de connaissances détaillées sur les phénomènes physiques et thermiques du bâtiment [59]. Cependant, ces modèles sont intéressants par leurs incapacités de prévoir les déviations non déterministes de la demande électrique causées par le changement brusque des conditions climatiques et les habitudes des occupants.

Un autre inconvénient des modèles de régression paramétriques est qu'ils sont fortement affectés par la forme retenue, ainsi, aux hypothèses contraignantes sur le choix des distributions des variables qui risquent de générer des incertitudes du modèle. Par exemple, il a été supposé que toutes les variables exogènes suivent une loi normale alors qu'en réalité une autre pourrait être appliquée, ce genre d'hypothèses peut produire des incertitudes sur le modèle lui-même.

Une première façon d'y arriver consiste à faire une régression polynomiale ou d'effectuer une transformation de la variable à prédire. Par exemple, appliquer des transformations logarithmiques naturelles, Box-Cox ou géométriques tel que proposé dans [87]. Il est à noter que l'emploi de la transformation logarithmique sur les observations de la variable d'intérêt est dans l'espoir de réduire sa variance et d'obtenir des résidus normalement distribués avec la nouvelle échelle [88]. Cependant, il est mentionné dans les travaux de [89] que l'utilisation de cette pratique n'est pas garantie tout le temps, ceci dépend de la nature de distribution des données.

#### 2.3.1.4 Modèles paramétriques non- linéaires

Une revue de littérature sur les modèles de régression non-linéaires est présentée dans [90], [91]. Ces modèles présentent une grande application dans la modélisation de la demande en électricité dans les bâtiments. Parmi ces modèles, les Réseaux de Neurones Artificielles (RNA) et leurs applications dans la prévision de la demande à court terme sont brièvement présentés. La revue réalisée dans [92] donne un aperçu sur les différentes applications des réseaux de neurones dans le domaine du bâtiment, tel que la modélisation du comportement thermique et la prévision de la température intérieure d'un bâtiment complexe [52], ainsi que les besoins en chauffage et en refroidissement dans un immeuble de bureaux. Ils ont aussi montré une capacité à capturer la relation non-linéaire dans la prévision de la demande d'électricité dans les bâtiments. Les RNA peuvent s'adapter en temps-réel pour prendre en compte de nouvelles mesures. Ils peuvent modéliser des systèmes complexes avec plusieurs entrées et plusieurs sorties.

La fonction générale d'une couche RNA est définie par la relation suivante [52] :

$$y = \varphi(\mathbf{w}\mathbf{x}) \quad (2-12)$$

Où  $\varphi()$  est la fonction d'activation,  $\mathbf{x}$  correspond au vecteur des entrées de la couche,  $\mathbf{w}$  est une matrice de dimensions  $n \times |\mathbf{x}|$ , où chaque ligne contient les poids associés aux  $n$  cellules assignées à la couche. Cette matrice configure les paramètres à déterminer lors de l'apprentissage ou l'entraînement [93]. Pour la phase d'entraînement, les poids peuvent être estimés hors ligne ou en ligne à chaque pas d'échantillonnage en fonction de l'application. Les poids sont normalement estimés de façon itérative ou récursive jusqu'à ce que la performance du réseau soit satisfaisante. Après l'apprentissage, une phase de validation avec des données non-utilisées pendant l'entraînement est employée afin d'évaluer la capacité de généralisation du réseau [52].

Cependant, deux problèmes majeurs sont rencontrés lors de l'utilisation des RNA. Les problèmes discutés dans la littérature parlent du choix de l'architecture optimale du réseau (nombre de couches et de neurones cachés et le type de fonction d'activation), et la durée de l'apprentissage, lorsque l'apprentissage est hors ligne [30]. Cette durée a une importance

primordiale, elle couvre, entre autres, la sélection de la bonne période de jeu de données afin d'éviter le sur-apprentissage. Par ailleurs, lors de sur-apprentissage, le RNA mémorise le bruit dans les mesures, et du coup, il réduit sa compétence de la généralisation. Une des méthodes les plus simples consiste à diviser les données d'apprentissage en deux, une partie des données étant réservée à l'identification des poids des neurones, et l'autre servant à vérifier à chaque itération, qu'il n'y a pas de dégradation de la prévision. Et si l'architecture conçue est dédiée à un bâtiment spécifique, il faut simplement reentraîner le réseau. Une alternative aux méthodes paramétriques est la régression non-paramétrique qui suppose moins d'hypothèses sur la nature de la relation entre les variables explicatives et la demande en électricité, ce qui lui donne plus de flexibilité.

### 2.3.2 Modèles non-paramétriques

Les modèles de régression non-paramétriques basés sur l'estimateur de noyau ou *Kernel Density Estimation en anglais* ont été introduits par Rosenblatt en 1956 [94] puis par Parzen en 1962 [95]. Ils ont été largement utilisés dans de nombreux domaines de sciences biomédicales, ensuite dans le domaine de l'énergie notamment dans la prévision de la vitesse du vent. Dans le contexte de la prévision de la demande d'électricité Rob Hyndman a réalisé divers travaux basés sur la régression non-paramétrique et l'estimation de la densité conditionnelle proposés dans [96]–[100]. La plupart de ces études utilisent l'estimateur par noyau pour expliquer la relation entre la demande d'électricité et les variables exogènes en estimant la densité conditionnelle  $f(y|x)$ . Cependant, les modèles de Hyndman n'étaient pas adaptatifs aux changements de la demande électrique, ce qui présente une limite au niveau du consommateur, si ce dernier veut suivre en ligne sa consommation. Pour le cas hors ligne, les estimateurs par noyau sont fondés sur un grand ensemble de données ou des mesures passées de la demande résidentielle [30], [48]. Le processus de modélisation commence avec l'analyse des données disponibles pour former le modèle et réaliser l'apprentissage. Une fois cette étape terminée, il réalise l'étape de la prévision de la demande. Alors que pour le cas en ligne, le modèle s'entraîne automatiquement au fur et à mesure que les nouvelles observations arrivent.

Les modèles non-paramétriques s'écrivent d'une manière similaire à la régression linéaire, mais avec une fonction inconnue à spécifier donné par 2-13 :

$$y_k = \mathbb{E}(y|x_k) + \varepsilon_k \quad (2-13)$$

Où,  $\mathbb{E}(y|x_k)$  est l'espérance conditionnelle de  $y$  sachant  $x_k$ , qu'il faut estimer à partir des données [101]. L'idée principale de la régression non-paramétrique consiste à estimer directement l'espérance  $\mathbb{E}(y|x_k)$  à partir d'une distribution conditionnelle  $f(y|x_k)$ . Ceci par un "lissage" qui va produire des moyennes pondérées de l'observation de la puissance électrique  $y_k$ . Pour le cas discret, si la loi  $p(y|x_k) = f(y|x_k)$  existe, l'espérance conditionnelle sera calculée par 2-14.

$$\mathbb{E}(y|x_k) = \sum_y y f(y|x_k) \quad (2-14)$$

La distribution  $f(y|x_k)$  pourrait être analysée à partir d'un estimateur de la densité par noyau. Contrairement aux modèles de régression paramétrique, l'estimateur par noyau considère le problème de détermination de la moyenne conditionnelle sans avoir une connaissance approfondie sur le type de liaison entre la demande et les variables exogènes. En effet, cet estimateur ne concerne plus les paramètres, mais directement la fonction de régression d'où le terme non-paramétrique, ce qui lui donne beaucoup plus de flexibilité [30].

À titre d'exemple de l'estimateur par noyau, l'estimateur de Nadaraya-Watson [102] est une des fonctions de pondération très utilisée pour établir la relation entre les variables. Sa probabilité conditionnelle  $p(y|x) = f(y|x)$  est estimée par la relation suivante (2-15) :

$$f(y|x) = \frac{1}{|\mathbf{y}_x|^\nu} \sum_{y_i \in \mathbf{y}_x} \mathcal{K} \left( \frac{y - y_i}{\nu} \right) \quad (2-15)$$

Où  $\mathbf{y}_x = \{y_k | x_k = x\}_{k=1 \dots N}$  correspond à l'ensemble de mesures de  $y_k$  quand  $x_k = x$ . Dans cette expression,  $\nu > 0$  est la fenêtre de lissage ou *bandwidth* du noyau et  $\mathcal{K}$  est la fonction noyau vérifiant les deux conditions suivantes :  $\sum \mathcal{K}(y) = 1$  et, la densité de probabilité doit être symétrique (*i.e.*  $\mathcal{K}(-y) = \mathcal{K}(y)$ ) [103].

La fenêtre de lissage  $\nu$ , a une influence assez évidente sur la qualité de l'estimation. Plus la variance ou la fenêtre de lissage  $\nu$  est petite, plus la répartition va être très resserrée autour de la

TABLEAU 2-1 Quelques exemples de noyaux classiques [105]

Noyau	$\mathcal{K}$
Parabolique (noyau d'Epanechnikov)	$\begin{cases} \frac{3}{4}(1-y^2), \text{ si }  y  < 1 \\ 0, \text{ sinon} \end{cases}$
Bicarré	$\begin{cases} \frac{3}{4}(1-y^2), \text{ si }  y  < 1 \\ 0, \text{ sinon} \end{cases}$
Triangle	$1 -  y , \text{ si }  y  < 1, 0 \text{ sinon}$
Gaussien	$\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$
Rectangle	$\frac{1}{2} \text{ si }  y  < 1, 0 \text{ sinon}$

moyenne de l'estimateur du noyau et ce dernier va donner plus de détails sur les observations. À l'inverse, plus  $\nu$  est grande, plus la répartition de la densité sera aplatie et étirée en largeur et l'estimateur devient moins lisse [104], comme illustré à la Figure. 2-6. La calibration de la fenêtre de lissage est donc très importante pour s'approcher de la vraie distribution des données observées.

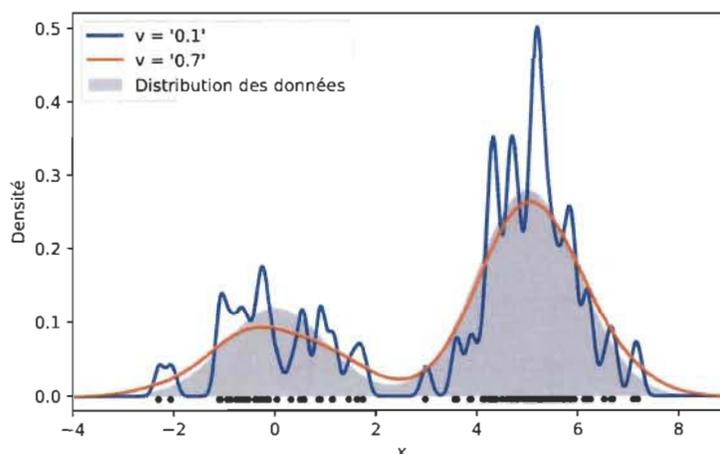


FIGURE 2-6 Comparaison de la distribution des densités pour différentes fenêtres de lissage [87].

Il existe plusieurs types de fonction à noyau pouvant satisfaire les conditions définies précédemment : les fonctions Uniforme, Triangle, Epanechnikov, Quadratique, Cubique, Gaussienne, Circulaire.etc [105]. Le Tableau 2-1 résume les propriétés typiques des noyaux les plus utilisés. Les noyaux les plus couramment employés pour la modélisation sont les noyaux gaussiens.

Selon [106], le type de noyau choisi n'est pas primordial, par rapport au choix de la fenêtre de lissage. Pour l'étude présente, la fonction à noyau gaussienne sera retenue, elle est définie par la relation suivante :

$$\mathcal{K}(y - y_k) = \frac{1}{v\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y - y_k}{v}\right)^2\right) \quad (2-16)$$

### 2.3.2.1 Cas multivarié

En outre, l'estimateur par noyau est bien introduit dans l'analyse statistique. Il permet d'estimer le lien entre les variables via les distributions jointes, conditionnelles et marginales, où la relation entre ces distributions est définie par 2-17 [103]. Ces possibilités favorisent des analyses plus élaborées ciblant le cas multivarié.

$$f(y|x) = \frac{f(y, x)}{f(x)} \quad (2-17)$$

Pour le cas de la densité marginale et jointe, elles sont estimées en utilisant une fonction mono-variable et bi-variables, respectivement. Toutefois, en fonction de la dimension de la variable d'entrée  $x$  l'estimation non-paramétrique recourt à des hypothèses permettant la factorisation. Pour le cas de la distribution de la variable  $x$ , l'estimateur est défini en assumant une indépendance entre les variables d'entrée.

$$f(x) = \frac{1}{N} \sum_{k=1}^N \prod_{i=1}^d \frac{1}{v_x^i} \mathcal{K}(x^i - x_k^i) \quad (2-18)$$

D'autre part, l'estimateur pour la distribution jointe est construit à partir de multiplications des noyaux individuels. À cet égard, la distribution jointe non-paramétrique adopte la forme suivante :

$$f(y, x) = \frac{1}{N} \sum_{k=1}^N \frac{1}{v_y} \mathcal{K}(y - y_k) \prod_{i=1}^d \frac{1}{v_x^i} \mathcal{K}(x^i - x_k^i) \quad (2-19)$$

Les estimateurs antérieurs permettent alors une expression pour la distribution conditionnelle de l'équation (2-17) :

$$f(y|x) = \frac{\sum_{k=1}^N \frac{1}{v_y} \mathcal{K}(y - y_k) \prod_{i=1}^d \frac{1}{v_x^i} \mathcal{K}(x^i - x_k^i)}{\sum_{k=1}^N \prod_{i=1}^d \frac{1}{v_x^i} \mathcal{K}(x^i - x_k^i)} \quad (2-20)$$

Il est à noter que pour le cas multidimensionnel pour lequel plusieurs valeurs du passé de la variable explicative sont utilisées. Le calcul de l'espérance conditionnelle  $\mathbb{E}(y|x)$  est très difficile pour une évaluation pour tout  $x$ , et ceci pose le problème de "malédiction de la dimension" [107]. Par contre, il y a la possibilité d'introduire une hypothèse d'indépendance entre les variables d'entrée pour évaluer l'espérance de la loi de probabilité.

## 2.4 Synthèse du chapitre

Cet état de l'art a permis de présenter de manière simplifiée des modèles prévisionnels élaborés pour être en mesure d'anticiper l'évolution de la demande d'électricité dans les bâtiments résidentiels. Ces modèles sont classifiés en trois grandes familles à savoir : les modèles boîte blanche, boîte noire et boîte grise. Comme nous n'avons aucune information sur la physique du bâtiment, nous avons mis l'accent sur les approches statistiques pour l'estimation et l'explication de la demande énergétique à court terme. Nous avons présenté les concepts de base de modélisation et de prévision de la demande associés aux régressions paramétriques et non-paramétriques. Sur la base de ces méthodes, nous nous sommes inspirés pour proposer une approche de pouvoir explicatif et prédictif cohérente et transposable à la problématique de la thèse. Il s'agit de construire un cadre de modélisation qui permet l'anticipation et l'explication de la puissance d'électricité totale du côté des clients. Ce modèle doit prendre moins d'information sur le bâtiment, mais exploite d'une façon optimale les données mesurées de température extérieure afin de nous permettre d'assimiler l'information contenue dans la consommation thermosensible des bâtiments. Une réponse à ce problème est exposée dans les chapitres 3 et 4. Elle consiste à exploiter les estimateurs par noyau uni et bi-variables pour estimer la densité de probabilité de la demande. Une caractéristique importante de ces estimateurs est liée à leurs capacités intrinsèques à s'adapter aux variations

de la puissance, et ceci pour mettre constamment à jour le modèle avec de nouvelles données au lieu de commencer avec une grande quantité de données.

## **Chapitre 3 - Modélisation et prévision de la composante thermosensible**

### **3.1 Introduction**

Le présent chapitre a pour objectif d'identifier la composante thermosensible et d'étudier le comportement thermique et énergétique des bâtiments résidentiels existants, à savoir : des maisons unifamiliales typiques québécoises. La section 3.1 expose la base de données utilisée dans ce projet. Nous décrivons aussi le flux d'énergie dans un bâtiment à partir d'un diagramme de Sankey [108]. La section 3.2 présente l'analyse de la corrélation entre la demande d'électricité et la température qui permet d'extraire la composante thermosensible. Ensuite, à partir de la composante capturée, nous calculons la valeur du coefficient de déperdition de chacune des maisons unifamiliales étudiées. De plus, un test d'orthogonalité a été établi entre les observations de la température et l'erreur produite entre le modèle et les mesures de la puissance, c'est-à-dire, en rendant l'erreur et les températures décorréliées, pour minimiser l'effet des facteurs indépendants de la température sur la consommation thermosensible. La section 3.3, fait l'objet d'une comparaison entre l'approche proposée avec des modèles adaptatifs des réseaux de neurones et un modèle linéaire pour prévoir la forme de la thermosensibilité de la puissance d'électricité. Finalement, il est question de valider la prévision de la composante thermosensible produite par le modèle proposé, à partir d'une analyse de sensibilité qui permet de connaître l'effet de l'incertitude des valeurs futures de la température sur la sortie du modèle proposé. Nous utilisons dans ce cas, des écarts-types de température prédite sur les prochaines 24 heures fournies par cinq stations météo. Ceci permet alors de quantifier la composante thermosensible en analysant la précision de la prévision de cette composante selon la précision de la prévision des températures des cinq stations météo.

### 3.2 Description de la base de données

Les données réelles de la température extérieure et de la puissance totale ont été fournies par le Laboratoire des Technologies de l'Énergie (LTE). Ces données ont été enregistrées sur deux ans et 9 mois du 01 janvier 1994 au 30 septembre 1996. Les mesures des données réelles de la puissance globale sont réalisées sur un total de 5 bâtiments résidentiels existants de la région de Montréal. Il faut savoir ici que toutes les cinq maisons ont été testées et validées. Toutefois, nous avons extrait les résultats d'une seule maison qui est la résidence numéro 5 comme un exemple pour montrer les résultats dans les chapitres 3 et 4. La Figure. 3-1 illustre un exemple des profils de puissance totale et de température extérieure pour des enregistrements effectués sur une période d'un an en 1995 pour une maison unifamiliale numéro 5. Nous trouvons à l'Annexe B l'ensemble des données présentées sous forme de graphique pour chaque bâtiment. Il est à noter qu'aucune information sur l'architecture, les dimensions et les caractéristiques thermiques des résidences était disponible.

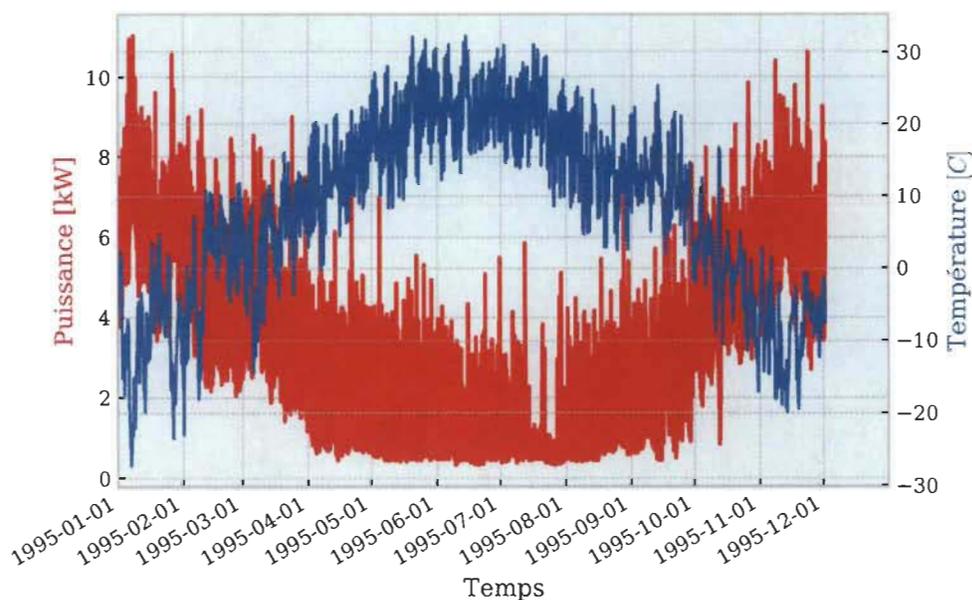


Figure. 3-1 Profils de la puissance totale et de la température extérieure pour une résidence unifamiliale.

Pour chacune des maisons, un mesurage a été réalisé à des intervalles de temps d'une heure. L'ensemble des données mesurées dans chaque maison comprend :

- la température extérieure ;
- la puissance totale de la résidence ;
- la puissance de chauffage électrique ;
- la puissance du chauffe-eau électrique ;
- la puissance totale de toutes les autres charges. Il est à noter que l'énergie consommée par certaines de ces charges finit par être transmis sous forme de chaleur qui contribue au chauffage de la maison pendant l'hiver, telles que l'éclairage, le réfrigérateur et le four électrique. Contrairement à certaines charges, comme la sècheuse où son énergie dépensée sort à l'extérieur sous forme de chaleur et d'humidité.

De la relation 3-1, nous pouvons déduire facilement que la puissance totale du bâtiment définit par  $Y$ , est la somme de la puissance électrique du système de chauffage  $Y_{ehs}$ , la puissance électrique des autres charges  $Y_{el}$  et la puissance électrique du chauffe-eau  $Y_{ewh}$ .

$$Y = Y_{ehs} + Y_{ewh} + Y_{el} \quad (3-1)$$

Peu importe le type et la forme des résidences québécoises, une consommation sensible aux variations de température est observable, mais ceci varie d'un client à l'autre. En effet, avant d'établir la méthodologie d'identification de cette consommation thermosensible, le bilan d'énergie a été pris en compte à partir de la puissance totale  $Y$ . Ce bilan est nécessaire d'une part pour définir les scénarios plausibles servant à la simplification des postulants qui collaborent dans la formulation du modèle boîte noire, et aussi, pour sélectionner la variable nécessaire permettant de prédire la demande d'électricité avec un modèle de pouvoir explicatif.

Nous présentons d'abord le bilan d'énergie sous forme d'un diagramme de Sankey d'un bâtiment, illustré dans la Figure. 3-2. Ce diagramme montre les différentes connexions des flux énergétiques en vue d'esquisser la demande d'électricité de divers facteurs. Le résumé des gains et des pertes de chaleur est défini par un bilan d'énergie incluant un ensemble de composantes déterministes et stochastiques, tel que présenté à l'équation (3-2).

$$C\dot{X}^{int} = \underbrace{UA(X^{ext} - X^{int}) - \dot{q}_p}_{\dot{Q}_{pertes}} + \underbrace{Y_{ehs} + Y_{el} + \dot{q}_s + \dot{q}_{occ}}_{\dot{Q}_{gains}} \quad (3-2)$$



Figure. 3-2 Les flux d'énergie dans le bâtiment.

Où  $C$  est la capacité thermique du bâtiment,  $U$  est la constante de déperdition thermique équivalente,  $A$  est la surface d'échange de chaleur,  $X^{ext}$  et  $X^{int}$  représentent la température extérieure et intérieure respectivement,  $\dot{q}_s$  et  $\dot{q}_{occ}$  sont respectivement les gains de chaleur externes et internes liés principalement à l'ensoleillement et aux occupants. Le terme  $\dot{Q}_{gains}$  est la somme des contributions individuelles des gains internes et externes générés par le soleil, les occupants et ceux dus à la puissance électrique de système de chauffage  $Y_{ehs}$  et de certains usages domestiques des charges  $Y_{el}$ , qui contribuent directement dans le contrôle de la température intérieure. La majorité des pertes de chaleur  $\dot{Q}_{pertes}$  proviennent souvent de  $\dot{q}_X$  définies par (3-3) qui représente les déperditions thermiques globales du bâtiment à travers les murs, de  $\dot{q}_v$  qui exprime le transfert de chaleur à l'extérieur du bâtiment tel que la sécheuse, de  $\dot{q}_{inf}$  qui définit les échanges de chaleur générés par la ventilation ou l'infiltration de l'air extérieur et enfin  $\dot{q}_{ewh}$  qui représente les pertes de chaleur de l'eau chaude, telles qu'exprimées par l'équation (3-4).

$$\dot{q}_X = UA(X^{ext} - X^{int}) \quad (3-3)$$

$$\dot{q}_p = \dot{q}_v + \dot{q}_{inf} + \dot{q}_{ewh} \quad (3-4)$$

Pour que le modèle de prévision de la puissance soit adéquat avec l'objectif principal de la thèse, nous n'allons considérer seulement la contribution de la température extérieure et l'intérieure. Concernant la température extérieure, elle est utilisée car elle représente le facteur prépondérant comparé aux facteurs du vent et de l'ensoleillement, surtout si on s'intéresse à

des résidences situées dans la région de Québec.

Toutefois la mesure de la température intérieure  $X^{int}$  n'est pas disponible dans les données réelles fournies par LTE. Pour cette raison, une hypothèse de départ a été émise, qui suppose une température intérieure constante à  $21^{\circ}C$  sur toute la période d'analyse, comme étant la température intérieure moyenne des maisons québécoises [109]. Cette valeur est utilisée pour calculer l'écart de température  $x_k$  entre l'extérieur et l'intérieur à chaque instant  $k$  donné par (3-5) :

$$x_k = x_k^{ext} - x_k^{int} \quad (3-5)$$

Pour cela, il a été question de réduire l'équation du bilan d'énergie en incluant la température extérieure et intérieure comme les seules variables exogènes données par 3-6. Ceci permet de cerner le problème de la modélisation en sélectionnant un modèle de régression non-paramétrique permettant d'extraire la composante thermosensible soumise aux aléas de la température.

$$Y = h(X = x_k) + r \quad (3-6)$$

L'objectif est de trouver la fonction  $h(X = x_k)$  qui explique au mieux la puissance  $Y$ . Pour cela, nous supposons  $y_k$  à temps discret  $k$  la réalisation de la puissance totale  $Y$ , vue comme une variable stochastique. Pour cela, la valeur à estimer  $y_k$  est calculée par  $x_k$  qui est la réalisation de la variable  $X$  mesurée sur un intervalle horaire pendant une période finie  $N$ , présentée par (3-7) :

$$y_k = h(x_k) + r_k, \quad k \in \{1, \dots, N\} \quad (3-7)$$

Ici  $h(x_k)$  représente la composante thermosensible de  $y_k$ . Elle définit la puissance consommée par le système de chauffage électrique  $y_{ehs}$ , la climatisation et d'autres charges affectées directement par le changement de l'écart de température  $x_k$ .  $r_k$  est la composante résiduelle qui est indépendante des observations des écarts de température. La composante  $r_k$  a deux propriétés qui vérifient : i)  $\mathbb{E}(r_k|x_k) = 0$ , et ii)  $\mathbb{E}(h(x_k)r_k) = 0$ . Autrement dit, ces propriétés expriment l'orthogonalité de  $r_k$  par rapport à  $x_k$ .

Dans ce cas, à partir de l'équation de régression (3-7) nous pouvons estimer  $h(x_k)$  qui est

la valeur attendue ou également l'espérance conditionnelle de  $y_k$  conditionnée par  $x_k$  définie par (3-8). Sa fonction de densité peut alors être écrite sous la forme  $f_k(y|x)$  [110]. Elle permet de connaître la distribution de probabilité de la puissance  $y_k$  quand l'écart de température  $x_k$  est connu.

$$h(x_k) = \mathbb{E}(y_k|x_k) \quad (3-8)$$

Notre motivation première est d'estimer la fonction de densité  $f$  afin d'avoir une explication réaliste de la nature de la relation entre la demande d'électricité et l'écart de température. Cette information servira plus tard à ajuster le modèle de prévision de la puissance totale. Cependant, nous ne possédons pas une connaissance a priori sur le type de la relation entre la puissance consommée de la maison étudiée et sa variable explicative.

Pour contourner cette contrainte, nous avons employé un modèle de régression non-paramétrique fondé sur l'Estimation Adaptative de la Densité Conditionnelle EADC de la fonction  $f$ . Ce modèle permet d'estimer la densité en tout point par l'estimateur par noyau ou Kernel Density Estimation (KDE) qui détermine la moyenne et la variance.

Prenant en compte la discussion ci-dessus et les contributions mentionnées au chapitre 1 sur la part thermosensible, nous présentons brièvement dans la Figure. 3-3 la structure générale de la stratégie d'identification de la consommation thermosensible, en tenant compte d'une seule mesure de la température pour expliquer la puissance totale. Une fois la composante  $h(x_k)$  sera identifiée, une prévision de cette puissance devrait s'effectuer sur un horizon  $\tau$  des vingt-quatre heures suivantes.

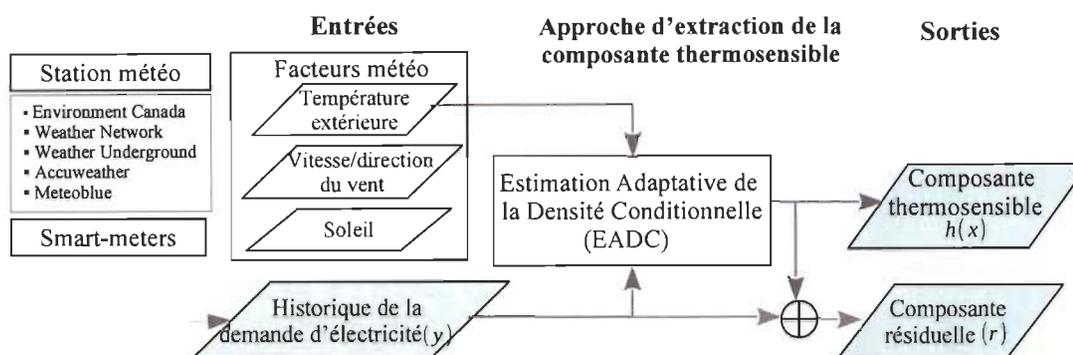


Figure. 3-3 Structure générale de la décomposition de la puissance d'électricité totale avec l'approche proposée.

### 3.3 Analyse de la relation entre la demande d'électricité et la température

Dans cette section, nous présentons l'approche explicative et prédictive EADC qui sert à extraire la part thermosensible  $h(x_k)$ , et étudier la nature de la relation entre la puissance électrique et l'écart de température. Cette approche a été conçue dans ce travail pour capturer de manière adaptative les valeurs de  $h(x_k)$  en termes de densité conditionnelle  $f(y|x)$ . Un aspect fondamental à comprendre est qu'ici, il y a deux quantités à estimer. La fonction de densité jointe du couple de variables de puissance/écart de température  $f(x, y)$  et la densité marginale de l'écart de température  $f(x)$ . En estimant ces deux quantités, nous pouvons immédiatement déduire l'expression de la fonction de densité conditionnelle  $f(y|x)$  définie en (3-9) :

$$f(y|x) = \frac{f(x, y)}{f(x)} = \frac{f(x, y)}{\sum_y f(x, y)} \quad (3-9)$$

En résumé, la méthodologie à suivre est la suivante :

- estimer la fonction de la densité jointe à partir des observations de l'écart de température et de puissance ;
- estimer la fonction marginale de l'écart de température à partir de la fonction jointe ;
- déduire la fonction de la densité conditionnelle et identification à partir de cette dernière de la composante  $h(x_k)$ .

Nous commençons d'abord par l'estimation de la fonction de densité jointe  $f_k(x, y)$  en tout

point de l'espace d'observation discret. Sur cet espace, nous construisons une grille de points de température et de puissance  $\mathcal{D}_X \times \mathcal{D}_Y$  sur laquelle la densité  $f$  va être estimée. Cette grille a été créée à partir d'un calcul du minimum et du maximum des points d'observations sur des intervalles, définis par  $\mathcal{D}_X = \{x_1, \dots, x_{N_x}\}$  et  $\mathcal{D}_Y = \{y_1, \dots, y_{N_y}\}$ , qui représentent l'ensemble discret de l'écart de température et la puissance totale, respectivement. De plus, dans la construction des ensembles  $\mathcal{D}_X$  et  $\mathcal{D}_Y$ , les propriétés suivantes doivent être vérifiées :

- la discrétisation des variables de température et de puissance est réalisée par la différence entre deux points consécutifs qui donne une constante qui définit le pas de la discrétisation :  $x_i - x_{i-1} = \Delta_x$ , pour tout  $i \in \{2, \dots, N_x\}$  et  $y_j - y_{j-1} = \Delta_y$ , pour tout  $j \in \{2, \dots, N_y\}$  ;
- le premier élément de l'ensemble correspond à la valeur minimale et le dernier correspond à la valeur maximale pour les variables de température et de puissance, c'est-à-dire :  $x_1 = x_{min}$ ,  $x_{N_x} = x_{max}$  et  $y_1 = y_{min}$ ,  $y_{N_y} = y_{max}$ .

### 3.3.1 Méthodologie d'apprentissage adaptatif

Étant donné la nature non-paramétrique de la distribution  $f(y|x)$ . Celle-ci peut être vue comme un ensemble d'histogrammes (fréquences relatives) sur l'espace discret de la puissance. Sur ce point, cette distribution peut s'associer à une loi discrète où la probabilité d'un élément de  $\mathcal{D}_Y$  conditionné à une observation  $x_k$  est :  $p(y = y_i|x_k) = f(y_i|x_k)$ . Dans le cadre statistique fréquentiste, cette probabilité est définie par le rapport entre le nombre d'observations  $y_k = y_i$  sur le nombre très grand (vers l'infini) d'observations. Cependant, dans la pratique les systèmes d'apprentissage peuvent disposer de très peu d'observations, c'est le cas de la phase initiale d'apprentissage. C'est pourquoi nous proposons une méthode d'estimation de probabilités non-paramétriques encadrée dans les méthodes d'inférence bayésienne. En effet, il est possible d'introduire une initialisation pour  $f(y|x)$  et au fur et à mesure que les nouvelles données deviennent disponibles, les valeurs de ces probabilités pourront être ajoutées. Une façon simple consiste à démarrer la distribution  $f(y|x)$  à partir des informations générales sur le comportement de la consommation à partir de la température, postérieurement, nous pouvons mettre à jour la fréquence relative à partir des observations  $y_k$ . Ceci est représenté par le

processus suivant :

$$f_k(y_i|x_k) = f_{k-1}(y_i|x_k) + \frac{1[y_i = y_k]}{N_y} \quad \forall y_i \in \mathcal{D}_Y \quad (3-10)$$

Où  $1[y_i = y_k]$  est une fonction indicatrice retournant 1 si la condition est vrai et 0 dans le cas contraire.

La représentation de  $f_k(y_i|x_k)$  en tant que fréquence relative n'est pas cohérente avec le processus de l'équation (3-10) lequel produit une distribution non-normalisée. Pour résoudre cela, nous pouvons tout simplement faire une re-normalisation de  $f_k(y_i|x_k)$ . Le processus pour la mise à jour de la distribution conditionnelle adopte la forme suivante selon [111] :

$$f_k(y_i|x_k) = \frac{N_y}{N_y + 1[y_i = y_k]} f_{k-1}(y_i|x_k) + \frac{1[y_i = y_k]}{N_y + 1[y_i = y_k]} \quad \forall y_i \in \mathcal{D}_Y \quad (3-11)$$

Nous pouvons constater que si  $1[y_i = y_k] = 0$  la valeur de  $f_k(y_i|x_k)$  est égale à son passé  $f_{k-1}(y_i|x_k)$ .

L'équation (3-11) représente un histogramme adaptatif, dont la valeur du segment de barre associée à  $y_i = y_k$  est mise à jour suivie dans une normalisation pour respecter la propriété  $\sum_y f_k(y|x_k) = 1$ . Il est donc évident que cette estimation non-paramétrique a un noyau associé défini par la fonction  $1[y_i = y_k]$ . Entant donnée que ce noyau a une fenêtre de lissage égale à zéro les probabilités des éléments voisins de  $y_i = y_k$  ne seront pas affectées par l'estimation. On peut alors remplacer ce noyau simple par un noyau permettant une distribution plus souple en récupérant le contrôle sur le paramètre de lissage. En réécrivant l'équation (3-10) selon un noyau arbitraire et en normalisant tel que fait pour avoir l'équation (3-11) nous obtenons :

$$f_k(y|x_k) = \frac{N_y}{N_y + 1} f_{k-1}(y|x_k) + \frac{1}{N_y + 1} \mathcal{K}(y - y_k) \quad (3-12)$$

Avec  $\sum_y \mathcal{K}(y - y_k) = 1$ .

Le processus de l'équation (3-12) configure l'apprentissage adaptatif d'une loi conditionnelle non-paramétrique. Il faut souligner que ceci met à jour la loi  $f_k(y|x_j)$  associée à l'observation  $x_k = x_j$ . L'estimation ne modifie pas le reste de distributions conditionnelles,

car elle n'est qu'une fonction tout au long de la puissance  $y$ . Cette situation engendre un problème de lissage au long de  $x$  dû à l'impossibilité d'assurer suffisamment d'observations dans tout le domaine de la température. En effet, certaines régions ou points particuliers de  $\mathcal{D}_X$  peuvent prendre beaucoup du temps avant d'être observée. Cela peut entraîner des discontinuités considérables dans l'espérance conditionnelle, laquelle détermine finalement le modèle régression.

Afin d'assurer une estimation plus douce à travers le domaine de la température, on propose de remplacer le noyau  $\mathcal{K}(y - y_k)$  pour une estimation non-paramétrique  $\hat{f}_k(y|x)$  à partir d'une estimation jointe. Pour cela, un estimateur de noyau bidimensionnel  $\hat{f}_k(y, x)$  est utilisé sur une mesure ponctuelle  $(x_k, y_k)$ . En effet, l'estimation ponctuelle de la densité conditionnelle est définie par :

$$\hat{f}_k(y|x) = \frac{\hat{f}_k(y, x)}{\sum_y \hat{f}_k(y, x)} \quad (3-13)$$

Pour ce faire, nous estimons la fonction de densité  $f_k(x, y)$  sur l'ensemble de points de la grille à l'aide d'un noyau gaussien bi-variable. Ce noyau doit être centré sur  $(y_k, x_k)$  qui représente le centre de la distribution [112], donnée par (3-14) :

$$f_k(y, x|\Sigma) = \mathcal{N} \left( \begin{bmatrix} x_k \\ y_k \end{bmatrix}, \Sigma \right) \quad (3-14)$$

Où  $\Sigma$  correspond à la matrice de covariance définie par (3-15), dont laquelle nous trouvons sur la diagonale les variances de l'écart de température et de la puissance, et sur l'extra-diagonale nous avons les covariances entre les deux variables. Ceci est calculé pour définir la fenêtre de lissage du noyau de l'estimateur, cette valeur est celle qui va faire de la répartition de la densité plus au moins pointue ou tassée [113]. Pour l'exemple de la maison étudiée, la matrice de covariance employée est une matrice symétrique définie par (3-15), dont les valeurs de la fenêtre de lissage ne sont pas diagonales. Donc pour pouvoir moduler le rayon d'influence de ce noyau en conservant la corrélation de variables, nous avons associé un facteur  $\lambda > 0$  à la matrice de covariance définie par  $\Sigma = \lambda \Sigma_{xy}$  [114].

$$\Sigma = \lambda \Sigma_{xy} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy < 0} \\ \sigma_{yx < 0} & \sigma_y^2 \end{pmatrix} \quad (3-15)$$

Ce facteur  $\lambda$  a été choisi de manière empirique en balayant l'intervalle de  $[0.01, 1]$ . Dans ce cas d'étude ce facteur a été fixé à  $\lambda = 0.07$ , ce qui donne une bonne résolution de contour de la densité estimée. Il est à noter que, la fenêtre de lissage sélectionnée pour la consommation de la maison étudiée n'est pas valable pour toutes les autres maisons. Elle changera selon la période d'analyse, la résolution des données et les mesures de puissance pour chaque type de bâtiment.

Sommairement, le diagramme de la Figure. 3-4 montre la structure primaire du modèle EADC employé pour la décomposition de la puissance totale en fonction de l'écart de température. Étant donné que l'approche proposée est adaptative en raison du changement de

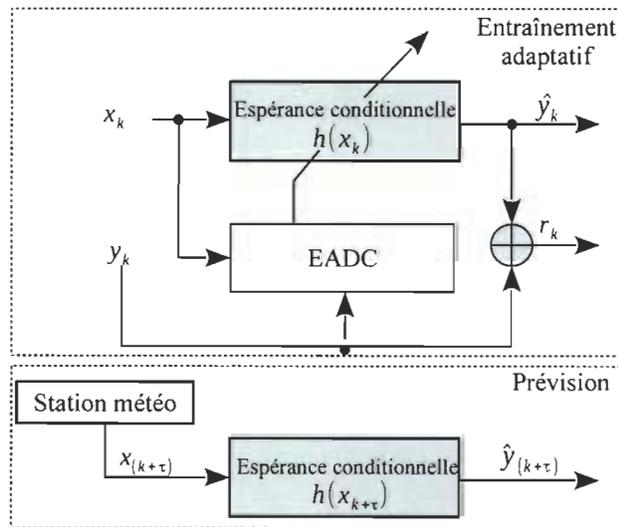


Figure. 3-4 Schéma du modèle EADC proposé pour identifier la thermosensibilité de la consommation.

la demande électrique, deux phases ont été considérées dans le processus de prévision de  $h(x_k)$ , premièrement, la phase d'entraînement en ligne qui utilise l'information du passé et du présent de  $x_k$  et  $y_k$  pour mettre à jour la densité jointe de l'équation (3-12) et estimer la densité conditionnelle de la puissance totale, deuxièmement, la phase de validation ou de prévision qui exploite le futur de l'écart de température  $x_{k+\tau}$  pour prédire la thermosensibilité sur un horizon

$\tau$ . Il est à noter que les observations utilisées dans l'entraînement du modèle n'ont jamais été employées dans la prévision de la demande d'électricité dans la phase de la validation.

### 3.3.2 *Caractéristiques de la composante thermosensible*

Deux scénarios ont été considérés pour estimer la fonction de densité conditionnelle  $f_k(y|x)$  et évaluer la composante  $h(x_k)$  de la puissance totale. Le premier scénario utilise les observations originales de la demande d'électricité, tandis que, le deuxième considère une transformation logarithmique à base 10 sur les mêmes données de la puissance totale. Ce scénario est souvent utile dans l'analyse des variables aléatoires ayant une distribution asymétrique [115], [116]. Dans le cas de la demande électrique, cette transformation pourra être établie, vu sa nature aléatoire perturbée par des facteurs non-déterministes. Une hypothèse a été posée sur l'application de la transformation logarithmique qui peut stabiliser la variance de la demande d'électricité, et cela lorsque ses données ne sont pas normalement distribuées [89]. Notre objectif dans ce scénario est de prédire la puissance électrique à l'échelle des watts. Il s'agit donc de convertir la composante  $h(x_k)$  estimée de l'échelle logarithmique à l'échelle originale de la puissance [117]. L'approche EADC a été appliquée pour les deux scénarios en estimant de manière adaptative les fonctions de densité jointe et marginale. À ce stade, nous faisons appel aux observations originales de la demande d'électricité  $y_1, \dots, y_N$  prise à l'instant  $k$ , et les observations transformées  $\tilde{y}_1, \dots, \tilde{y}_N$ , qui sont le résultat de la transformation suivante :  $\tilde{y}_k = \log(y_k + 1)$  [117]. En effet, pour que la transformation logarithmique soit appropriée, les données observées doivent être strictement positives [87]. C'est pour cette raison que la valeur de 1 a été ajoutée pour éviter les valeurs négatives en échelle logarithmique. Exemple, si nous appliquons la transformation logarithmique sur une valeur de puissance  $y_k = 0.1$  kW, cela signifie que le logarithme à base de 10 sera  $\tilde{y}_k = -1$ .

Par conséquent, l'analyse appliquée génère un map de densité pour la résidence unifamiliale étudiée, ce qui est présenté dans les Figure. 3-5 et Figure. 3-6. Elles illustrent la relation entre la demande d'électricité et l'écart de température analysée avec le modèle EADC sans et avec la transformation logarithmique. Chacune de ces figures communique la distribution de la densité de la demande d'électricité conditionnée par l'écart de température pendant une année d'analyse.

La Figure. 3-5. (A) et (B) représente la densité jointe et conditionnelle estimée des couples  $f_k(x, y)$  et  $f_k(y|x)$ , respectivement. Les régions en brun et rouge qui apparaissent sur le map (A) de la figure signifient qu'il est plus probable d'observer une distribution de densité élevée de la demande d'électricité et ceci autour de  $-3^\circ\text{C}$  (ce qui correspond à une température extérieure  $x_k^{ext} = 18^\circ\text{C}$ ). Cette densité élevée reflète la quantité annuelle de besoins en puissance qui a plus de chance de se répéter conjointement à la fréquence d'apparition de l'écart de température. Par exemple la caractéristique thermique dominante de la résidence étudiée présentée par la densité jointe la plus élevée liée au couple  $(-3^\circ\text{C}, 1.09\text{kW})$ .

D'un autre côté, sur le map (B), nous observons une relation inverse entre la demande d'électricité et la température, qui est non-linéaire. Dans cette situation, la puissance a tendance à augmenter de façon plus remarquable lorsque la température diminue, comme est montrée par la partie à gauche sous forme d'une pente. Ainsi dans cette période une grande portion de la puissance totale a été identifiée, et s'explique principalement par l'usage important du chauffage électrique. Tandis que pour les températures plus élevées, il est observé une partie plus constante de la puissance électrique. Il est constaté aussi que la résidence étudiée ne fait pas appel à la charge de climatisation. C'est la raison pour laquelle la corrélation entre la demande et la température ne rapporte pas une relation en forme de vallée sur toute la plage de variation de température.

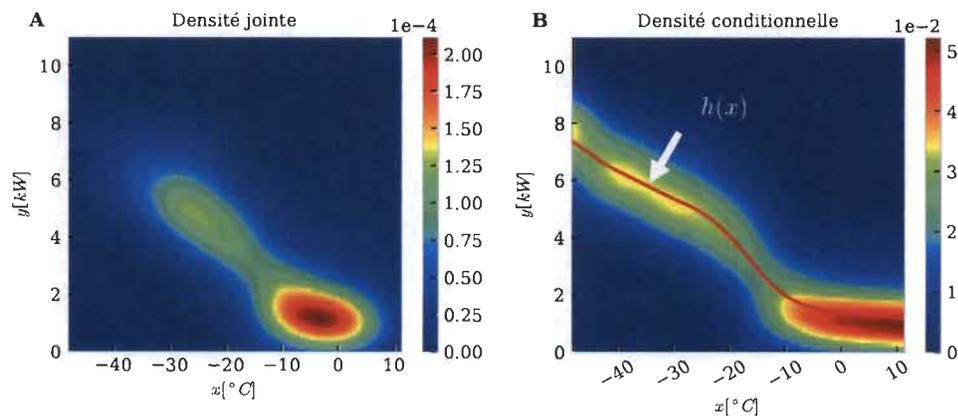


Figure. 3-5 La densité jointe et conditionnelle de la demande d'électricité estimée : (A) jointe, (B) conditionnelle.

Pour le deuxième scénario, la Figure. 3-6. (A) montre la densité jointe  $f_k(x, \bar{y}_k)$  estimée de

la demande d'électricité et l'écart de température sur une année. La densité conditionnelle estimée est observée dans le map (B). Elle visualise deux régions de la relation entre la demande et la température, connectées entre elles par un point noté  $x_b$ . Il est à noter que l'application de la transformation logarithmique sur la puissance peut amener des hypothèses sur l'interprétation thermique ou physique du point  $x_b$ . Nous pouvons affirmer que les deux régions sont séparées par une température d'équilibre du bâtiment. Cette dernière est la température extérieure à partir de laquelle les apports internes sont égaux aux apports externes [118]. Pour la maison

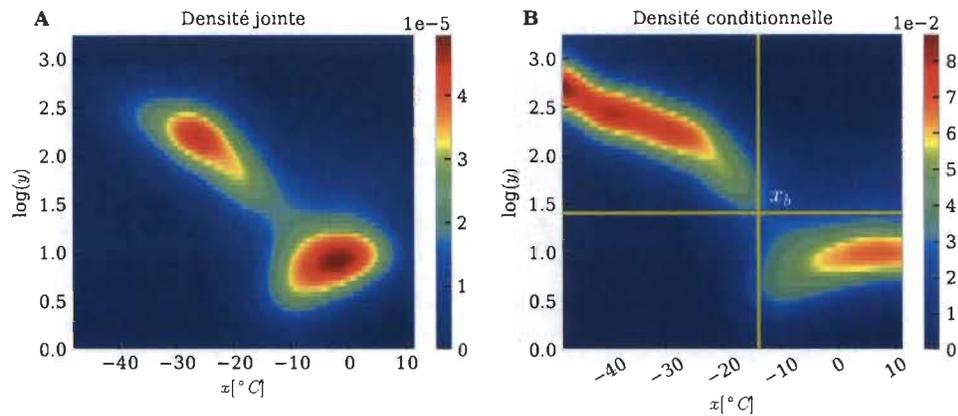


Figure. 3-6 La densité jointe et conditionnelle de la demande d'électricité estimées après une transformation logarithmique : (A) jointe, (B) conditionnelle.

étudiée, la température  $x_b$  interprète l'arrêt d'utilisation du système de chauffage qui se produit pour un écart de température de  $-13^{\circ}\text{C}$  qui correspond à une température extérieure de  $8^{\circ}\text{C}$ . Le complément de chaleur nécessaire pour maintenir la température intérieure est fourni par les apports internes liés à l'occupation, et l'utilisation des autres charges, ainsi que les apports externes liés à l'ensoleillement [21].

Dans le scénario de la transformation logarithmique, la température d'équilibre est plus visible et interprétable sur le map obtenu après la transformation logarithmique (Figure. 3-6). Toutefois, la température d'équilibre demeure un paramètre inconnu et spécifique pour chaque bâtiment résidentiel étudié. Elle dépend de plusieurs facteurs, parmi l'isolation du bâtiment qui a un impact sur la déperdition thermique, et les variations de la température intérieure. Nous présentons dans la sous-section 3.3.3, le Tableau 3-1 qui donne les valeurs moyennes de la température d'équilibre pour chacun des bâtiments.

Afin de démontrer l'hypothèse de la capacité de la transformation logarithmique dans la stabilisation de la fluctuation de variance, nous prenons dans l'exemple illustré sur la Figure.3-7 l'évolution de la distribution de la densité conditionnelle de la puissance en connaissant l'écart de température. Dans la Figure.3-7, la ligne rouge continue correspond à la composante  $h(x_k)$  entourée par des points en verts associés à la mesure de puissance totale. Il est clair que les distributions conditionnelles montrées sur le map (A) ne suivront pas la distribution normale durant la saison chaude comparée à la saison froide. Ainsi, la fluctuation de la variance est faible en hiver par rapport à celle caractérisée par une variabilité plus grande en été. Ceci est dû au fait que la puissance dépend très fortement de la température en période froide, tandis qu'en été, elle dépend en grande partie du comportement et de l'activité stochastique des occupants.

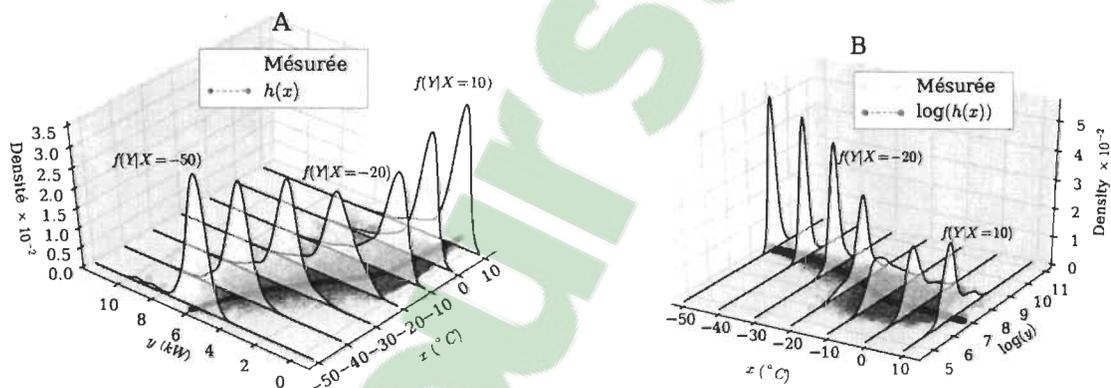


Figure. 3-7 Distributions de la densité conditionnelle sachant l'écart de température : (A) données de puissance, (B) données de puissance transformées.

En effet, sur la même figure, de nombreuses distributions de la puissance totale à l'échelle originale ainsi que logarithmique sont observées pour des valeurs de température connues. Pour bien comprendre l'effet de la transformation, nous prenons l'exemple de la densité conditionnelle de la puissance  $y$  prise à une valeur spécifique d'écart de température  $x$  à 10 degrés notée par  $f(y|X = 10^\circ\text{C})$  tel que montré sur les map (A) et (B) de la Figure.3-7. Nous observons d'abord que la distribution de cette puissance est asymétrique autour de sa moyenne avant la transformation logarithmique. La fluctuation de sa variance se stabilise, et sa distribution suivra une forme symétrique pour cette température et ceci après l'application de la fonction logarithmique. Il est à noter qu'une telle amélioration peut faciliter le processus

de modélisation de la demande dans la saison chaude qui dépend principalement de l'activité des occupants.

### 3.3.3 *Évaluation de la performance thermique du bâtiment*

L'étude de la corrélation entre la puissance totale et l'écart de température permet de déterminer le comportement thermique et énergétique de la maison unifamiliale étudiée. Plus spécifiquement, la composante thermosensible identifiée à partir de laquelle nous pouvons calculer le coefficient de déperdition thermique équivalent  $UA$  et ainsi évaluer approximativement la performance d'isolation thermique du bâtiment. En fait, l'évaluation exacte de cette isolation nécessite des informations à savoir : la superficie totale de l'enveloppe, les murs, le type et la superficie de fenestration du bâtiment, etc. Étant donné l'indisponibilité de ces éléments avec la modélisation boîte noire, nous avons essayé avec le minimum de mesures dont nous disposons de développer des connaissances sur les maisons étudiées en évaluant le niveau d'isolation globale de la résidence par le calcul du coefficient de déperdition équivalent tel que décrit dans l'Annexe C. À partir des données mesurées de la puissance totale des 5 maisons disponibles, nous avons appliqué une campagne de modélisation en utilisant l'approche EADC sur toutes les maisons unifamiliales en identifiant la composante thermosensible par la même relation définie par (3-8). Il est à noter que le même exercice a été répété sur une période d'un an en 1995. Dans cette période, les valeurs de la fenêtre de lissage et le facteur d'apprentissage du modèle EADC ont été ajustés pour chacun des bâtiments. Parallèlement, la composante  $h(x_k)$  a été capturée pour l'ensemble des bâtiments afin d'évaluer leur performance thermique, en calculant la valeur du coefficient déperdition global,  $UA$  en utilisant la relation (3-16) :

$$UA = \frac{\Delta h(x)}{\Delta x} \quad (3-16)$$

Où,  $\Delta h(x)$  et  $\Delta x$  sont les variations en puissance et en température, respectivement. En outre, afin de déterminer la valeur de la constante  $UA$ , le critère d'orthogonalité a été vérifié. Ce critère a été établi entre les observations de l'écart de température et le résiduel produit entre le modèle EADC et les mesures de la puissance totale. La résultante de l'orthogonalité peut démontrer si la variable de l'écart de température est non corrélée au résiduel, en se basant

sur l'évaluation de la corrélation  $\text{corr}(x, r)$ . Par conséquent, une corrélation de valeur faible (c'est-à-dire proche de zéro) démontre que le modèle EADC est capable de générer deux composantes  $h(x_k)$  et  $r_k$  convenablement décorréliées. Ceci veut dire aussi que  $h(x_k)$  a été bien capturée avec un effet minime des facteurs indépendants de l'écart de température. Ainsi, ce test peut montrer que l'incertitude du modèle EADC a été minimisée, et cela par rapport aux impacts d'ajustement des paramètres de la fenêtre de lissage  $\Sigma$ , le facteur d'apprentissage  $\alpha$  ainsi que les pas  $\Delta_x$  et  $\Delta_y$  choisis de la discrétisation des variables continues qui constituent la grille sur laquelle les densités ont été estimées. Ce test considère le calcul de la covariance de l'écart de température  $x_k$  et de résiduel  $r_k$  pour mesurer le taux de décorrélation exprimé par (3-17) :

$$\text{corr}(x, r) = \frac{\mathbb{E}([x - \mathbb{E}(x)][r - \mathbb{E}(r)])}{\sigma_x \sigma_r} \quad (3-17)$$

Où  $\sigma_x$  et  $\sigma_r$  sont les écarts-types de la température et le résiduel, respectivement. En effet, à partir de l'équation (3-16), nous avons tracé la courbe des variations de puissance en fonction des variations de l'écart de température présentée dans la Figure. 3-8. Nous avons sélectionné ensuite l'évolution de la constante de déperdition thermique équivalente  $UA$  en hiver où une grande partie de la puissance dépend principalement de la température extérieure. Dans cette période, nous pouvons déduire des informations sur l'isolation de la résidence étudiée. Techniquement, cette période a été sélectionnée afin de contourner les effets indépendants de la température extérieure. Plus spécifiquement, minimiser les effets introduits par les gains solaires et l'impact des activités des occupants, tels que l'ouverture des portes et des fenêtres entre autres. Donc le coefficient de déperdition obtenu sera moins biaisé et l'étude du comportement thermique de la résidence sera moins difficile. De la Figure. 3-8, nous pouvons déduire la valeur moyenne, minimale et maximale de  $UA$ . Ainsi, la valeur moyenne se situe plus sur la région stable. Nous pouvons constater que la valeur maximale de coefficient est obtenue sur la région qui s'approche de la température d'équilibre thermique  $x_b$ , et ceci parce que la probabilité que le chauffage électrique arrête de fonctionner à la température  $x_b$  est faible. Tandis que plus loin de ce point, dans la zone autour de moins  $7^\circ\text{C}$  de la température extérieure la probabilité est assez élevée.

La Figure. 3-9, montre l'ensemble des profils des composantes thermosensibles capturées pour les 5 maisons par le modèle EADC. À partir de ces composantes, nous présentons dans la

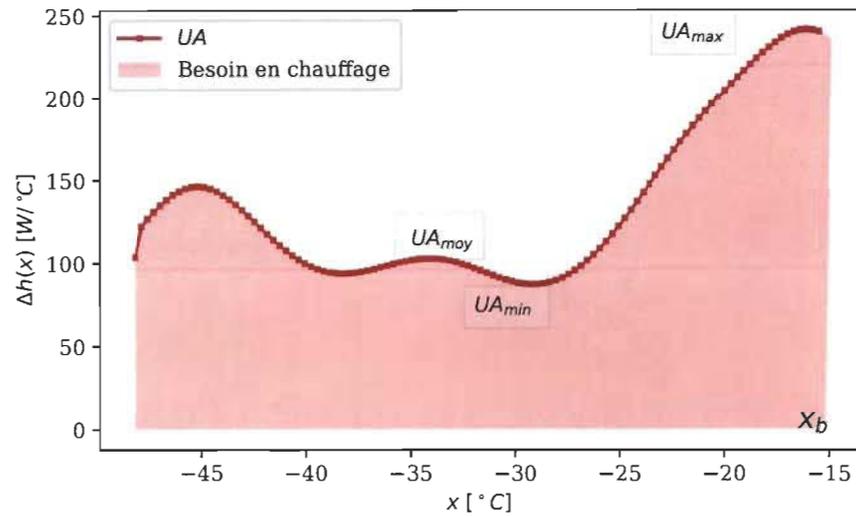


Figure. 3-8 L'évolution de coefficient de déperdition  $UA$  pour la maison étudiée.

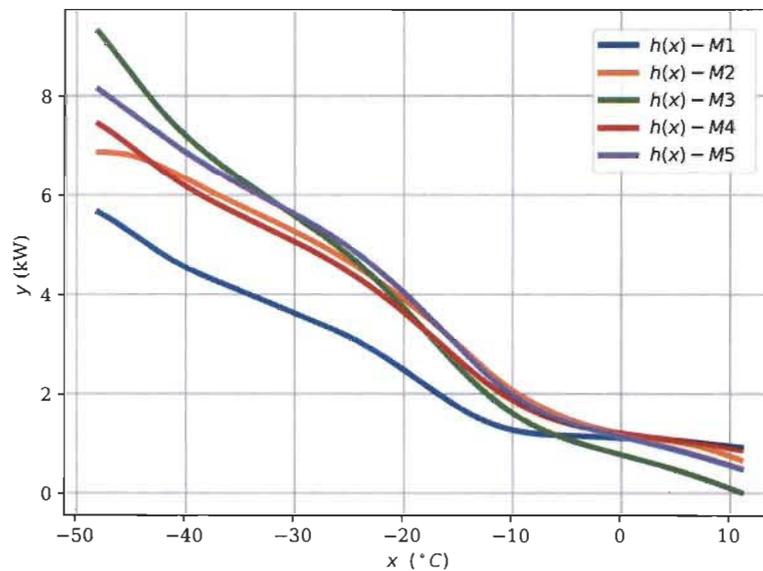


Figure. 3-9 Profils de la composante thermosensible des 5 maisons.

Figure. 3-10 les déperditions thermiques de l'ensemble de ces maisons en fonction des écarts de températures extraites.

De ces graphiques, nous observons que l'évolution du coefficient de déperdition de la maison 2 est différente comparée aux autres maisons. Son coefficient tend vers zéro pour des écarts de températures très basses. Cela signifie que pendant les jours de grand froid, cette maison a une variation d'écart de température mais sa puissance électrique demandée demeure

constante. Cela est expliqué par la saturation des systèmes de chauffage électrique de cette maison, et puisque la puissance ne diminue pas cela veut dire que la maison est probablement dotée d'un chauffage d'appoint (pas à l'électricité) comme le chauffage à gaz, ou au bois qui peuvent combler les besoins en chaleur durant le froid extrême.

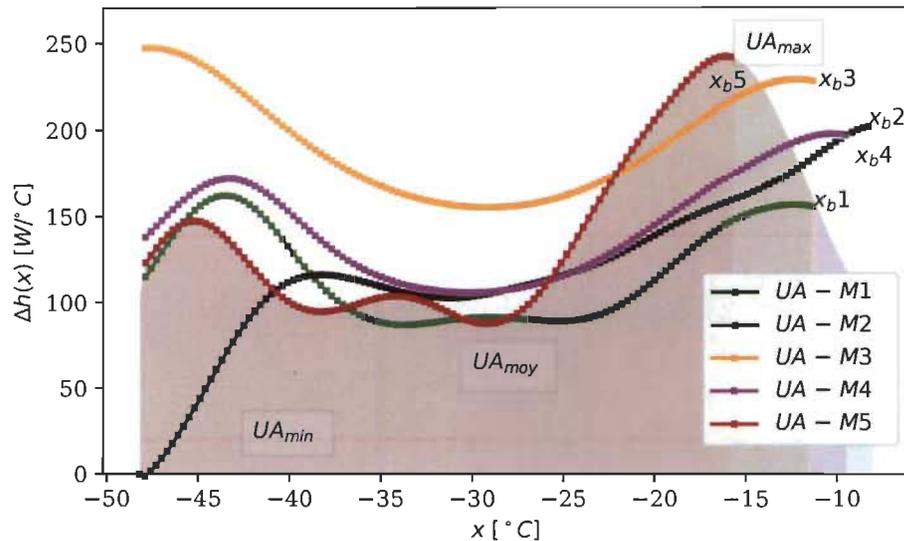


Figure. 3-10 L'évolution de coefficient de déperdition  $UA$  pour l'ensemble des maisons.

Le Tableau 3-1 montre les résultats numériques de test d'orthogonalité entre l'écart de température  $x_k$  et le résiduel  $r_k$ . Nous trouvons également les températures d'équilibre  $x_b$ , la moyenne ainsi que l'écart minimal et maximal des valeurs  $UA$  déduites des composantes thermosensibles des 5 maisons unifamiliales. Le calcul de ces paramètres permet de comparer les 5 maisons du point de vue de leur isolation thermique. Il faut préciser que les températures

TABEAU 3-1 Valeurs moyennes des constantes  $UA$  de déperdition thermique de l'ensemble des maisons

Maison	$\text{corr}(x, r)$	$UA_{\text{moy}}$ (W/°C)	$UA_{\text{min}}$ (W/°C)	$UA_{\text{max}}$ (W/°C)	$x^{\text{ext}}$ (°C)
1	0.0391	117.6	86.6	161.5	10
2	0.0280	117.9	0.0	200.1	13
3	0.0552	191.50	154.1	246.5	10
4	0.0155	143.40	105.5	196.6	12
5	0.0251	133.9	87.8	241.1	7

d'équilibre des maisons, sont pour des températures extérieures  $x_k^{ext}$  qui se situent en moyennes entre  $7^{\circ}C$  et  $13^{\circ}C$ . Il est à noter qu'au dessous de ces températures d'équilibre, le système de chauffage est requis pour maintenir les conditions intérieurs. Par contre, au dessus de ces températures, il y a une forte probabilité que le chauffage électrique arrête de fonctionner. Une autre propriété des températures d'équilibre des cinq maisons étudiées montre que plus la température d'équilibre s'approche des températures extérieures négatives, plus la maison est caractérisée de manière globale par une bonne isolation et une grande capacité thermique. Ceci s'applique plus sur la maison numéro 5 où sa température d'équilibre indique une meilleure isolation comparée aux autres résidences.

*Cas d'une température intérieure variable :* La corrélation entre la puissance totale et l'écart de température permet d'estimer la valeur de la déperdition thermique globale. D'autre part, l'incorporation de la donnée réelle de la température intérieure dans l'analyse de la corrélation aide aussi à déduire la capacité thermique  $C$  d'un ensemble de résidences. Cette constante est obtenue en faisant le produit suivant :

$$C = UA \cdot \tau_c \quad (3-18)$$

Ceci peut se réaliser si une coupure d'électricité de 7 heures minimum par exemple est effectuée sur les bâtiments. Pour cela, un scénario à suivre pour estimer le niveau d'inertie thermique d'un bâtiment consiste à effectuer une panne de courant pour quelques heures en mesurant le taux de décroissance de la température intérieure. À partir de la variation de cette dernière, nous pouvons déduire la constante de temps  $\tau_c$  à l'aide de la relation suivante [119] :

$$\frac{(x^{int,k} - x^{int,f})}{(x^{int,i} - x^{int,f})} = \exp\left(-\frac{UA}{C}k\right) \quad (3-19)$$

Où  $x^{int,k}$  est la température intérieure à l'instant  $k$ ,  $x^{int,i}$  la température intérieure initiale au début de la coupure de courant et  $x^{int,f}$  la température intérieure finale à un instant infini après le commencement de la coupure. Cette dernière correspond à la température qui prévaut à l'extérieur pendant la coupure.  $UA$  est le coefficient de transfert de chaleur global du bâtiment tel que spécifié à la section 3.3.3,  $C$  la valeur de la capacité thermique du même bâtiment et

finalement  $k$  le temps écoulé depuis le début de la coupure. En réorganisant l'équation 3-19 [119], nous obtenons :

$$\ln \frac{(x^{int,k} - x^{int,f})}{(x^{int,i} - x^{int,f})} = \frac{UA}{C} k = \frac{k}{\tau_c} \quad (3-20)$$

Où le rapport  $\frac{C}{UA}$  représente la constante de temps  $\tau_c$ . En connaissant les valeurs de  $UA$  et de  $\tau_c$  pour chacun des bâtiments, nous pouvons déduire la capacité thermique par la relation  $C = UA \cdot \tau_c$ .

### 3.4 Résultats et discussions

#### 3.4.1 Méthodes de comparaison

Il est à noter que les résultats et les discussions présentés dans cette section sont liées à la résidence numéro 5 comme un exemple sur l'ensemble de tests et validations faites sur les cinq maisons.

En effet, afin d'évaluer la performance de l'approche EADC, une étude comparative a été appliquée dans cette section avec des modèles adaptatifs linéaires et non linéaires basés sur le filtre RLS et un modèle de réseaux de neurone ANN, en analysant leur capacité à identifier la thermosensibilité de la demande d'électricité. Le ANN utilisé est de type Multi-Layer Perceptron (MLP) dont la fonction d'activation est la tangente hyperbolique. La méthodologie de calcul du facteur d'apprentissage a été décrite dans la référence [120]. Il est à noter que cette analyse a été réalisée pour la maison numéro 5 et que le choix de la structure de MLP n'est pas aléatoire. Plusieurs structures ont été testées de manière manuelle pour trouver la bonne topologie qui converge bien avec la sortie du modèle. Au début des tests, le réseau avait une structure qui possède un nombre de couches et de neurones proche du nombre d'entrée. Ensuite, un nombre supérieur de neurones avec des couches supplémentaires ont été ajoutés en vérifiant à chaque fois si la sortie du modèle a été améliorée par une diminution de l'erreur. À la fin et pour le cas de la maison 5 présentée dans ce manuscrit, c'est avec la structure présentée en dessous qui donne une faible erreur qui a été adoptée pour réaliser la comparaison avec l'approche proposée. Elle possède deux couches cachées avec 7 neurones et un facteur d'apprentissage  $\alpha = 0.012$ . Le réseau a une couche de sortie associée à la puissance totale  $y_k$  et six entrées des écarts de température  $x_k$ . Le modèle linéaire établi avec l'algorithme RLS et

MLP s'appuie sur des équations récurrentes des écarts de température, définies par le vecteur d'entrée  $x_k$  suivant :

$$x_k = [1, x_k, x_{k-1}, \dots, x_{k-5}]^T \quad (3-21)$$

Le premier échantillon de l'écart de température a été pris à l'instant  $k$ , les cinq autres entrées représentent les échantillons horaires antérieurs de la température, avec une constante de 1 ajoutée pour que  $\mathbb{E}[e_k]$  définie ultérieurement, soit nulle, donc si  $\mathbb{E}[e_k]$  valait 1 cela veut dire que nous avons oublié une constante dans le vecteur d'entrée. Le modèle RLS, a été implémenté selon [121], qui considère le même vecteur d'entrée de l'équation (3-21), sa sortie est définie par :

$$y_k = w_k x_k \quad (3-22)$$

Où  $w_k$  représente le vecteur des paramètres du filtre RLS, avec  $w_k = [w_{0k}, w_{1k}, w_{2k}, \dots, w_{6k}]^T$ . La mise à jour de ces paramètres est réalisée selon (3-23) :

$$w_{k+1} = w_k + R_k x_k^T e_k \quad (3-23)$$

Où  $e_k$  est l'erreur et  $R_k$  présente la matrice d'autocorrélation inverse [121], qui est définie par :

$$R_k = \frac{1}{\mu} \left( R_{k-1} - \frac{R_{k-1} x_k^T x_k R_{k-1}}{\mu + x_k R_{k-1} x_k^T} \right) \quad (3-24)$$

Le paramètre  $\mu$  est le facteur d'oubli dont la valeur est comprise entre 0 et 1. La valeur de ce paramètre pour ce cas d'étude est de  $\mu = 0.92$ .

Il est à noter que nous avons réalisé le processus de prévision de  $h(x_{k+\tau})$ , en supposant d'abord une prévision avec les valeurs exactes des températures ensuite avec les valeurs futures. Nous présentons dans la section 3.4.1.2 les prévisions de la composante thermosensible en utilisant les modèles décrits ci-dessus sur les prochaines 24 heures.

### 3.4.1.1 Évaluation de la précision de la prévision

Pour mesurer la précision de la prévision de la demande d'électricité, nous proposons la métrique de l'erreur absolue normalisée (NMAE). Cette métrique a été utilisée pour mesurer l'erreur et obtenir un résumé clair sur sa distribution par rapport à la puissance électrique observée. Cette métrique considère l'erreur absolue de prévision normalisée par la somme des mesures de la demande d'électricité, comme est défini en (3-25). Cette métrique a été choisie, car elle ne donne pas des valeurs infinies ou indéfinies pour le cas de la série de la puissance totale étudiée. Ceci est vrai seulement dans le cas où toutes les données historiques de la demande ne sont pas égales à zéro. Dans le cas contraire Hyndman et Koehler [122] propose une alternative à cette métrique qui sera décrite dans l'Annexe C.

$$NMAE = \frac{\sum_{k=1}^N |y_k - h(x_k)|}{\sum_{k=1}^N y_k} \quad (3-25)$$

Où  $y_k$  est la puissance totale mesurée à chaque instant  $k$ .

### 3.4.1.2 Résultats d'un cas d'étude

La Figure. 3-11 ci-dessous nous fournit la composante thermosensible prédite sur un horizon  $\tau = 24$  par les trois modèles en utilisant l'historique des températures. Il est clair que le modèle EADC produit la meilleure performance qui décrit une relation non-linéaire réaliste entre la demande totale d'électricité et l'écart de température. Comparer ce résultat avec celui obtenu par le modèle RLS, démontre que ce dernier est incapable de traiter les problèmes de la relation non-linéaire qui apparaît dans la composante  $h(x_k)$ . D'un autre côté, le modèle MLP tente de suivre la tendance de la forme non-linéaire des changements en puissance liée à la température. Cependant, sa sortie est toujours incapable de fournir des prévisions réalistes proches des résultats de EADC.

De plus, la corrélation entre l'écart de température et le résiduel a été évaluée avec les trois modèles afin de vérifier si la composante  $h(x)$  a été capturée de manière efficace. Pour cela, nous avons mesuré l'orthogonalité des deux variables aléatoires à chaque horizon de prévision et nous avons tracé par la suite les courbes d'analyse de corrélation présentée dans la Figure. 3-12. Le résultat de corrélation du modèle MLP est meilleur dans les horizons les plus

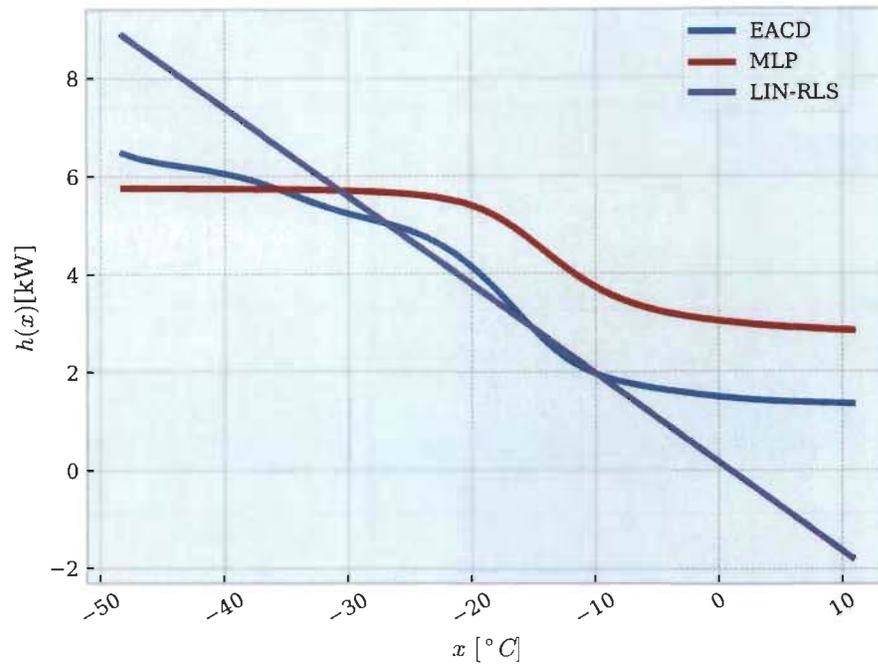


Figure. 3-11 Performance des trois modèles utilisés dans la prévision de  $h(x)$  sur un horizon de  $\tau = 24$  heures. MLP (2 couches, 7 neurones, et un facteur d'apprentissage  $\alpha = 0.012$ ), RLS ( $\mu = 0.92$ ).

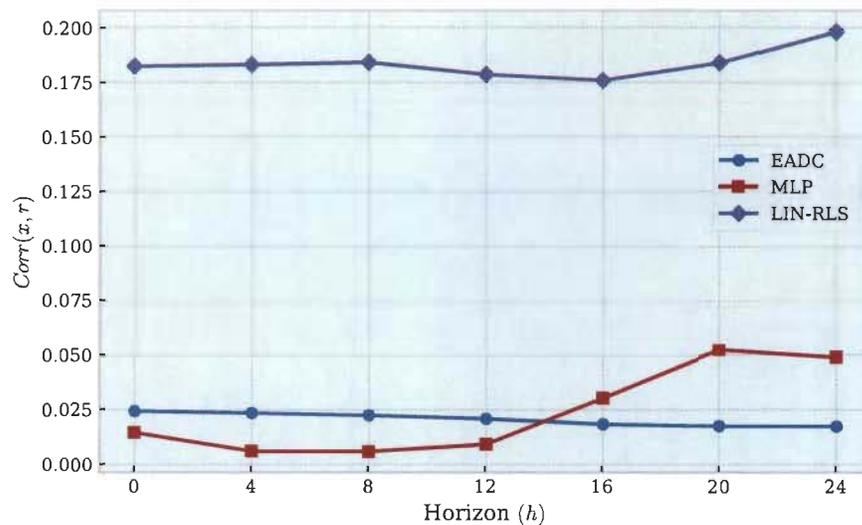


Figure. 3-12 Comparaison des résultats d'analyse de corrélation entre l'écart de température et le résiduel des trois modèles.

courts comparés au modèle EADC, ses valeurs de covariance augmentent considérablement après quelques heures. Le modèle EADC a pour résultat une consommation thermosensible bien décorrélée de la part résiduelle avec des covariances petites sur toutes les 24 heures,

tandis que, la performance de la sortie de MLP se dégrade quand l'horizon s'éloigne.

D'autre part, le modèle EADC, commence avec une covariance plus au moins petite, sa courbe de covariance diminue jusqu'à une valeur de 0,0236 à l'horizon de  $\tau = 24$  heures. En conséquence, le modèle proposé fournit une corrélation stable sur tous les horizons avec une valeur de covariance remarquablement faible.

Nous employons la métrique NMAE comme un autre indicateur pour évaluer l'erreur de la prévision et déduire la précision d'identification et la précision de la prévision de la composante thermosensible en termes de pourcentage d'erreur.

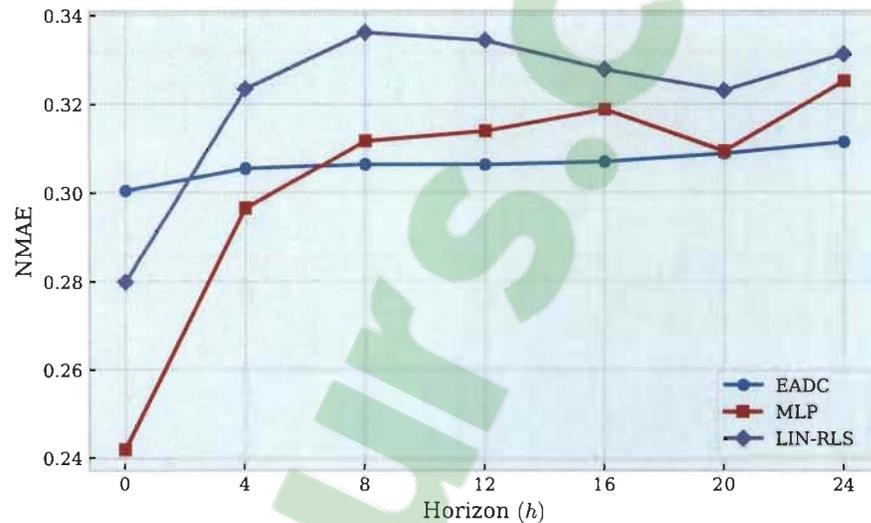


Figure. 3-13 Comparaison de la précision de la prévision des modèles.

Nous observons sur la Figure. 3-13 une différence notable dans la précision des trois modèles analysés. Les modèles adaptatifs linéaires et MLP indiquent une bonne précision durant les trois premières heures. Cependant, leur NMAE augmente rapidement en comparaison aux NMAE de EADC qui est capable d'expliquer environ 68% de la demande en électricité sur plusieurs horizons. En effet, la stabilité de EADC, démontre qu'il possède une capacité consistante pour capturer une relation réaliste de la composante  $h(x_k)$  sur l'ensemble des horizons.

En s'appuyant sur les statistiques de la répartition de la puissance d'électricité par usage au Québec [9], la valeur prédite de la composante thermosensible peut être expliquée par les faits suivants : 64% de cette proportion est liée essentiellement à la consommation annuelle

moyenne de chauffage électrique connexe à la maison québécoise étudiée, où environ 4% de la consommation associée aux autres charges électriques (p.ex. la climatisation) ont un lien avec la température et contribuent probablement au chauffage de la maison.

### 3.4.2 Analyse de sensibilité

La prévision de  $h(x_{k+\tau})$  faite jusqu'à présent a été effectuée et évaluée avec l'historique de la température mesurée. Nous répétons le même exercice, mais cette fois-ci nous considérons la présence des incertitudes sur des valeurs futures de la température extérieure qui proviennent des différentes stations météo. Ainsi, nous analysons leurs impacts sur la sortie du modèle EADC proposé, à partir d'une analyse de sensibilité. Pour cela, nous avons généré une température extérieure bruitée de distribution normale et de variance  $\sigma_{x^{ext}}$ , de telle sorte que  $\delta_k \sim \mathcal{N}(0, \sigma_{x^{ext}}^2)$ . Nous ajoutons ce bruit aux valeurs des températures mesurées définies en (3-26). Cette analyse de sensibilité est fondée sur la considération de l'incertitude de prévision de température fournie par 5 stations météo sur un horizon de 24 heures, comme indiqué dans le Tableau 3-2 [33]. Ce dernier montre l'écart type de la prévision en degrés pour chaque station météo sur plusieurs horizons. Cette analyse permet alors de quantifier la composante thermosensible et de déterminer quels sont les écarts-types moyens des températures des cinq stations météo ayant le plus d'impact sur la thermosensibilité de la demande, en considérant les modèles EADC, MLP et RLS.

$$x_{k+\tau}^{ext} = x_{k+\tau}^{ext} + \delta_{k+\tau} \quad (3-26)$$

TABLEAU 3-2 Les écarts-types moyens de la température extérieure fournis à partir des cinq stations météo [33].

Station météo	L'écart-type moyen journalier de la température $\sigma_{x^{ext}}^2$		
	1-2 jours prévisions	3-4 jours prévisions	5-7 jours prévisions
Environment Canada (EC)	2.5	3.2	3.5
Weather Network (WN)	2.6	3.0	3.8
Weather Underground (WU)	1.5	2.4	3.7
Meteoblue (MB)	1.9	3.1	4.3
Accuweather (AW)	1.7	2.5	3.6

### 3.4.2.1 Résultats d'un cas d'étude

Les résultats de l'analyse de sensibilité obtenus à partir des trois modèles sont illustrés sur la Figure. 3-14. Cette dernière reflète l'impact des incertitudes introduites sur les variations de la température extérieure, et donc sur les composantes thermosensibles produites par les modèles. D'un point de vue général, le profil de la thermosensibilité prédit par le modèle EADC possède une allure très similaire au profil réel pour le scénario sans incertitude présenté dans la Figure. 3-12. Cependant, l'allure des prévisions avec les modèles MLP et RLS montre quelques variations différentes à celles présentées dans la Figure. 3-12.

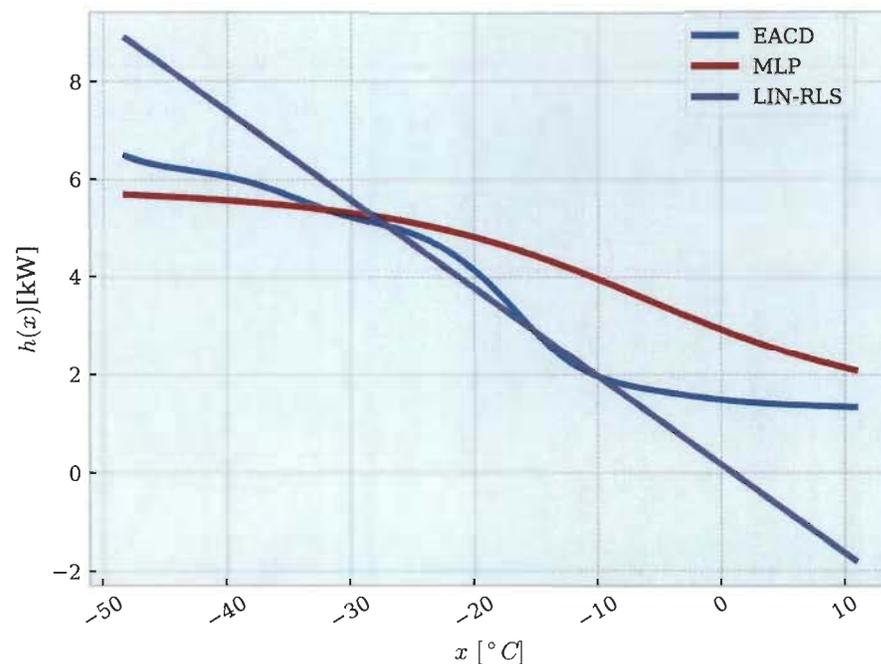


Figure. 3-14 Impact de l'incertitude de la température sur la composante  $h(x)$ .

L'analyse de corrélation entre les températures des cinq stations météo et le résiduel a été effectuée pour chaque modèle. Les résultats sont illustrés dans la Figure. 3-15. Cette dernière montre que le modèle EADC donne les meilleurs résultats pour toutes les stations de prévision avec une faible covariance qui est maintenue sur tout l'horizon. Ainsi, la précision en présence d'incertitude de température en termes de NMAE sur un horizon de 24 heures pour les cinq sites est présentée dans la Figure. 3-16. L'erreur NMAE avec le modèle EADC est petite, et ceci en présence des incertitudes de prévision de température des 5 stations météo considérées. Ceci justifie que le modèle EADC est moins affecté par les incertitudes

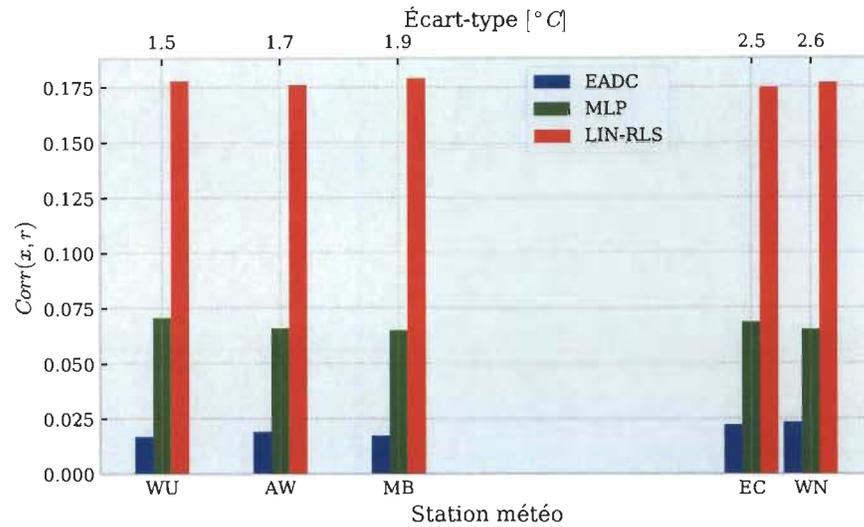


Figure. 3-15 Comparaison de l'analyse de corrélation des trois modèles en fonction des écarts-types des cinq stations météo à  $\tau = 24$  heures.

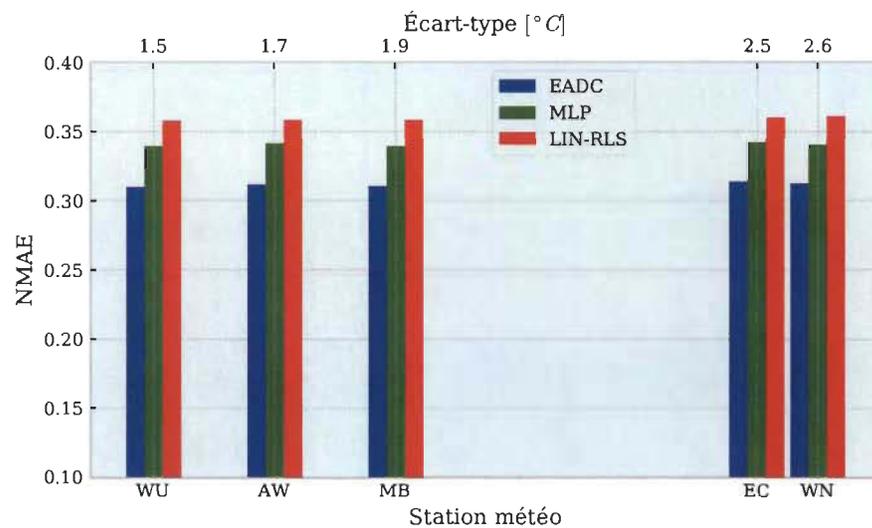


Figure. 3-16 Comparaison de la précision de la prévision des trois modèles en fonction des écarts-types des cinq stations météo à  $\tau = 24$  heures.

des prévisions de température considérant les écarts-types de cinq sites de prévision. Pour ce scénario, la puissance totale expliquée est environ 68%, encore très similaire au profil de la thermosensibilité sans incertitudes.

Les résultats de la prévision de la demande d'électricité totale obtenus à partir des modèles EADC, MLP et RLS sur l'horizon temporel de 24 heures sont présentés sur la Figure. 3-17.

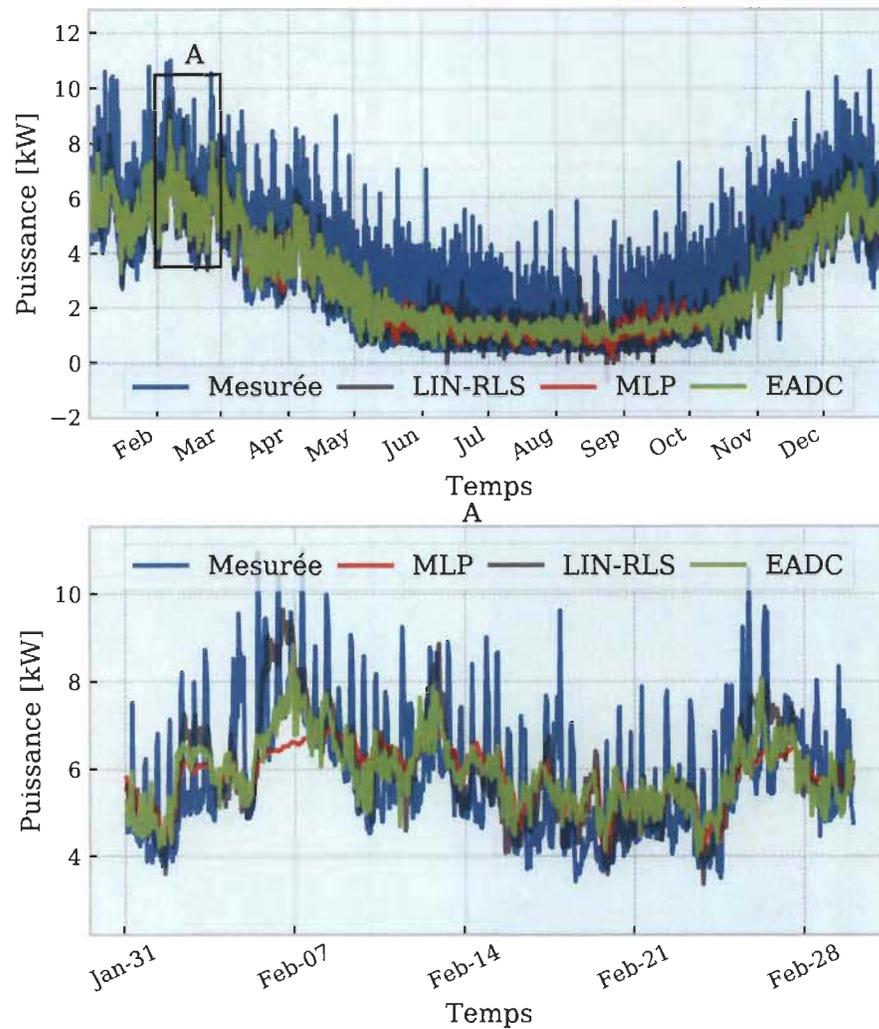


Figure. 3-17 La puissance mesurée et prédite avec les trois modèles sur un horizon  $\tau = 24$  heures (A) zoom d'un mois en hiver.

Ces résultats correspondent à la demande énergétique d'une maison unifamiliale suite à une année de simulation du mois de janvier au mois de décembre 1995. Il est observé que le modèle linéaire adaptatif RLS ne concorde pas au profil de la puissance totale, et le modèle MLP n'offre pas une meilleure performance comparée au modèle EADC. La capacité de ce dernier à suivre les variations de la puissance totale, en particulier pendant la période froide est justifiée par la forte corrélation entre l'appel de puissance et la température comme c'est illustré en détail sur la Figure. 3-17-(A). En effet, le EADC peut expliquer avec succès la contribution de la composante thermosensible dans la demande d'électricité totale, dont la température extérieure demeure le facteur prépondérant en hiver dans la région de Québec.

Il est à noter que, si le modèle EADC produit en particulier une bonne prévision de la demande totale, c'est parce qu'il a la capacité de fournir une estimation entière de la densité conditionnelle sur une grille bien adaptée aux données observées, comparée au modèle MLP caractérisé par une estimation ponctuelle de l'ensemble des valeurs [123]. En outre, le modèle EADC est non paramétrique, sa modélisation pour la prévision de la puissance ne requiert aucune définition d'ordre de paramètres ou une connaissance préalable de la nature des relations entre la demande et les facteurs explicatifs. Au contraire le modèle MLP nécessite l'identification du bon nombre de neurones dans chaque couche, d'une manière qui peut justifier l'efficacité de sa topologie à produire des performances souhaitables de prévision.

Cependant, le modèle EADC démontre des difficultés à capter les pointes de la demande en raison de l'augmentation de la puissance d'électricité totale causée par l'utilisation des charges non liées aux variations de la température et ainsi aux comportements stochastiques des occupants. Un tel comportement nécessite un autre processus de modélisation comme l'analyse de la composante résiduelle, qui est plus corrélée avec les activités des occupants notamment pendant la période estivale. Par conséquent, une analyse des résidus dans le chapitre suivant peut être envisagée comme un processus complémentaire au modèle EADC qui mène à une amélioration de précision de la prévision. D'autre part, une tâche de prévision réaliste requiert à la fois deux processus de modélisation, un lié à la composante thermosensible et l'autre correspond à l'analyse de la composante résiduelle en raison de sa capacité de développer des connaissances en scrutant le comportement de l'occupant, et l'utilisation d'électricité, entre autres.

### 3.5 Synthèse du chapitre

Ce chapitre présente une approche explicative et prédictive de la demande énergétique à court terme d'une maison unifamiliale par régression non paramétrique. La méthodologie présentée permet d'établir une prévision adaptative sur les vingt-quatre prochaines heures en utilisant des données de puissance horaire et de température extérieure d'une année de mesure. La méthodologie proposée utilise d'une part une régression non paramétrique bi-variable pour l'estimation en ligne de la densité jointe entre la demande et l'écart de température. D'autre part, une régression non paramétrique mono-variable est employée afin d'estimer la

densité marginale de l'écart de température. À partir de ces deux densités, une Estimation Adaptative de la Densité Conditionnelle EADC a été déduite, pour expliquer la composante thermosensible. Cette dernière est la consommation thermosensible liée aux systèmes de chauffage, à la climatisation et à toutes charges affectées directement par les variations de température extérieure.

De plus, à partir des caractéristiques de la composante thermosensible, le modèle explicatif a montré une contribution originale de passer d'un modèle type boîte noire à un modèle gris en offrant des connaissances au client sur sa consommation thermosensible, la température d'équilibre de sa résidence ainsi que la constante de déperdition thermique équivalente  $UA$  du bâtiment. L'évaluation de la performance de la prévision avec le modèle EADC par comparaison avec les modèles adaptatifs linéaires et non linéaires de type RLS et MLP a montré que l'approche proposée est plus favorable pour la tâche de prévision en plusieurs heures en avance. Les résultats de l'étude comparative avec les données réelles mesurées sur des bâtiments résidentiels québécois ont montré que le modèle EADC peut produire une précision améliorée des profils de la puissance prévue sur tous les horizons par rapport aux autres méthodes.

Le chapitre suivant traitera de l'amélioration de la précision de la prévision de la puissance totale en intégrant l'analyse du résiduel. En effet, l'utilisation de la variable temporelle autrement dit l'effet calendaire comme un outil permet de capter l'activité périodique des occupants et d'obtenir une idée sur les profils cycliques d'appel de puissance pour les phénomènes indépendants de la température. Et ceci pourra être utilisé pour justifier ce que la température n'arrive pas à expliquer.

## **Chapitre 4 - Modélisation et prévision de la composante résiduelle**

### **4.1 Introduction**

Nous avons étudié, dans le chapitre précédent, les effets de la température extérieure sur la puissance électrique. Nous avons montré aussi que la consommation électrique Québécoise est fortement thermosensible en raison de la part importante du chauffage électrique lors des vagues de froid hivernal. L'autre partie non thermosensible est obtenue par l'écart entre le modèle et la mesure de la consommation totale. Cette partie représente la composante résiduelle, inexplicable par la température extérieure.

Ce chapitre portera sur la modélisation et la prévision de la composante résiduelle obtenue selon les résultats du chapitre 3. Le résiduel est une composante dynamique, qui est en interaction continue avec l'aspect comportemental, l'environnement lié aux conditions météorologiques ainsi que les variations dues à la masse thermique du bâtiment. À cet effet, le résiduel est considéré de nature plurielle, il peut se décomposer en quatre éléments : la tendance, la part cyclique, la part saisonnière et la part irrégulière. Parmi celles-ci, c'est sur la part cyclique que se concentre l'étude du chapitre 4.

Bien que le résiduel est défini comme un ensemble des phénomènes non-déterministes, sa composante cyclique est soumise à un rythme périodique, calé sur des cycles journaliers, hebdomadaires ou annuels. Or, les fluctuations de cette composante indiquent que les phénomènes proviennent à la fois des changements saisonniers de la météo et des rythmes de vie journaliers des occupants.

Tel que mentionné dans la section 2.1 du chapitre 2, l'horizon de prévision dans le contexte de la GLDD d'énergie est de court terme. Il est donc nécessaire de mieux gérer la

consommation résiduelle via des cycles saisonniers de l'ordre journalier et hebdomadaire. Pour ce cas, le cycle journalier est supposé dans cette étude pour refléter les pointes ou les heures de forte demande, ainsi que, les creux de consommation au cours de la journée. Tandis que, le cycle hebdomadaire a été supposé pour refléter les variations de la demande d'électricité entre jours ouvrables et le week-end [21].

Or, la tendance de la composante résiduelle est non-linéaire (non-stationnaire) à court terme, donc elle a peu d'effet sur les fluctuations de résiduel. Dans ce sens, il est difficile d'estimer séparément la part tendance de la part cyclique. Cependant, elles sont supposées fusionnées pour être étudiées ensemble comme une composante tendance-cycle. Les deux autres parts saisonnière et irrégulière sont indépendantes de la part tendance-cycle, donc, ce sont elles qui masquent l'état de cyclicité du résiduel.

Le concept de modélisation de la composante résiduelle étudié dans ce chapitre ne se limite pas à faire ressortir l'effet calendaire cyclique des changements de la consommation résiduelle. Il consiste en fait à diviser le résiduel en deux séries sous-estimé et surestimé et capturer les fluctuations régulières récurrentes dans les deux séries, notamment les changements dans la consommation liés à l'utilisation des charges et aux phénomènes météorologiques.

En effet, ce chapitre est divisé en trois sections : dans la première section, nous présentons tout d'abord les sources d'incertitudes qui causent l'écart dans le processus de modélisation. Puis nous décrivons les caractéristiques et le comportement périodique du résiduel. Dans la deuxième section, l'approche de modélisation sera présentée et détaillée, en considérant la saisonnalité cyclique journalière et hebdomadaire de la demande d'électricité du résiduel. Enfin, les résultats obtenus de cette modélisation seront utilisés pour la reconstruction de la composante résiduelle par le biais d'un modèle linéaire. Une comparaison entre le résiduel original et prédit sera également présentée. Dans la dernière section, nous construisons la prévision de la demande totale d'électricité, en utilisant les résultats de prévision de la consommation thermosensible traitée dans le chapitre 3 et la prévision de la consommation résiduelle. Les résultats globaux seront comparés avec d'autres modèles, afin de conclure les performances des approches de prévision proposées.

## 4.2 Caractéristiques de la composante résiduelle

### 4.2.1 Incertitude agissant sur le résiduel

De la même manière que dans la modélisation de la composante thermosensible, la composante résiduelle est affectée par une combinaison de nombreux phénomènes météorologiques, comportementaux et un ensemble d'incertitudes, comme les incertitudes liées aux choix du modèle, l'incertitude liée à l'estimation des paramètres, l'incertitude liée aux mesures, et les prévisions météo. Par conséquent, la démarche de modélisation de la composante résiduelle ne peut proposer qu'une approximation, c'est-à-dire, sa représentation demeure imparfaite, mais simplifiée par rapport à la réalité.

#### 4.2.1.1 Incertitude liée au modèle et à l'estimation des paramètres

Dans le cas du choix du modèle, les incertitudes de type des modèles paramétriques ou non-paramétriques ne sont pas les mêmes. Le fait de choisir un modèle paramétrique s'accommodera des incertitudes sur le choix de l'ordre des paramètres qui permettra d'obtenir un résultat de calcul se rapprochant au plus près des valeurs mesurées de la puissance d'électricité. Ainsi, plus un modèle paramétrique sera complexe (multivariées), plus il nécessitera une grande quantité de données et un grand nombre de paramètres à ajuster et donc plus d'incertitudes. Dans le cas d'un modèle non-paramétrique, nous avons besoin d'une grille beaucoup plus détaillée et beaucoup plus fine. Il nous faudra alors choisir la bonne valeur de la fenêtre de lissage, car la résolution de la grille peut parfois cacher des incertitudes très importantes sur les résultats [24], [124].

En ce qui concerne les incertitudes sur l'estimation des paramètres, pour notre cas d'analyse non-paramétrique adaptative, il s'agira de l'incertitude sur le choix de la bonne valeur de la fenêtre de lissage et du facteur d'apprentissage. Cependant, dans le cas non adaptatif, seule l'incertitude sur la fenêtre de lissage est considérée.

#### 4.2.1.2 Incertitude liée aux mesures

Les incertitudes sur les données d'observation, comprennent à la fois les incertitudes liées aux erreurs de mesures et également aux instruments utilisés. Par exemple, dans le cas où les

prévisions des données météo sont disponibles, l'incertitude doit alors se baser sur la qualité et la précision des mesures (incertitudes liées aux capteurs utilisés) [125].

Il y a aussi la contribution de l'incertitude sur l'échantillonnage qui peut donner une idée correcte sur toutes les données temporelles. Selon l'application, un pas de temps plus court ou plus large risque de masquer des sollicitations des données très importantes [124].

#### 4.2.1.3 *Incertaince liée aux données météo*

L'incertitude liée aux conditions météorologiques y compris la température extérieure est une des grandes problématiques de la prévision de la demande d'électricité. Dans un contexte de prévision, la prise en compte du futur du rayonnement solaire, les effets du vent, ainsi que la localisation de la station météo peuvent générer des incertitudes non négligeables sur la puissance d'électricité. En outre, la prédominance des incertitudes sur les données météo est basée sur l'horizon temporel de prévision choisi, ce qui permet de caractériser le type d'incertitude à quantifier. En outre, l'incertitude sur les données météo peut être élevée et remarquable, si le but est de prévoir une puissance d'électricité moyenne sur des horizons des semaines, mois voir une année [125].

En pratique, il est difficile de supprimer complètement les incertitudes dans le processus de modélisation et prévision. Cependant, la quantification de cette barre d'incertitudes est incontournable si nous voulons améliorer dans une certaine mesure la qualité de la prévision et les décisions sur la puissance pour une bonne GLDD électricité.

En ce qui concerne la composante résiduelle, la division de cette dernière en résiduel sous-estimé et surestimé demeure nécessaire si nous voulons disposer d'un outil robuste à la quantification des incertitudes liées à chaque série des résiduels. Évidemment, les sources d'incertitudes ne sont pas les mêmes pour chaque série. D'une part, les incertitudes agissantes sur le résiduel sous-estimé sont surtout dues au type comportemental de la consommation qui fait intervenir le nombre, la présence, l'activité de l'occupant. Ainsi que les actions telles que l'ouverture/fermeture des fenêtres qui caractérisent les incertitudes par pertes d'infiltration et de ventilation. D'autre part, les incertitudes du résiduel surestimé sont généralement liées aux gains comportementaux, des appareils électrique et solaire.

#### 4.2.2 *Données et méthodes*

##### 4.2.2.1 *Cyclicité dans la composante résiduelle*

Dans cette sous-section, nous nous intéressons à la description de la part tendance-cyclique de la composante résiduelle qui se répète plus ou moins régulièrement de jour en jour et de semaine en semaine. Cette part regroupe la partie tendance non linéaire d'où l'intérêt adaptatif serait de permettre une adéquation entre la croissance ou la décroissance de la demande d'énergie.

Donc, comme il est difficile de mesurer les variétés comportementales en termes de présence, et d'utilisation des appareils électriques, nous avons proposé d'analyser séparément le résiduel sous-estimé et surestimé pour deux raisons principales, d'abord (i) s'assurer que la cyclicité est présente dans les deux résiduels et (ii) qu'il est possible de les capter correctement en exprimant la relation entre les deux résiduels en fonction du temps calendaire vu comme temps de consommation d'électricité.

Sur le plan pratique, il existe une corrélation de l'activité cyclique entre les phénomènes météorologiques de nature stochastique et le comportement des occupants d'une variance fortement variable. Les exemples de cette corrélation sont innombrables : sous certaines conditions climatiques, en présence des apports solaires, les occupants ont tendance à modifier leur environnement en agissant sur l'ouverture/fermeture des fenêtres, modification de la consigne de chauffage ou de climatisation ou l'usage de ventilateurs.

En revanche, projeter ces exemples sur les résiduels sous-estimé et surestimé s'avère un bon moyen de dégager la part tendance-cycle utile à analyser. Le résiduel sous-estimé est dérivé de la demande d'électricité due à l'utilisation des appareils ménagers. Alors que, le résiduel surestimé est lié aux gains énergétiques générés principalement par le rayonnement solaire à travers les fenêtres et les gains et les actions des occupants comme l'ajustement de thermostats électroniques pendant la nuit.

##### 4.2.2.2 *Données de la composante résiduelle*

La donnée de la composante résiduelle utilisée dans ce chapitre est celle de la maison unifamiliale numéro 5 étudiée dans le chapitre 3 sur la période d'un an en 1995. Cette

composante est obtenue par l'écart entre les mesures réelles de la puissance et le modèle de la thermosensibilité identifiée précédemment donnée par 4-1.

$$r_k = y_k - h(x_k) \quad (4-1)$$

L'approche de modélisation de résiduel est basée sur une division de la composante  $r_k$  en deux séries temporelles données par (4-2) comme l'étape principale de la phase 2 illustrée dans la Figure. 4-1.

$$r_k = r_k^+ + r_k^- \quad (4-2)$$

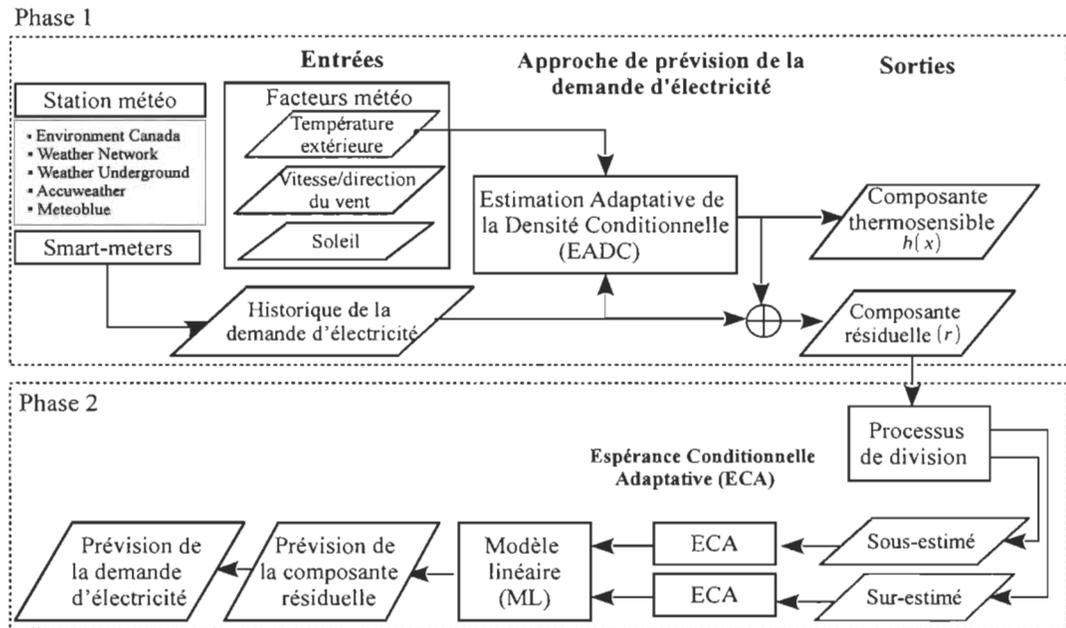


Figure. 4-1 Structure générale de l'approche de la décomposition et prévision de la demande d'électricité résidentielle.

Deux types de résiduels de nature stochastique sont distingués dans la composante  $r_k$ , le résiduel sous-estimé noté  $r_k^+$ , et le résiduel surestimé indiqué par  $r_k^-$ . En effet, le résiduel de la consommation est de moyenne nulle, avec une hypothèse qu'il a une distribution approximative à la distribution gaussienne. La convention des signes pour le concept de résiduel sous-estimé signifie que les valeurs prédites de la composante thermosensible  $h(x_k)$  sont petites aux valeurs mesurées de la puissance totale consommée, et pour le résiduel surestimé spécifier que les valeurs prédites de la thermosensibilité sont élevées à la mesure de la demande.

Les deux séries sont obtenues sous la forme suivante :

$$r_k^+ = \begin{cases} r_k, & \text{si } y_k > h(x_k) \\ 0, & \text{sinon} \end{cases} \quad (4-3)$$

$$r_k^- = \begin{cases} r_k, & \text{si } y_k < h(x_k) \\ 0, & \text{sinon} \end{cases} \quad (4-4)$$

Pour la composante résiduelle  $r_k$ , illustrée dans la Figure. 4-2, nous avons réalisé une analyse graphique qui montre d'une manière générale, l'évolution dans le temps la composante  $r_k$  centrée autour de sa valeur moyenne. Dans sa forme aléatoire, il n'est pas possible d'expliquer tous les phénomènes de résiduel agrégé, mais la grande majorité des cas, les consommations réelles dépassent les prévisions. En revanche, quelques fluctuations apparaîtraient de façon périodique, ce qui peut justifier le processus de division de la composante  $r_k$  pour les raisons suivantes : l'évolution de la puissance d'électricité provenant des pertes de chaleur et l'utilisation des appareils électriques n'est pas symétrique avec celle causée par les gains de chaleur. Ainsi, le résiduel  $r_k^+$  révèle des pics de puissance électrique annuelle élevés, et de courtes durées, tandis que les pics de résiduel  $r_k^-$  sont moins significatifs.

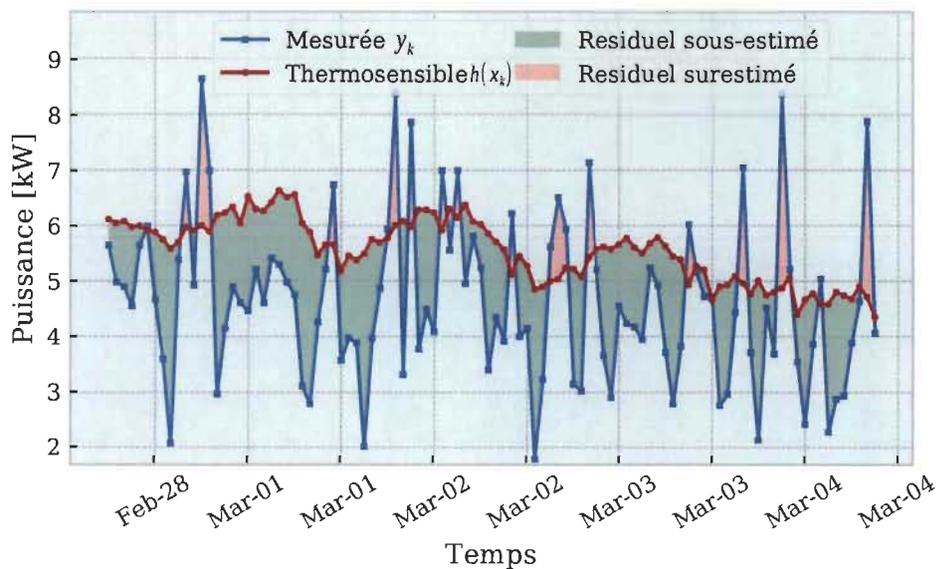


Figure. 4-2 Représentation d'une section de la puissance d'électricité totale, et la composante résiduelle connexe en hiver.

La connaissance des propriétés des résiduels  $r_k^+$  et  $r_k^-$  dans une résidence en fonction du temps d'utilisation d'énergie journalière/hebdomadaire est primordiale pendant le processus de modélisation. Pour cela, nous avons proposé d'utiliser une approche statistique non-paramétrique basée sur l'estimation de l'Espérance Conditionnelle Adaptative (ECA). Cette dernière, ne possède aucune donnée a priori sur la direction de la tendance-cycle à court terme liée résiduels  $r_k^+$  et  $r_k^-$ . Par contre, elle envisage de donner une meilleure approximation de la composante  $r_k$  à partir de l'information fournie par le temps calendaire défini par  $t_k$ .

La structure proposée dans la phase 2 de la Figure. 4-1 constitue la base de ce chapitre. Il s'agit de présenter le lien entre la modélisation et l'explication de la composante thermosensible et le résiduel. Ceci dans le but de voir comment garder la cohérence des approches proposées dans les étapes de prévision afin de les intégrer dans le processus de prévision de la demande d'électricité globale.

Nous commençons tout d'abord par la définition de l'espérance de résiduel  $r_k$  conditionnée par le temps  $t_k$  selon (4-5) :

$$g(t_k) = \mathbb{E}(r|t_k) = \sum_r r f_k(r|t_k) \quad (4-5)$$

Pour chaque composante résiduelle  $r_k^+$  et  $r_k^-$ , nous estimons la distribution de l'espérance conditionnelle  $\mathbb{E}(r^+|t_k)$  et  $\mathbb{E}(r^-|t_k)$  en fonction du temps d'utilisation qui le caractérisent. Nous utilisons ensuite ces distributions pour modéliser la nouvelle composante  $r$ .

### 4.3 Méthodologie de modélisation de la composante résiduelle

Présentement, aucune proposition d'étude d'extraction de la composante tendance-cyclique des résiduels sous-estimé et surestimé n'avait été entreprise dans l'analyse de résiduel. Dans cette étude nous nous basons sur la périodicité des phénomènes non-déterministes constituant le résiduel original  $r_k$ , et nous proposons une approche qui combine la régression non-paramétrique ECA avec la régression paramétrique de type Modèle Linéaire (ML) pour capturer la composante tendance-cyclique à partir des fluctuations de la consommation électrique de résiduel. En fait, l'intérêt de cette approche est d'avoir proposé un moyen d'initier

des améliorations dans l'explication de la partie résiduelle et de fournir une meilleure précision de la prévision de la demande électrique totale.

Dans ce cas, la loi conditionnelle suivie s'inspire de celle utilisée dans l'estimation de la composante thermosensible étudiée au chapitre 3. Pour cela l'espérance conditionnelle  $g(t_k)$  décrite par l'équation (4-5) étant donné le temps d'utilisation journalier/hebdomadaire d'électricité pourrait être calculée en identifiant d'abord la distribution jointe  $f_k(r, t_k)$ , ensuite nous pouvons calculer la distribution marginale du temps calendaire  $f_k(t_k)$  à partir de la fonction de la distribution jointe. Or, l'idée fondamentale de l'approche non-paramétrique ECA utilisée dans ce chapitre est de générer les valeurs d'espérances  $g(t_k)$  en puissance directement sans passer par le calcul des distributions jointe et marginale. Elle est plutôt proposée afin de simplifier la complexité en termes de coût de calcul.

À partir de l'équation (4-5), un estimateur par noyau a été utilisé pour saisir la relation entre les résiduelles  $r_k^+$  et  $r_k^-$  et le temps d'utilisation d'énergie  $t_k$ . Pour cela, un estimateur par noyau mono-variable a été construit, en utilisant l'ensemble de données constitué du temps  $t_k$  comme une variable d'observation du temps calendaire établi pour une période de 24 heures. Cela revient à construire d'abord une fonction périodique du temps d'utilisation d'électricité  $t_k$  définit sur l'ensemble des nombres réels positifs  $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  avec une période  $T$  définit par (4-6).

$$g(t_k) = g(t_k + T) \quad (4-6)$$

Ensuite, pour estimer la distribution circulaire du temps d'utilisation en regardant les résiduels  $r_k^+$  et  $r_k^-$ , l'estimateur par noyau discret basé sur la fonction gaussienne cyclique a été défini par l'équation (4-7). Dans cette expression, la fonction noyau  $\mathcal{K}(t_i)$  utilisée est centrée sur  $t_i$  avec  $t_i = \frac{T}{2}$ , et  $i = \{0, 1, \dots, 23\}$ . Ce point est considéré comme le point de départ de l'estimateur qui représente le milieu de la période du temps, c'est-à-dire à midi ; pour (i) assurer la symétrie dans la distribution du noyau et (ii) assurer la continuité entre tous les points d'observation de l'horloge [126]–[128] présentée dans la Figure. 4-3.

$$\mathcal{K}(t_i) = \exp\left(-\frac{1}{2} \frac{(t_i - t_k)^2}{v^2}\right) \quad (4-7)$$

Où  $v$  est la fenêtre de lissage de la fonction noyau, qui a été fixée à 1 heure pour les résiduels

$r_k^+$  et  $r_k^-$  au bout d'un intervalle de temps régulier [ 1 h, 24 h]. Pour ce cas d'étude, l'estimateur

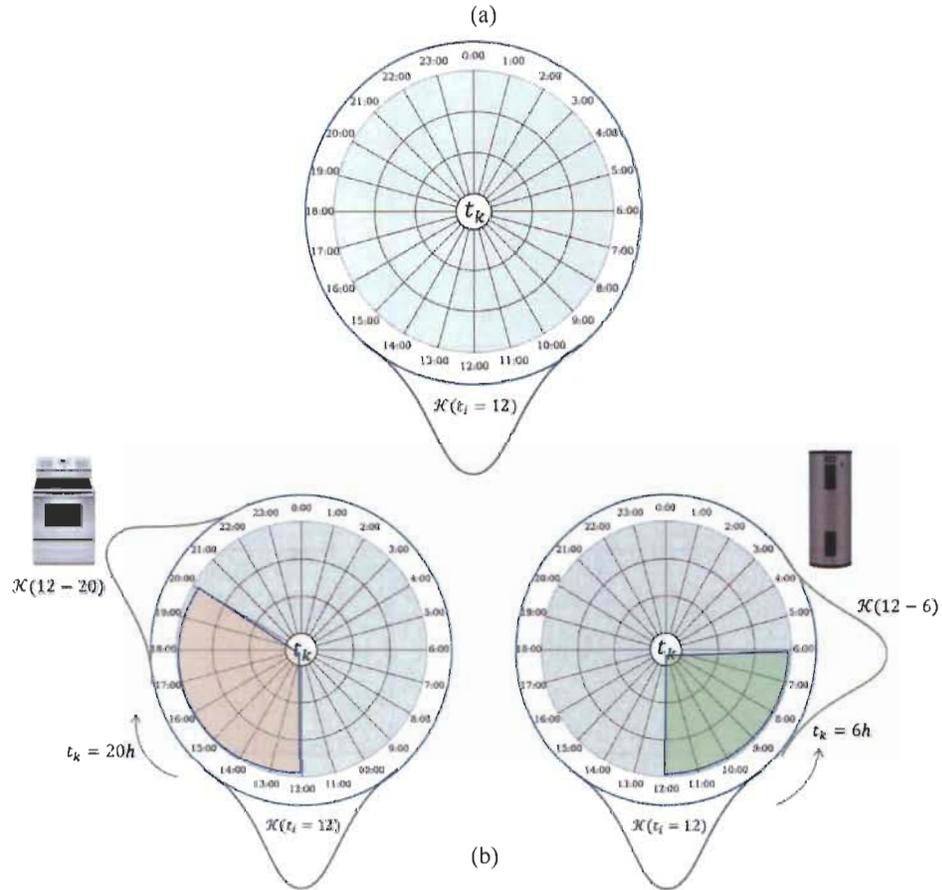


Figure. 4-3 (a) : Estimateur par noyau centré en point  $t_i$ , (b) Déplacement circulaire de l'estimateur par noyau vers l'événement d'utilisation d'électricité.

par noyau est estimé seulement une fois sur le temps  $t_i$ , ensuite, pour chaque événement d'utilisation d'électricité lié aux résiduels  $r_k^+$  ou  $r_k^-$  qui apparaît dans le temps  $t_k$ , le noyau va se modifier par un déplacement de  $\mathcal{K}(t_i - t_k)$  vers le moment d'apparition de l'événement d'utilisation d'électricité, afin d'estimer la distribution de cet événement. Cette opération est réalisée par le déplacement circulaire de l'estimateur par noyau. Une fois que l'estimateur arrivé au moment de l'événement de la demande d'électricité, nous estimons les espérances conditionnelles  $g_k(t_k)$  de manière adaptative, en utilisant la formulation (4-8). Ce mécanisme est adaptatif, il se répète à chaque apparition d'un événement pour exprimer la fréquence de l'utilisation des appareils électriques. Dans ce cas, l'entraînement de la formulation (4-8) requiert les données du passé et du présent des résidus sous et surestimé et le temps calendaire.

L'entraînement considère aussi le noyau  $\mathcal{K}(t_i - t_k)$  pour mettre à jour tous les points de la grille et pas seulement un seul point. Ce noyau va donner donc la plus grande densité à l'événement de résiduels sous ou surestimé qui apparaît à l'instant  $t_k$ , et une faible densité aux événements de voisinage de  $t_k$ . Toutefois, il faut noter que les données utilisées dans l'apprentissage du modèle n'ont jamais participées à sa validation.

$$g_k(t_k) = g_{k-1}(t_k) + \alpha (r_k - g_{k-1}(t_k)) \mathcal{K}(t_i - t_k) \quad (4-8)$$

Où  $0 < \alpha < 1$ , mesure la vitesse d'apprentissage et assure la convergence du modèle [129], [130] [131]. Le facteur d'apprentissage doit être autour de  $\frac{1}{N_r+1}$  [111]. Donc pour 100 points dans la grille, le facteur est fixé à 0.01 pour les valeurs de l'espérance conditionnelles des résiduels connexes à  $r^+|t_k$  et  $r^-|t_k$ .

Par exemple, pour un événement lié à l'utilisation du four électrique à l'instant  $t_k = 20$  h représenté en (b) dans la Figure. 4-3, nous déplaçons le noyau ou le kernel à ce moment et nous estimons les espérances conditionnelles  $g_k(t_k)$  de manière adaptative pour corriger les erreurs passées. Ce concept s'appuie sur l'idée de la correction de la valeur future de l'espérance conditionnelle à partir de sa valeur présente et de l'erreur d'espérance faite à l'instant passé  $g_{k-1}(t_k)$  [132], en tenant compte de l'écart de déplacement effectué par l'estimateur par noyau.

Nous appliquons le mécanisme de l'approche ECA sur les résiduels sous-estimés et surestimés qui proviennent des mesures d'une année d'une puissance électrique totale d'une maison uni-familiale située à Montréal, et nous estimons les espérances conditionnelles journalière et hebdomadaire présentées dans les Figures 4.4 et 4.5, respectivement.

Les Figures. 4-4 (a) et (b) respectivement, montrent le profil hebdomadaire des espérances conditionnelles  $r^+|t_k$  et  $r^-|t_k$  affichées sur toutes les périodes de l'année. Les valeurs d'espérances conditionnelles de  $r^-|t_k$  sont représentées en valeur absolue. Les régions en rouge-brun illustrent la fin de l'apprentissage adaptative de l'approche ECA, durant lesquelles la demande espérée est élevée sur toute l'année. Toutefois, la puissance d'électricité espérée pour le résiduel sous-estimé pendant la fin de semaine est plus forte que les jours de semaine en raison de l'utilisation plus susceptible des appareils énergivores en hiver et en été. Avec les valeurs espérées de résiduel surestimé, le modèle proposé montre qu'il a bien capté de

manière globale l'aspect comportemental et les variations dues à la masse thermique de la maison étudiée. Ces variations permettent de savoir si la maison va tout de suite réagir aux changements de la température ou elle va prendre un délai qui explique qu'elle a une bonne capacité thermique de stockage de chaleur. La Figure. 4-4 (c) représente le profil d'utilisation

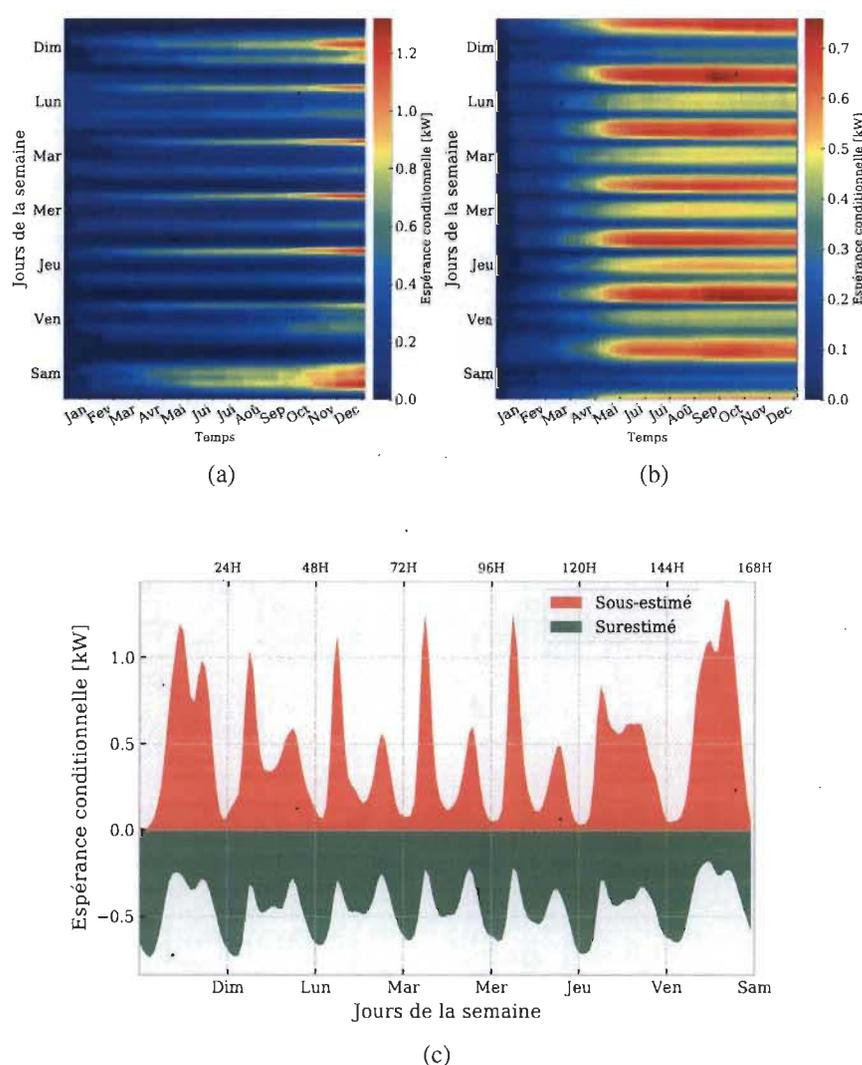
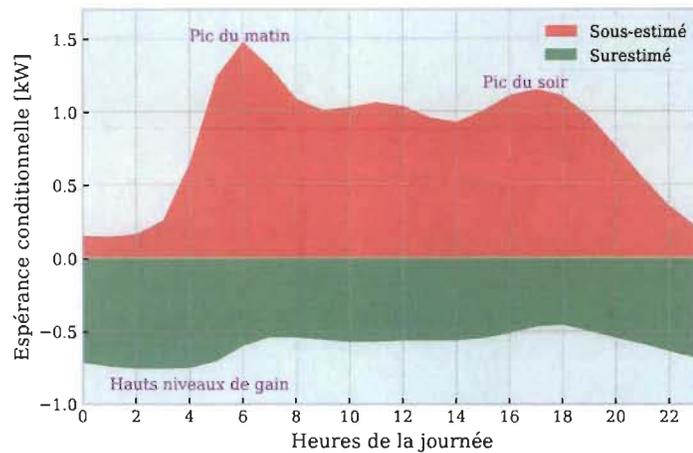


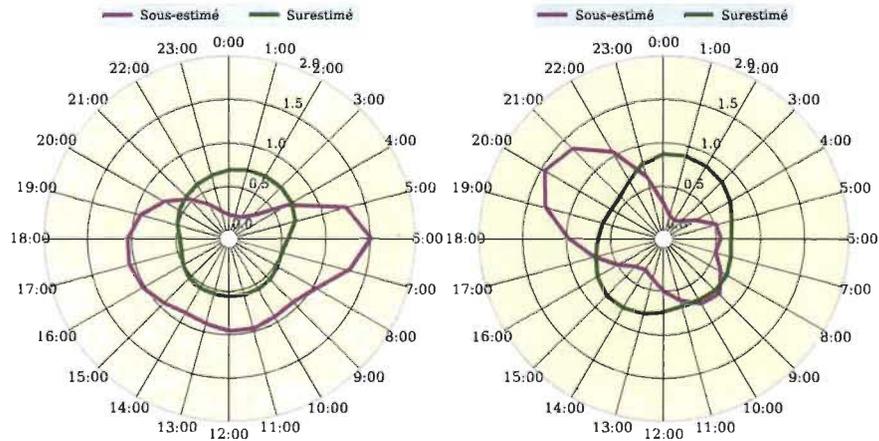
Figure. 4-4 Distribution journalière/hebdomadaire de l'espérance conditionnelle au cours de l'année du dimanche au samedi, (a) résiduel sous-estimé, (b) résiduel surestimé en valeur absolue et (c) profil de la dernière semaine de décembre.

périodique d'électricité journalière et hebdomadaire probable observée à la fois sur les résiduels  $r_k^+$  et  $r_k^-$  pour une semaine au mois de décembre. Elle présente les périodes de pointes journalières sur les résiduels  $r_k^+$  et  $r_k^-$ . Pour le résiduel sous-estimé, la puissance d'électricité

espérée atteint son pic deux fois par jour en matinée et en soirée, et cela lorsque la plupart des occupants se lèvent, prennent une douche, font du café, etc., et le soir, ils allument les lumières, préparent le dîner, etc. De l'autre côté, l'espérance conditionnelle de résiduel surestimé, atteint son pic dans la nuit, ceci peut s'interpréter par davantage d'occupants diminuent le réglage de leurs thermostats à ce moment de la journée.



(a)



(b)

(c)

Figure. 4-5 Distribution horaire de l'espérance conditionnelle des résiduels sous-estimé et surestimé : (b) 15 décembre, et (c) 30 juin.

La Figure. 4-5 (a), représente les distributions horaires des  $r^+$  et  $r^-$  conditionnée par le temps d'utilisation horaire d'électricité réalisée par l'approche ECA sur des mesures d'une année. La figure. 4-5 (a) affiche une journée d'une période hivernale (c.-à-d. le 15 décembre)

qui démontre une puissance élevée pendant la matinée et le soir et ainsi des niveaux de gains qui sont attribués aux différents phénomènes tels que la réduction de la consigne de température et le gain de chaleur généré par les occupants. D'autre part, l'espérance conditionnelle de résiduel sous-estimé contient deux périodes de pointe au cours du matin et le soir, entraînant une demande de pointe plus élevée que les espérances de demande dans le surestimé.

En conséquence les Figures 4.5 (b) et (c) montrent les histogrammes circulaires du comportement périodique horaire pour le 15 décembre et le 30 juin. Les distributions indiquent que le temps d'utilisation d'électricité qui est seulement présent dans le cas de résiduel  $r_k^+$  et dans l'autre il s'agit des gains de chaleur connexe au résiduel  $r_k^-$ . On peut également constater que, le résiduel surestimé  $r_k^-$  est susceptible d'avoir une distribution uniforme en hiver en raison de l'habitude des occupants de régler la température de points de consigne la nuit pendant toute la saison froide. Cette distribution change de forme à cause des gains solaires qui font diminuer la demande d'électricité.

À partir de cette analyse, il est plus facile de fournir des calendriers pratiques de temps de d'utilisation d'énergie et les espérances de puissance demandée dans les résiduels  $r_k^+$  et  $r_k^-$ , en donnant des indications réalistes sur les périodes de sommeil des occupants, les pointes typiques et occasionnelles et ainsi que les activités des occupants dans les heures de pointe et creuses journalières.

#### 4.3.1 *Modèle linéaire*

Afin de répondre à l'objectif de la prévision de la composante résiduelle, une reconstruction de cette composante est nécessaire et ceci à parti des coefficients des espérances conditionnelles  $\mathbb{E}(r^+|d_k)$  et  $\mathbb{E}(r^-|d_k)$  estimées avec l'approche ECA. Il existe plusieurs façons permettant de reconstruire le résiduel à savoir les modèles de série temporelle ou les réseaux de neurones, entre autres. Dans cette étude, le modèle linéaire (ML) a été choisi en fonction de ses propriétés vis-à-vis de sa capacité d'adaptation et de son apprentissage rapide qui lui permet de converger rapidement vers la bonne solution. Ce modèle a été implémenté conjointement avec l'approche ECA pour former une approche mixte notée ECA/ML.

Pour ce faire, le ML a été conçu en utilisant les espérances conditionnelles estimées  $\mathbb{E}(r^+|d_k)$  et  $\mathbb{E}(r^-|d_k)$  reçues de l'approche ECA, comme une séquence d'entrée du ML. Dans

ce cas, une technique de filtrage adaptative basée sur le filtre RLS a été utilisée pour identifier les paramètres du modèle et calculer la sortie  $\hat{r}_k$  [121], [133].

L'entrée du filtre RLS est donnée par (4-9) et la sortie du filtre est défini par l'équation (4-10) :

$$\mathbf{u}_k = [\mathbb{E}(r^+ | d_k), \mathbb{E}(r^- | d_k)]^T \quad (4-9)$$

$$\hat{r}_k = \mathbf{w}_k \mathbf{u}_k \quad (4-10)$$

Où  $\mathbf{u}_k$  est le vecteur d'entrée des coefficients du filtre connexes aux espérances conditionnelles  $\mathbb{E}(r^+ | d_k)$  et  $\mathbb{E}(r^- | d_k)$ , représentées sous la forme  $\mathbf{w}_k = [w_{1k}, w_{2k}]^T$ .

La mise à jour des coefficients peut se faire à partir de l'erreur  $e_k$  calculée entre la sortie du filtre  $\hat{r}_k$  et le signal désiré  $r_k$ , elle utilise un gain  $g_k$  en fonction de l'itération instantanée :

$$\mathbf{w}_{k+1} = \mathbf{w}_k + g_k e_k \quad (4-11)$$

Avec la formule de  $g_k$  donnée par :

$$g_k = \frac{\mu^{-1} R_{k-1} \mathbf{u}_k^T}{1 + \mu^{-1} \mathbf{u}_k R_{k-1} \mathbf{u}_k^T} \quad (4-12)$$

Où  $\mu$  est le facteur d'oubli du modèle de 0.98 et  $R_k$  est la matrice de covariance donnée par :

$$R_k = \mu^{-1} [R_{k-1} - g_k \mathbf{u}_k R_{k-1}] \quad (4-13)$$

Le diagramme présenté dans la Figure. 4-6 résume l'approche mixte ECA/ML appliquée à la modélisation et la prévision de la composante résiduelle. L'apprentissage se fait à chaque pas de temps  $k$  sur une période d'une année, durant laquelle le modèle utilise les données d'entrées des résiduels  $r_k^+$  et  $r_k^-$  et le temps d'utilisation d'électricité  $t_k$  à l'instant  $k$  pour estimer les espérances conditionnelles. Ces dernières seront utilisées dans le modèle ML pour la mise à jour des coefficients du filtre RLS. Dans la phase de prévision, le modèle utilise l'information future de temps d'utilisation  $t_{k+\tau}$  afin de prédire la composante résiduelle  $\hat{r}_{k+\tau}$  sur un horizon  $\tau$ .

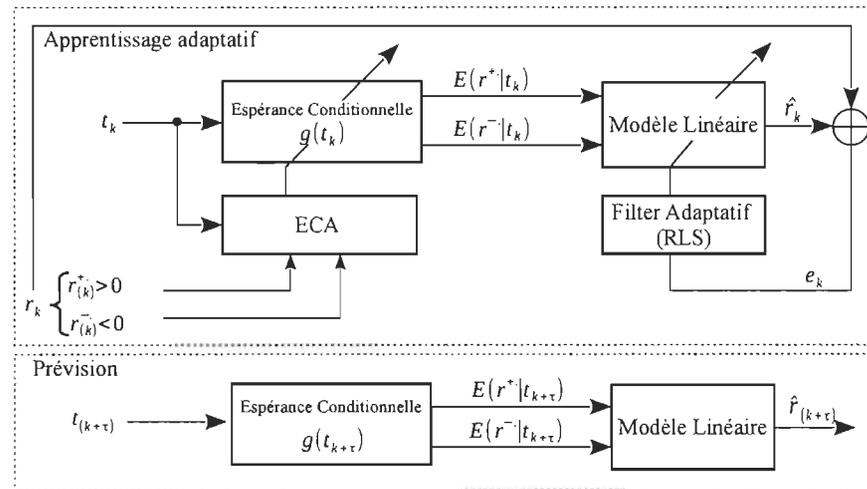


Figure. 4-6 Schéma bloc de la modélisation et prévision de la composante résiduelle.

#### 4.4 Cadre comparatif

La première application utilisée pour évaluer les performances de l'approche ECA/ML est la prévision de la composante résiduelle  $r_k$  sur un horizon d'un jour, en faisant une comparaison avec les modèles autorégressifs AR et ARX, couramment utilisés dans le traitement de résiduel. L'application des deux modèles sur la composante  $r_k$  donnée par l'équation (4-1) doit être adaptative. Ainsi, l'utilisation de cette composante ne prend pas en compte le processus de division en résiduels surestimé et sous-estimé. D'abord, nous considérons le modèle AR avec la sortie définie par :

$$r_{k+\tau} = \sum_{j=1}^p \theta_j r_{k+\tau-j} + \varepsilon_{k+\tau} \quad (4-14)$$

Où est  $\varepsilon_k$  est l'erreur de prévision, et  $\Theta_k = [\theta_{1k}, \theta_{2k}, \dots, \theta_{pk}]^T$  est le vecteur des coefficients. Le résiduel  $r$  estimé est calculé en fonction des données  $r_k$  et du vecteur des poids  $z_k$ .

$$z_k = [r_{k-1}, r_{k-2}, \dots, r_{k-p}]^T \quad (4-15)$$

Dans ce cas, l'entrée du modèle a été fixée à six données de résiduel  $r_k$  et les paramètres du modèle ont été identifiés de manière adaptative en utilisant le filtre RLS présenté dans [121].

L'équation de mise à jour des paramètres est donnée comme suit :

$$\Theta_{k+1} = \Theta_k + K_k \xi_k \quad (4-16)$$

Où  $K$  est le vecteur de gain, défini par l'équation suivante :

$$K_k = \frac{\rho^{-1} X_{k-1} z_k^T}{1 + \rho^{-1} z_k X_{k-1} z_k^T} \quad (4-17)$$

$X$  est la matrice de covariance du signal d'entrée, donnée par l'équation (4.18), avec un facteur d'oubli  $\rho$  sélectionné à 0.98 :

$$X_k = \rho^{-1} [X_{k-1} - K_k z_k X_{k-1}] \quad (4-18)$$

La deuxième méthode utilisée dans la comparaison est le modèle ARX direct mis en œuvre selon [134], [135]. Ce modèle est utilisé pour saisir la relation entre le résiduel  $r_k$  et le temps d'utilisation d'électricité  $t_k$  comme une variable explicative. La sortie est décrite par l'équation (4-19) :

$$r_{k+\tau} = \sum_{j=1}^p \theta_j r_{k+\tau-j} + \sum_{i=1}^b \varphi_i d_{k+\tau} + c_{k+\tau} \quad (4-19)$$

Où  $t_k$  et  $r_k$  sont les séquences d'entrées et de sortie, et  $\varphi_i$  représente les coefficients de temps d'utilisation, et  $c$  représente l'erreur. Dans ce modèle, le nombre de données d'entrée de résiduels est supposé  $p = 6$ . Les coefficients du modèle ont été estimés en utilisant l'algorithme adaptatif RLS présenté dans [121], afin de produire la sortie  $r_{k+\tau}$ , en appliquant un facteur d'oubli  $\eta = 0.92$ .

En terme de précision l'NMAE, MAE et R-Square données par 4-20, 4-21 et 4-22 ont été utilisées comme indicateurs pour faire référence à la qualité de la prévision.

$$NMAE = \frac{\sum_{k=1}^N |q_k - \hat{q}_k|}{\sum_{k=1}^N |q_k|} \quad (4-20)$$

$$MAE = \frac{1}{N} \sum_{k=1}^N |q_k - \hat{q}_k| \quad (4-21)$$

$$R^2 = \frac{\sum_{k=1}^N (\hat{q}_k - \bar{q}_k)^2}{\sum_{k=1}^N (q_k - \bar{q}_k)^2} \quad (4-22)$$

Où  $q_k$  représente  $r_k$  dans le cas de la prévision de la composante résiduelle et  $y_k$  pour le cas de la prévision de la puissance totale électricité. La section suivante permet de comparer et discuter les performances des trois modèles utilisés dans la prévision de la composante  $r_k$ .

#### 4.4.1 Résultats et Discussion

##### 4.4.1.1 Résultats de prévision de la composante résiduelle

La comparaison des résultats de prévision sur un horizon de 24 heures avec les méthodes ECA/ML, les modèles AR et ARX sont présentés dans la Figure. 4-7. Les résultats obtenus montrent la capacité de l'approche ECA/ML à extraire la part cycle-tendance de caractère non-linéaire de la composante résiduelle par rapport aux autres modèles sélectionnés.

Nous observons également que, au cours de l'année d'analyse, l'approche ECA/LM peut fournir parfois une bonne qualité de prévision de résiduelle en suivant de manière réaliste les fluctuations de la puissance d'électricité. L'approche montre que l'accord entre le résiduel original et la prévision est parfois excellent sur les pics, particulièrement pour les saisons chaudes où la consommation dépend surtout du comportement stochastique des occupants. Puisque cette demande notable peut avoir un comportement périodique et compte pour une part importante de la composante résiduelle, elle peut s'expliquer en appliquant l'approche ECA/LM.

L'exercice de prévision de la composante  $r_k$  est répété cette fois sans le processus de division en considérant plutôt les modèles adaptatifs AR et ARX. Dans ce cas, les deux modèles échouent dans la capture des changements de résiduels par rapport à l'approche ECA/ML. Ainsi, leurs capacités de s'adapter à la tendance du résiduel et de saisir la forme non-linéaire des variations dynamiques de la composante résiduelle n'est pas adéquate. Les non-linéarités périodiques introduites dans la composante résiduelle rendent moins efficaces l'utilisation des modèles AR et ARX. Ces non-linéarités sont principalement dues aux comportements des charges électriques qui sont plus perceptibles dans le résiduel sous-estimé  $r_k^+$ . En effet, notre approche ECA/ML excelle dans l'amélioration de la prévision du résiduel par rapport aux

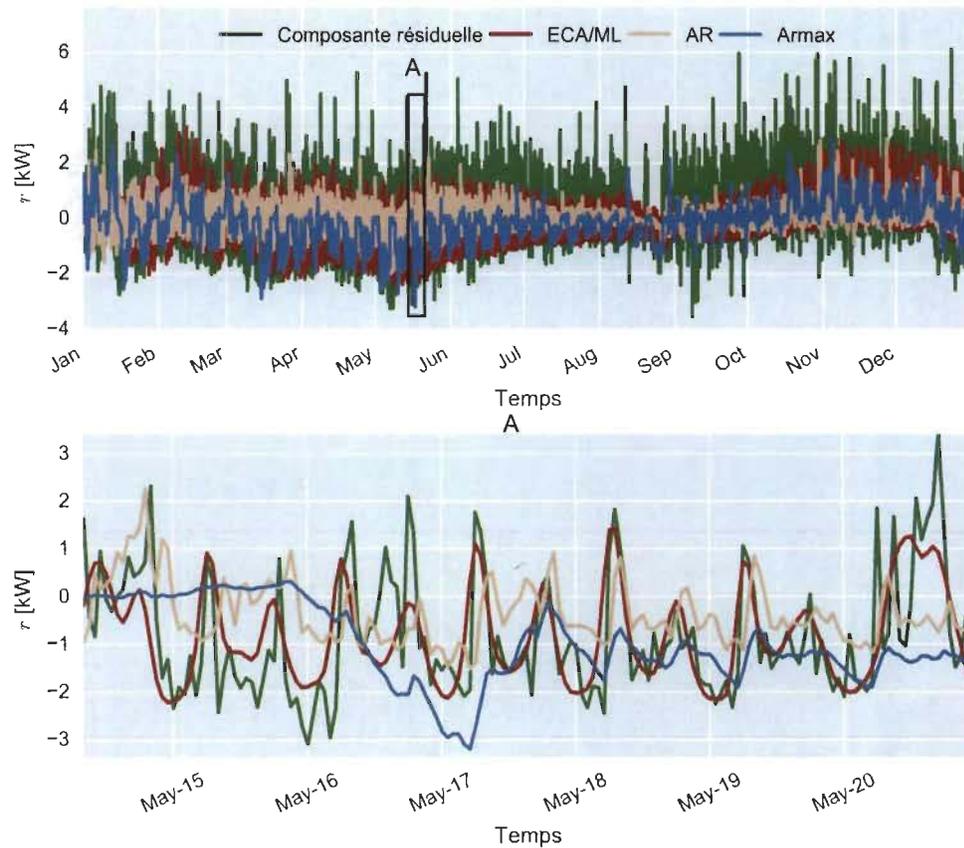


Figure. 4-7 Performance des modèles adaptatifs dans la modélisation et la prévision de la composante résiduelle sur un horizon de  $\tau = 24$  heures. AR ( $\rho = 0.98$ ), ARX ( $\lambda = 0.92$ )

autres modèles en raison de sa capacité à fournir une distribution d'espérance conditionnelle réaliste sur la prévision de la composante résiduelle. D'autre part, l'inconvénient des modèles AR et ARX peut-être attribué au fait qu'ils sont des méthodes paramétriques, qui requièrent une définition préalable de la bonne structure mathématique qui converge bien avec la sortie du modèle. Cette structure suppose la connaissance a priori du bon ordre des coefficients de la série.

Les courbes présentées sur la Figure. 4-8 permettent d'observer le résumé des performances de précision de la prévision de la composante  $r_k$  obtenue pour différents horizons. Au regard des valeurs de NMAE, on remarque très rapidement que l'approche ACE/ML est plus performante que les autres modèles, car quelque soit l'horizon de prévision considéré, le NMAE est toujours inférieur à 0.78. Seule l'approche ACE/ML est capable de rivaliser avec les AR et ARX pour la prévision de la composante résiduelle. De manière générale, le modèle

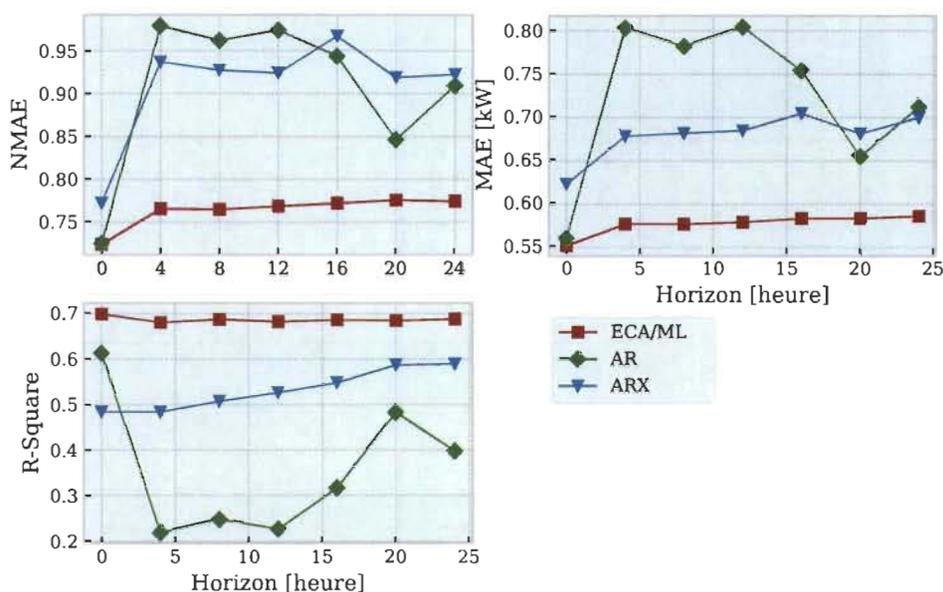


Figure. 4-8 Comparaison de la précision de la prévision de la composante résiduelle en fonction d'horizon de prévision.

proposé fournit un NAME stable comme il est capable d'expliquer environ 23% du résiduel sur tous les horizons.

#### 4.4.1.2 Résultats de prévision de la demande d'électricité

Afin de construire un modèle global de prévision de la demande d'électricité, les résultats de prévisions des deux phases liées à la composante thermosensible et la composante résiduelle, respectivement, sont ajoutés à chaque horizon  $\tau$  tel que montré à la Figure. 4-1, en utilisant l'équation suivante :

$$\hat{y}_{k+\tau} = \hat{h}(x_{k+\tau}) + \hat{r}_{k+\tau} \quad (4-23)$$

Où  $\hat{y}_{k+\tau}$  est la prévision de la puissance totale, et  $\hat{h}(x_{k+\tau})$  est la prévision de la composante thermosensible obtenue à partir de l'approche EADC. En outre, les résultats de prévision de cette composante sont ajoutés aux prévisions des résiduels réalisés avec les modèles AR et ARX. La Figure. 4-9 présente une comparaison de la puissance d'électricité totale prédite pour chacun des modèles EADC, EADC+ECA/ML, EADC+AR et EADC+ARX, sur un horizon de 24 heures.

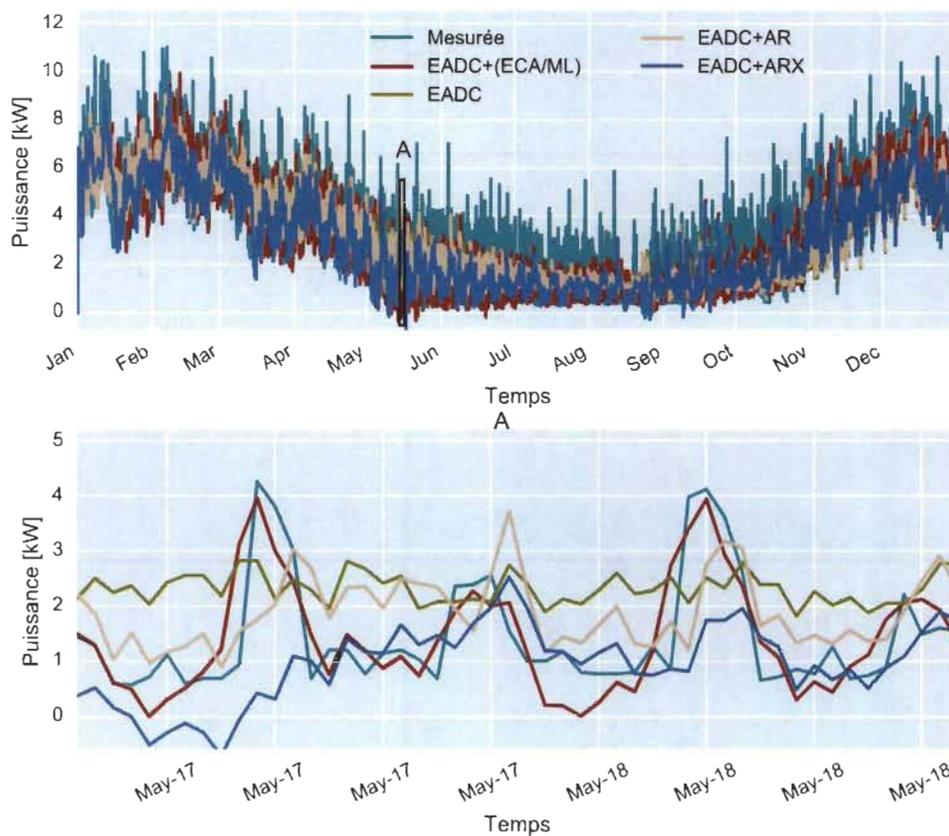


Figure. 4-9 Comparer la demande d'électricité mesurée et prédite par les trois modèles sur l'horizon  $\tau=24$  heures

Une analyse plus détaillée de la Figure. 4-9 permet de constater que le modèle EADC tout seul éprouve particulièrement beaucoup de difficulté à suivre les pics de la consommation réelle. Cette difficulté a été expliquée dans le chapitre 3. Seul le EADC+ECA/ML est capable de montrer une capacité à prévoir la puissance liée aux activités des occupants, l'utilisation de l'électricité, ainsi que l'ensemble des gains et des pertes, mais de manière agrégé. Les modèles EADC+AR et EADC+ARX sont inopérants pour suivre les changements de la puissance totale et identifier correctement la non-linéarité de la demande d'électricité.

Nous voyons ici que malgré le temps d'utilisation de l'énergie considérée dans le modèle ARX, les modèles EADC+ARX fournissent des résultats qui sont différents de ceux obtenus avec le EADC+ECA/ML.

Les courbes des erreurs sont réalisées pour divers horizons sur une journée telles que montrés à la Figure. 4-10. Les résultats de la précision obtenus dans la prévision de la

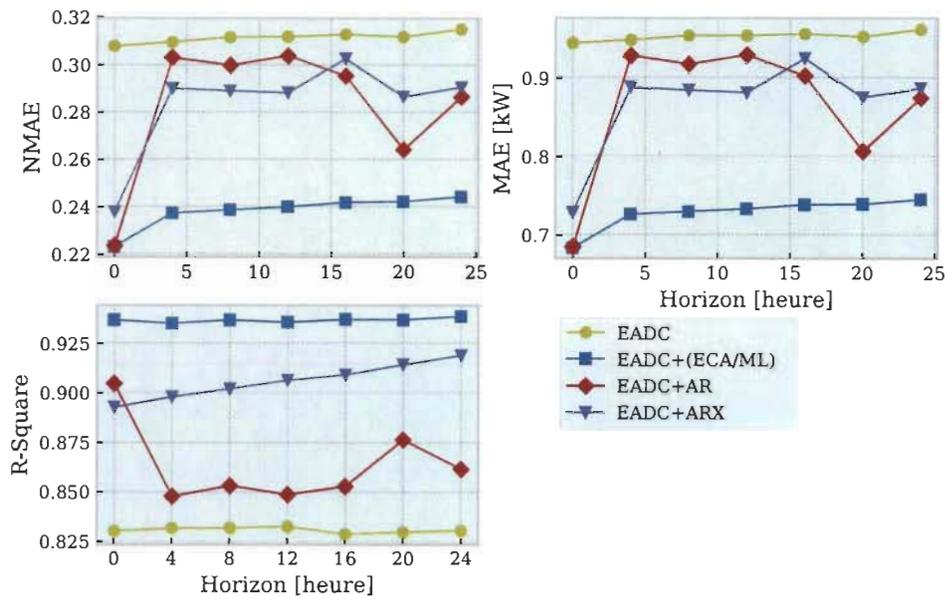


Figure. 4-10 Comparaison de la précision de la prévision de la demande d'électricité réalisée avec les trois modèles en fonction d'horizon de prévision.

demande d'électricité totale vont dans le même sens que ceux obtenus dans la prévision de la composante  $r_k$ , et démontrent une fois de plus la performance des modèles EADC+ECA/ML qui fournissent un bon résultat. D'autre part, l'erreur en NMAE produite de la combinaison des modèles EADC et ECA/ML est non seulement petite et demeure stable sur tous les horizons. En terme de précision, elle représente 75% de la prévision de la puissance totale pour l'horizon  $\tau=24$  heures. Alors que l'erreur en termes de MAE en (kW) avec l'emploi du modèle EADC tout seul a diminuée de 0.991 kW à 0.754 kW pour l'horizon  $\tau=24$  heures et ceci par rapport aux modèles couplés EADC+ECA/ML. Tandis que, le coefficient de détermination  $R^2$  signifie que la combinaison des modèles EADC+ECA/ML est capable de déterminer 92.7 % de la distribution des points de la prévision sur la droite de régression sur l'ensemble des horizons.

La Figure. 4-11 présente le taux d'amélioration de la précision de la prévision de la puissance d'électricité totale réalisée avec EADC+ECA/ML et comparée avec les taux obtenus avec les modèles EADC+AR, et EADC+ARX. Ce taux est calculé par l'écart entre deux valeurs de précision de la prévision. La première est  $(1 - NMAE) \times 100$  qui distingue l'écart entre l'unité et l'erreur de la prévision de la puissance totale en termes de NMAE. La deuxième représente la précision de la prévision de la composante thermosensible estimée dans le

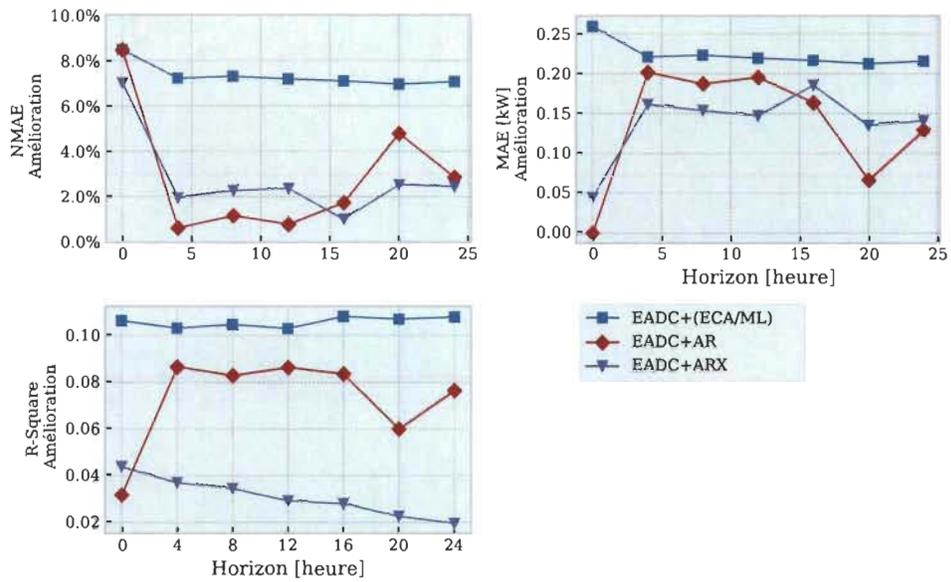


Figure. 4-11 Comparaison de l'amélioration de la prévision entre les trois modèles

chapitre 3 qui correspond à 68 %, en utilisant la relation  $(1 - NMAE) \times 100 = 68\%$ .

Sur la figure, on observe une amélioration significative sur tous les horizons d'environ  $\sim 7\%$  obtenue avec le couplage des résultats de prévision de la composante thermosensible et la composante résiduelle EADC+ECA/ML. Par contre les prévisions faites avec les modèles EADC+AR et EADC+ARX sont capables d'améliorer seulement 3% et 2,3% de la précision, respectivement, à l'horizon de 24 heures. Ces résultats montrent que l'approche proposée donne les meilleures précisions de la prévision, avec le plus important taux d'amélioration pour le cas de la maison étudiée. Il faut savoir que ce taux d'amélioration change d'une maison à une autre, et sur l'ensemble des cinq maisons testées et validées, l'approche proposée a donné des résultats similaires comparativement aux méthodes classiques.

#### 4.4.1.3 Explication du second résidu

Après l'estimation de la partie périodique de la composante résiduelle, nous sommes intéressés à voir le comportement de la partie restante appelée le second résidu, définie par :

$$r_{res,k} = |r_k| - |\hat{r}_k| \quad (4-24)$$

Dans ce cas, le même exercice est répété en considérant plutôt la division de  $r_{res,k}$  en deuxième résiduel sous-estimé et surestimé, comme indiqué par :

$$r_{res,k}^+ = \begin{cases} r_{res,k}, & \text{si } r_k > \hat{r}_k \\ 0, & \text{sinon} \end{cases} \quad (4-25)$$

$$r_{res,k}^- = \begin{cases} r_{res,k}, & \text{si } r_k < \hat{r}_k \\ 0, & \text{sinon} \end{cases} \quad (4-26)$$

Les deux résiduels  $r_{res,k}^+$  et  $r_{res,k}^-$  sont illustrés dans la Figure. 4-12. Leurs explications peuvent aider à mieux comprendre la demande d'électricité totale donnée par :

$$y_k = h(x_k) + r_k + r_{res,k} \quad (4-27)$$

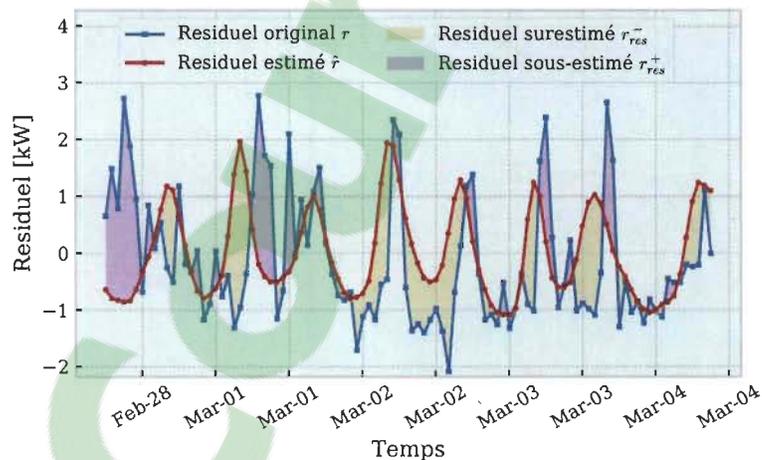


Figure. 4-12 Représentation d'une section de la composante résiduelle, et le résiduel  $r_{res,k}$  en hiver.

Par ailleurs, afin de réaliser une explication du second résidu  $r_{res,k}$ , nous avons suggéré d'exploiter une des sources de données que nous possédons. Cette source consiste à utiliser des informations générées par les charges électriques à savoir les mesures de chauffe-eau électrique  $y_{ewh}$  et les autres charges  $y_{el}$ , fournies par le LTE.

Par contre, avant de réaliser cette analyse, nous proposons d'étudier l'autocorrélation des sous-composantes  $r_{res,k}^+$  et  $r_{res,k}^-$  à partir de l'historique de leurs observations que nous disposons. Le but de l'autocorrélation ici est de mesurer la ressemblance entre les séries temporelles du second résidus et quantifie leur association interne. Le tracé de l'autocorrélation

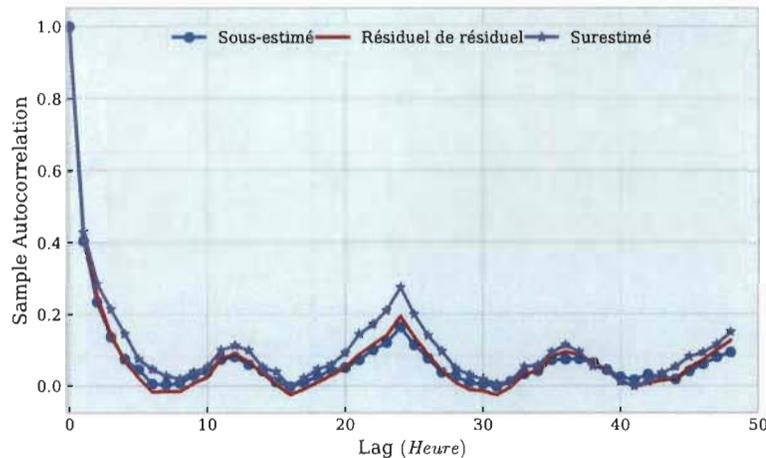


Figure. 4-13 Autocorrélation de résiduel du résiduel sous-estimé et surestimé .

est illustré sur la Figure. 4-13. Il montre les autocorrélations de chaque résiduel retardé par un délai noté  $lag$ . Ce dernier définit le paramètre de décalage temporel choisi arbitrairement à 48 heures.

Cette analyse a montré qu'il a une forte ressemblance interne dans la série du second résidu surestimé  $r_{res,k}^-$  au délai de  $lag = 24$  heures. Ce qui montre qu'il a une chance de 26% qu'il soit prévisible à ce délai comparé à 16% pour le second résidus sous-estimé.

Nous avons par la suite vérifié l'autocorrélation de chacune des séries temporelles de ces résiduels et les charges de chauffe-eau, et les autres charges comme montré dans la Figure. 4-14. Le but derrière cette analyse est de révéler les relations existantes entre les second résidus  $r_{res,k}^+$ ,  $r_{res,k}^-$  et les mesures des charges électriques. Pour cela, nous avons trouvé une forte ressemblance de 40% pour le chauffe-eau et 20% pour les autres charges, pour une valeur de décalage significative correspond au  $lag = 24$  heures.

Pour ce  $lag$ , nous avons analysé la corrélation entre les 4 séries temporelles des charges de chauffe-eau  $y_{ewh,k}$ , les autres charges  $y_{el,k}$  et les seconds résidus sous-estimés et surestimés  $r_{res,k}^+$  et  $r_{res,k}^-$ , respectivement.

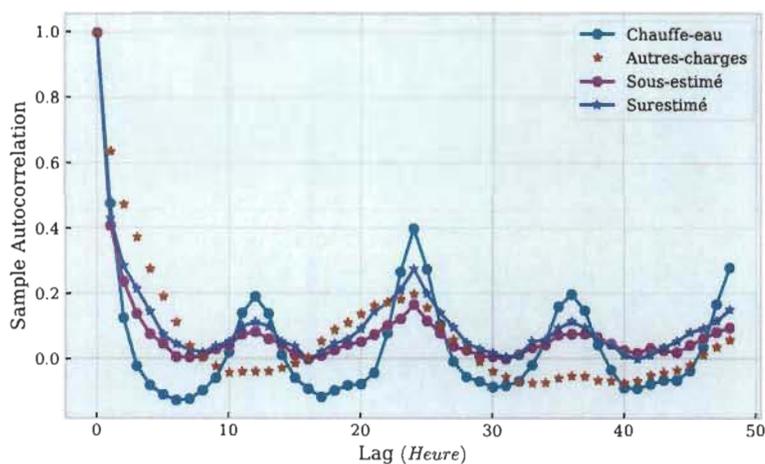


Figure. 4-14 Autocorrélation de résiduel sous-estimé et surestimé et leurs interactions avec les charges de chauffe-eau et les autres charges.

Cette analyse a été présentée par une matrice de corrélation utilisée pour évaluer la dépendance entre les 4 séries temporelles, comme est illustré dans la Figure. 4-15. La matrice

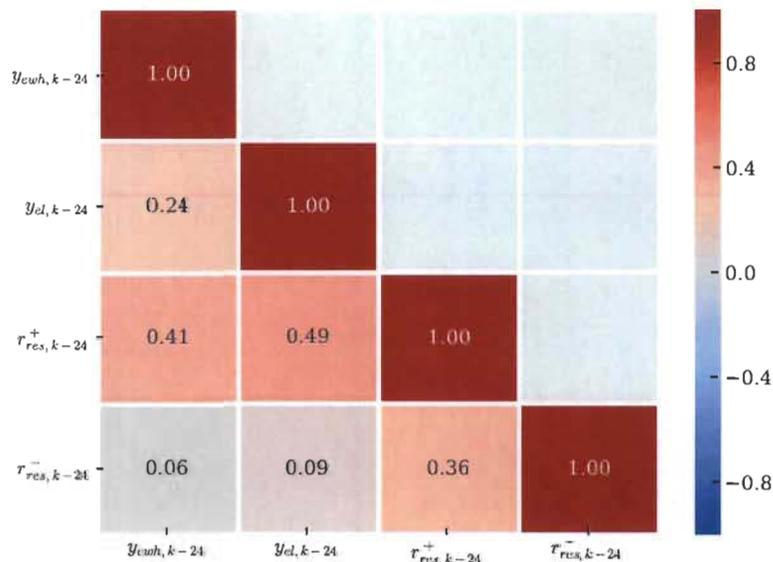


Figure. 4-15 Matrice de corrélation pour évaluer la dépendance entre les charges électriques et les seconds résidus sous-estimé et surestimé.

illustrée contient le test d'autocorrélation pour chaque série en rouge et les autres couleurs montrent la corrélation entre les 4 séries temporelles.

L'observation de la corrélation entre le second résidu sous-estimé et les mesures de chauffe-eau  $\text{corr}(r_{res,k-24}^+; y_{ewh,k-24})$  montre que la présence de cette mesure peut aider à expliquer 41% du second résiduel  $r_{res,k}^+$ . Tandis que, la présence des mesures des autres charges peuvent expliquer 49% du second résidu  $r_{res,k}^+$ .

Par conséquent, les prochaines améliorations du second résidu  $r_{res,k}$  sont envisageables seulement en présence des observations des charges électriques et de chauffe-eau. Ces mesures peuvent être fournies soit par deux voies : (1) la désagrégation de la puissance totale d'électricité totale ou (2) une mesure directe de ces charges en installant des capteurs auprès du client.

#### 4.5 Synthèse du chapitre

L'objectif de ce chapitre était de proposer une stratégie parallèle de modélisation et de prévision de la composante résiduelle en vue de prédire la puissance totale d'électricité. Les caractéristiques périodiques d'une partie de la composante résiduelle font en sorte que ce travail a introduit un nouveau processus d'analyse de résiduel. Tout d'abord, la composante résiduelle a été divisée en résiduels sous-estimé et surestimé pour extraire la périodicité des phénomènes environnementaux et comportementaux.

Cette stratégie nous a menés à la proposition d'un modèle mixte qui combine une approche adaptative de l'espérance conditionnelle (ECA) pour repérer et expliquer séparément ces phénomènes périodiques. Or, la connaissance du comportement des occupants est considérable pour identifier une des sources d'incertitude qui causent l'écart entre la composante thermosensible et les mesures de la demande d'électricité. En outre, dans les modèles basés sur les logiciels de simulation des bâtiments résidentiels comme TRNSYS et EnergyPlus, la présence et les activités des occupants sont modélisées de manière sommaire par des scénarios déterministes. Il est donc possible de remplacer ces scénarios par le calendrier du temps de puissance d'électricité capté avec l'approche proposée, afin de simuler une puissance électricité totale proche de la réalité.

Enfin, une prévision de la composante résiduelle a été réalisée avec le modèle linéaire (ML) en utilisant les espérances conditionnelles estimées avec l'approche ECA. Ceci nous

a permis de présenter une étude comparative de la performance d'ECA/ML par rapport aux autres méthodes adaptatives comme le AR et ARX.

Pour l'ensemble de ces stratégies, un modèle global a été construit en ajoutant les résultats de prévision de la composante thermosensible avec ceux obtenus de la composante résiduelle pour prédire la demande d'électricité totale. Une comparaison des résultats avec les données réelles d'une résidence occupée a montré que la structure d'ensemble composée des approches EADC et ECA/ML proposées permet une amélioration significative de 7% de la précision de la prévision de la demande d'électricité sur un horizon d'un jour.

Finalement, une étude de corrélation a été réalisée pour identifier le lien entre le second résidus et les informations générées par les charges électriques et le chauffe-eau. Cette analyse montre que l'explication et les prévisions de la puissance d'électricité totale peuvent s'améliorer en présence des données désagrégées des charges électriques.

## Chapitre 5 - Conclusions et recommandations

Cette thèse a étudié le problème de la prévision à court terme de la demande électrique résidentielle dans le contexte de GLDD. Dans cette étude, nous avons présenté une approche statistique modélisant la demande d'électricité des bâtiments résidentiels. Celle-ci permet l'anticipation de la consommation sur un horizon de vingt-quatre heures. Les modèles utilisés sont majoritairement basés sur l'analyse des distributions conditionnelles non-paramétrique. La procédure de modélisation considère un minimum de mesures, mais elle exploite de ces données le maximum de connaissance sous une forme nouvelle et automatisée. De façon générale, le développement est axé sur l'identification de la thermosensibilité du bâtiment. Cette thermosensibilité permettra de quantifier la dépendance entre la puissance et l'écart de température intérieure-extérieure. Plusieurs applications pourront être dérivées de cette analyse. Dans l'un, l'utilisateur peut avoir une information plus précise sur la performance thermique de sa maison. Dans l'autre, le modèle peut être utilisé dans un cadre de prévision à partir des données météo. Le travail montre aussi une contribution supplémentaire qui consiste à l'analyse des moments de consommation et qui permet d'améliorer l'algorithme de prévision à partir d'un modèle statistique liant le résiduel et le temps calendaire.

### 5.1 Contributions principales

Les contributions principales de cette thèse sont d'ordre méthodologique.

- Les analyses réalisées visent à construire une technique d'exploration de données en tant qu'étape de pré-traitement pour aider au développement d'un cadre de modélisation permettant d'explicitier une partie de la demande agrégée. À notre connaissance, l'identification de la composante thermosensible avec la technique d'exploration des

données n'a pas fait l'objet de recherches détaillées. Jusqu'à présent, aucune étude expliquant en ligne la consommation thermosensible n'avait été entreprise en employant l'approche non-paramétrique pour l'apprentissage adaptatif. Cette approche utilise la régression non-paramétrique bi-variable pour l'estimation de la densité jointe entre la demande et l'écart de température. À partir de cette densité, une Estimation Adaptative de la Densité Conditionnelle EADC a été proposée. Celle-ci cherche à expliquer la consommation thermosensible au auprès du client. L'étude a permis de démontrer la capacité d'une telle démarche de passer de la modélisation boîte noire au modèle gris en fournissant des paramètres de signification physique. Plus spécifiquement, il s'agissait de la température d'équilibre ainsi que de la constante de déperdition thermique équivalente UA et ceci à partir des caractéristiques de la composante thermosensible. Cette modélisation physico-statistique a permis de tenir compte des incertitudes associées à la prévision de la température sans avoir un grand effet sur la précision de la prévision de la part thermosensible. Une évaluation des performances de l'approche proposée ainsi que des comparaisons avec des méthodes de l'état de l'art ont été réalisées. Ceci rend possible son application à des maisons uni-familiales sans recours à des mesures supplémentaires.

- Le modèle EADC est complété par un deuxième processus d'exploration de données en employant une deuxième approche non-paramétrique spécifique à l'explication de la composante résiduelle. Ceci utilise l'effet calendaire comme un outil qui permet de capter les moments périodiques de l'utilisation de l'électricité et obtenir une compréhension sur les profils cycliques d'appel de puissance pour les phénomènes indépendants de la température extérieure. L'approche consiste à diviser le résiduel en deux sous-composantes associées aux sous et sur estimations. Chacune de ces composantes peuvent être en interaction avec des facteurs météorologiques autres que la température telle que l'activité des occupants et leurs habitudes. Un modèle mixte qui combine une approche d'Espérance Conditionnelle Adaptative(ECA) avec une régression paramétrique ML a été adopté. À partir du résiduel sous-estimé et surestimé, le modèle ECA génère en ligne les moments périodiques d'utilisation d'électricité

de l'ordre journalier et hebdomadaire lié aux phénomènes environnementaux et comportementaux non déterministes. Conjointement, la régression paramétrique ML a été conçue de manière adaptative pour fournir des prévisions de la partie cyclique du résiduel en employant les espérances conditionnelles estimées par l'approche ECA.

- Au final, les modèles EADC et ECA/LM sont articulés entre eux et couplés pour former un modèle global de la puissance totale d'électricité résidentielle. Nous pouvons conclure que l'approche statistique non-paramétrique permet de développer des modèles de prévision de pouvoir explicatifs, auxquels nous pouvons nous référer pour articuler des connaissances avec les données disponibles.

## 5.2 Limites et difficultés

### 5.2.1 *Hypothèse de la température intérieure constante*

La première limite de notre étude concerne la supposition d'une température intérieure constante des maisons sur toute l'année d'analyse. En effet, si nous avions la mesure réelle des températures intérieure, nous pourrions l'intégrer dans le modèle EADC afin d'avoir le bon calcul d'écart de température et aussi expliquer de manière précise la composante thermosensible des maisons.

### 5.2.2 *Hypothèse d'un estimateur multi-dimensionnel*

La formulation du modèle de thermosensibilité pourrait être améliorée, en tenant compte du cas un estimateur multi-dimensionnel de la température extérieure et de certaines variables climatiques qui influencent la demande comme le soleil et le vent.

### 5.2.3 *Fenêtre de lissage de l'estimateur par noyau*

Une des difficultés majeures de l'analyse non-paramétrique adaptative est de trouver la bonne valeur de la fenêtre de lissage au fur et à mesure que les nouvelles données deviennent disponibles. Or, dans le cas d'un estimateur hors ligne, l'estimateur par noyau est fondé sur un grand ensemble de données d'où la fenêtre de lissage sera facile à paramétrer. En effet, le choix de la fenêtre est un facteur déterminant sur les résultats finaux de l'estimateur non

paramétrique. Notamment, une fenêtre trop faible ou trop grande provoque l'apparition ou l'effacement de détails sur la distribution. Dans l'approche proposée, nous avons fait un balayage des valeurs sur l'intervalle  $[0.01, 1]$ . Cependant, la valeur de la fenêtre change en fonction des mesures et des maisons à étudier. Ceci est une limite qui impose de balayer chaque fois l'intervalle pour chercher la valeur de la fenêtre spécifique à chaque maison. Enfin, la manière de trouver la bonne fenêtre de lissage doit être améliorée pour qu'elle soit automatique, notamment si nous voulons estimer la densité d'un ensemble des maisons ou un parc résidentiel.

### 5.3 Recommandations

Plusieurs perspectives d'amélioration de ce travail sont prometteuses :

#### 5.3.1 *Modélisation multi-dimensionnelle*

Chaque composante de la demande d'électricité totale pourrait faire l'objet d'une modélisation plus complexe et spécifique. Nous suggérons que les modèles EADC et ECA/LM pourraient être améliorés en prenant en compte d'autres variables climatiques telles que l'ensoleillement et la vitesse du vent. Nous pourrions alors comparer les résultats à ceux obtenus par notre démarche en utilisant seulement la température extérieure et l'effet calendaire comme seules variables explicatives.

En ce qui concerne la composante thermosensible, il est aussi possible d'utiliser d'autres valeurs des écarts des températures. Pour cela, il est suggéré de choisir différentes températures de consigne moyennes dans les résidences plutôt que  $21^{\circ}\text{C}$  et voir leurs impacts sur la consommation thermosensible et la part résiduelle.

#### 5.3.2 *Amélioration de l'estimateur par noyau*

##### 5.3.2.1 *Choix du noyau*

Bien que le choix de la fonction noyau à utiliser soit considéré comme moins important, il est suggéré de tester les différents types de noyaux cités dans le tableau 2-1, comme il en a été discuté dans le chapitre 2. Il serait aussi intéressant de comparer la performance des

estimateurs pour tout noyau  $\mathcal{K}$  dans des situations différentes afin de voir leurs impacts sur l'estimation obtenue.

Puisque la fenêtre de lissage  $\nu$  joue un rôle très important dans la performance des estimateurs adaptatifs des densités à noyaux associées, un sujet intéressant serait d'étudier les sélecteurs automatiques de la fenêtre, en traitant le problème de choix de la fenêtre optimale avec d'autres méthodes de la littérature comme l'utilisation des critères bayésiens ou les méthodes adaptatives de Lepski modifiée [136].

### 5.3.2.2 *Dimension du noyau*

En plus des noyaux univariés et bivariés, il serait intéressant d'augmenter la dimension de l'estimateur afin d'englober plus de données. En effet, l'augmentation de la dimension permet d'ajouter de la complexité des estimateurs par noyau et demander plus de mémoire dans les calculs. Il serait donc intéressant d'étudier comment étendre l'estimateur par noyau utilisé pour le cas des données multidimensionnelles. Ici l'aspect fondamental de la notion multivariée est de trouver une solution pour analyser les données dans des espaces de grande dimension. Dans ce cas, il devient intéressant d'étudier prioritairement la manière de résoudre le phénomène de fléau de la dimension ou malédiction de la dimension.

Les perspectives d'amélioration présentées ci-dessus sont envisageables en présence des données supplémentaires. Il pourrait s'agir d'utiliser les données déjà disponibles, mais également d'exploiter des nouvelles sources des infrastructures de mesurage avancé à savoir les compteurs électriques intelligents, les informations générées par les charges électriques et les données météorologiques.

### 5.3.3 *Explorer les limites du modèle*

- Puisque la composante thermosensible est essentiellement liée au chauffage électrique, il serait judicieux de comparer cette consommation avec la puissance de chauffage électrique connexe et ceci pour les cinq maisons étudiées.
- Pour le cas adaptatif, il est aussi possible de faire une étude de la variance de la demande électrique qui pourrait être un bon indicateur de stabilité de l'approche. Ceci pour explorer la robustesse du modèle et voir comment il peut se réadapter lorsqu'un

client par exemple baisse sa température de consigne pendant un voyage de longue période.

- Pour que le modèle de prévision proposé soit plus représentatif, il est suggéré de générer plus de données de la puissance électrique totale en prenant une maison avec des mesures réelles de la puissance totale, et générer l'équivalent de la consommation typique à cette maison en utilisant la méthode de Monte-Carlo par exemple. Sur l'ensemble des nouvelles données générées, il est possible d'identifier la limite de l'approche. Il est aussi possible de créer des centaines de modèles des bâtiments simulés par un modèle boîte blanche et les comparer avec l'approche proposée pour la validation de l'approche de la prévision.

### 5.3.4 Amélioration de l'explication de la composante résiduelle

#### 5.3.4.1 Sélection de l'ordre des paramètres

Il serait intéressant d'analyser la façon pour sélectionner l'ordre des paramètres et choisir le bon nombre à inclure dans le ML de l'approche ECA/ML. Ceci pour une meilleure performance de la prévision de résiduel  $\hat{r}_k$  à partir des espérances conditionnelles des résiduelles sous-estimées et surestimées. Il semble important de s'appuyer sur certains critères d'information, exemple (AIC, BIC, Hannan Quinn,...) [137], [138] pour mesurer la qualité relative du modèle.

#### 5.3.4.2 Explication de résiduel du résiduel

Dans le contexte de la modélisation du second résiduel  $r_{res,k}$  sans la présence des données des autres charges, il est possible d'utiliser un modèle auto-regressif non-linéaire de type réseau de neurones ou le modèle Nonlinear Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) [139]. Nous pensons que ces modèles peuvent analyser la relation non-linéaire de la valeur à prévoir  $r_{res,k}$  et ses valeurs précédentes dans la série temporelle. Nous suggérons aussi de ne pas utiliser directement le modèle AR pour ce cas. Ceci à cause de la nature très aléatoire de la série temporelle de  $r_{res,k}$  qui n'est pas facilement estimable avec un modèle linéaire simple.

## Bibliographie

- [1] T. HONG, « Short Term Electric Load Forecasting », thèse de doct., 2010.
- [2] T. HONG et M. SHAHIDEHPOUR, « Load Forecasting Case Study », rapp. tech., 2015.
- [3] E. a. FEINBERG et D. GENETHLIOU, « Load Forecasting », in *Applied Mathematics for Restructured Electric Power Systems*, 2005, p. 269–285. DOI : 10.1007/0-387-23471-3\_12.
- [4] STATISTICS CANADA, « L'enquête sur les ménages et l'environnement : utilisation de l'énergie, 2013 », rapp. tech., 2016, p. 8300.
- [5] S. ROY, « La réduction de la dépendance au pétrole en Mauricie », rapp. tech., 2013, p. 1–99. adresse : <http://www.rncreq.org/images/UserFiles/files/04-Mauricie-Portrait-energetique.pdf>.
- [6] MINISTÈRE DE L'ÉNERGIE ET DES RESSOURCES NATURELLES ET STATISTIQUE, *Répartition de la consommation d'électricité par secteur*, 2013. adresse : <http://mern.gouv.qc.ca/energie/statistiques/statistiques-consommation-electricite.jsp>.
- [7] MINISTÈRE DES RESSOURCES NATURELLES ET DE LA FAUNE, « Évolution de la demande d'énergie au Québec », rapp. tech., 2005.
- [8] NATURAL RESOURCES CANADA, « Energy Efficiency Trends in Canada 1990 to 2013 », rapp. tech., 2016. adresse : <https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/energy/pdf/trends2013.pdf>.

- [9] RESSOURCES NATURELLES CANADA-L'OFFICE DE L'EFFICACITÉ ÉNERGÉTIQUE, « Évolution de l'efficacité énergétique au Canada, de 1990 à 2008 », rapp. tech., 2012. adresse : <http://oeenrncan.gc.ca/publications/statistiques/evolution10/chapitre3.cfm?attr=0>.
- [10] R. CANADA, *Froid record et pointe de consommation d'électricité*. adresse : <http://ici.radio-canada.ca/nouvelle/701259/consommation-electricite-pres-des-records>.
- [11] HYDRO-QUÉBEC, « Rapport annuel 2016 », rapp. tech., 2017, p. 1–94.
- [12] S. A. ERIC DUMONT, MICHAEL FOURNIER, « Réalisation des maisons d'expérimentation en énergétique du bâtiment », rapp. tech., 2011, p. 194.
- [13] G. de L'ONTARIO, « Plan énergétique à long-terme de l'Ontario », rapp. tech.
- [14] Y AGARWAL, T WENG et R. K. GUPTA, « Understanding the role of buildings in a smart microgrid », *2011 Design, Automation & Test in Europe*, p. 1–6, 2011. DOI : 10.1109/DATE.2011.5763195.
- [15] A. KAUR, L. NONNENMACHER et C. F. COIMBRA, « Net load forecasting for high renewable energy penetration grids », *Energy*, t. 114, p. 1073–1084, 2016. DOI : 10.1016/j.energy.2016.08.067.
- [16] CHAIRE DE GESTION DU SECTEUR DE L'ÉNERGIE - HEC MONTRÉAL, « État de l'énergie au Québec », rapp. tech., 2015, p. 1–36.
- [17] J. MARCEAU, *Chauffer, mais à quel prix ?*, 2015. adresse : <http://ici.radio-canada.ca/nouvelle/701276/chauffage-froid-quebec>.
- [18] DOMINIQUE FROMENT, *Palmarès des tarifs d'électricité : le Québec champion des bas prix*, 2013. adresse : <http://www.lesaffaires.com/secteurs-d-activite/ressources-naturelles/palmares-des-tarifs-d-electricite-le-quebec-champion-des-bas-prix/554900>.
- [19] H.-Q. DISTRIBUTION, « Avis de la régie de l'énergie sur les mesures susceptibles d'améliorer les pratiques tarifaires dans le domaine de l'électricité et du gaz naturel », rapp. tech., 2016, p. 1–65.

- [20] MINISTÈRE DE L'ÉNERGIE ET DES RESSOURCES NATURELLES, « Politique énergétique 2016-2025 », rapp. tech., 2016, p. 1–63.
- [21] Ö. ÖZKIZILKAYA, « Thermosensibilité de la demande électrique : identification de la part non linéaire par couplage d'une modélisation bottom-up et de l'approche bayésienne », thèse de doct., 2015.
- [22] S. JOVANOVIĆ, S. SAVIĆ, M. BOJIĆ, Z. DJORDJEVIĆ et D. NIKOLIĆ, « The impact of the mean daily air temperature change on electricity consumption », *Energy*, t. 88, p. 604–609, 2015. DOI : 10.1016/j.energy.2015.06.001.
- [23] TRANSITION ÉNERGÉTIQUE QUÉBEC, *Isolation*, 2018. adresse : [http://www.transitionenergetique.gouv.qc.ca/mon-habitation/conseils-pratiques/isolation/{\#}.WwlpRiDA{\\\_}IU](http://www.transitionenergetique.gouv.qc.ca/mon-habitation/conseils-pratiques/isolation/{\#}.WwlpRiDA{\_}IU).
- [24] É. VORGER, « Étude de l'influence du comportement des habitants sur la performance énergétique du bâtiment », thèse de doct., Ecole Nationale Supérieure des Mines de Paris, 2015, p. 1–475.
- [25] M. ALBORS, *Placer le consommateur québécois au centre d'un réseau électrique intelligent*, 2017. adresse : <https://novae.ca/2017/05/placer-consommateur-centre-dun-reseau-electrique-intelligent/>.
- [26] B. LEMAIGNAN, J.-Y. WILMOTTE et N. GAULT, « Études des tensions sur les énergies de réseau lors de la pointe de consommation hivernale », rapp. tech., 2014.
- [27] B. BAGHDAD, « La communication sans fil dans un réseau électrique intelligent (smart grid) : méthodologie de développement », thèse de doct., 2013, p. 1–116.
- [28] L. MA, « Projet pilote de mégadonnées pour les données des compteurs intelligents (version abrégée) », *Recueil du Symposium 2014 de Statistique Canada*, t. 6, 2018.
- [29] SYLVAIN LAROCQUE ; LA PRESSE CANADIENNE, *La Régie donne le feu vert aux compteurs intelligents d'Hydro-Québec*, 2012. adresse : <https://ici.radio-canada.ca/nouvelle/581764/energie-hq-compteur>.
- [30] M. LE CAM, « Short-term forecasting of the electric demand of HVAC systems », thèse de doct., 2016, p. 1–192. DOI : 10.1016/j.energy.2016.11.064.

- [31] AGENCE DE L'ENVIRONNEMENT ET DE LA MAÎTRISE DE L'ÉNERGIE, « Les compteurs communicants d'électricité et de gaz », rapp. tech., 2016.
- [32] O. E. MONITOR, *Optical Utility Meter LED Pulse Sensor*, 2018. adresse : <https://shop.openenergymonitor.com/optical-utility-meter-led-pulse-sensor/>.
- [33] S. WESTERVELD AND V., *Forecasting Frost Events – Comparison of Weather Forecast Sites*, 2016. adresse : <https://onspecialtycrops.wordpress.com/2016/05/13/forecasting-frost-events-comparison-of-weather-forecast-sites/>.
- [34] L. SAUNOIS, *Tests en boîte noire, grise ou blanche : quelles différences*, 2016. adresse : <https://www.nbs-system.com/blog/tests-en-boite-noire-grise-ou-blanche-quelles-differences/>.
- [35] T. HONG et D. A. DICKEY, *Electric load forecasting : fundamentals and best practices. OTexts*. 2014. adresse : <https://www.otexts.org/elf/1>.
- [36] ACTU-ENVIRONNEMENT, *Les smart grids portés par une myriade d'expériences locales*, 2012. adresse : <https://www.actu-environnement.com/ae/dossiers/smart-grid/projets-locaux.php4>.
- [37] A. JAIN et B. SATISH, « Short term load forecasting by clustering technique based on daily average and peak loads », *2009 IEEE Power and Energy Society General Meeting, PES '09*, p. 1–7, 2009. DOI : 10.1109/PES.2009.5275738.
- [38] W. LEI et M SHAHIDEHPOUR, « A Hybrid Model for Day-Ahead Price Forecasting », *Power Systems, IEEE Transactions on*, t. 25, n° 3, p. 1519–1530, 2010. DOI : 10.1109/tpwrs.2009.2039948.
- [39] P. R. J. CAMPBELL et K. ADAMSON, « Methodologies for load forecasting », *IEEE Intelligent Systems*, n° September, p. 800–806, 2006. DOI : 10.1109/IS.2006.348523.
- [40] T. HONG et S. FAN, « Probabilistic electric load forecasting : A tutorial review », *International Journal of Forecasting*, t. 32, n° 3, p. 914–938, 2016. DOI : 10.1016/j.ijforecast.2015.11.011.

- [41] H. HAHN, S. MEYER-NIEBERG et S. PICKL, « Electric load forecasting methods : Tools for decision making », *European Journal of Operational Research*, t. 199, n° 3, p. 902–907, 2009. DOI : 10.1016/j.ejor.2009.01.062.
- [42] H.-x. ZHAO et F. MAGOULÈS, « A review on the prediction of building energy consumption », *Renewable and Sustainable Energy Reviews*, t. 16, n° 6, p. 3586–3592, 2012. DOI : 10.1016/j.rser.2012.02.049.
- [43] X. LI et J. WEN, « Review of building energy modeling for control and operation », *Renewable and Sustainable Energy Reviews*, t. 37, p. 517–537, 2014. DOI : 10.1016/j.rser.2014.05.056.
- [44] L. G. SWAN et V. I. UGURSAL, « Modeling of end-use energy consumption in the residential sector : A review of modeling techniques », *Renewable and Sustainable Energy Reviews*, t. 13, n° 8, p. 1819–1835, 2009. DOI : 10.1016/j.rser.2008.09.033.
- [45] D. B. CRAWLEY, J. W. HAND, M. KUMMERT et B. T. GRIFFITH, « Contrasting the capabilities of building energy performance simulation programs », *Building and Environment*, t. 43, n° 4, p. 661–673, 2008. DOI : 10.1016/j.buildenv.2006.10.027.
- [46] A. YEZIORO, B. DONG et F. LEITE, « An applied artificial intelligence approach towards assessing building performance simulation tools », *Energy and Buildings*, t. 40, n° 4, p. 612–620, 2008. DOI : 10.1016/j.enbuild.2007.04.014.
- [47] A. MERABTINE, « Modélisation Bond Graphs en vue de l'efficacité énergétique du bâtiment », thèse de doct., 2012, p. 1–397.
- [48] T. BERTHOU, P. STABAT, R. SALVAZET et D. MARCHIO, « Comparaison de modèles linéaires inverses pour la mise en place de stratégies d'effacement », rapp. tech., 2012, p. 1–12.
- [49] C. ZAYANE, *Identification d'un modèle de comportement thermique de bâtiment à partir de sa courbe de charge*, 2011.

- [50] GREEN BUILDING CONSULTANTS, *Passive Building Design*. adresse : <http://sustainablybuilt.com/passive-building-design/>.
- [51] S. GAALOUL CHOUIKH, « Interoperabilite sur les standards Modelica et composant logiciel pour la simulation energetique des sytemes de bâtiment. », thèse de doct., 2012.
- [52] T. BERTHOU, « Développement de modèles de bâtiment pour la prévision de charge de climatisation et l'élaboration de stratégies d'optimisation énergétique et d'effacement », thèse de doct., 2013.
- [53] X. LI, J. WEN et E.-W. BAI, « Building energy forecasting using system identification based on system characteristics test », *2015 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES)*, p. 1–6, 2015. DOI : 10.1109/MSCPES.2015.7115401.
- [54] J. GONG et C.-Z. XU, « A Gray-Box Feedback Control Approach for System-Level Peak Power Management », *2010 39th International Conference on Parallel Processing*, p. 555–564, 2010. DOI : 10.1109/ICPP.2010.63.
- [55] A. EBADAT, G. BOTTEGAL, M. MOLINARI, D. VARAGNOLO, B. WAHLBERG, H. HJALMARSSON et K. H. JOHANSSON, « Multi-room occupancy estimation through adaptive gray-box models », *IEEE Conference on Decision and Control (submitted)*, n° Cdc, p. 3705–3711, 2015.
- [56] O. M. BRASTEIN, « Grey-box models for estimation of heating times for buildings », thèse de doct., 2016.
- [57] J. BRAUN et N. CHATURVEDI, « An Inverse Gray-Box Model for Transient Building Load Prediction », *HVAC&R Research*, t. 8, n° 1, p. 73–99, 2002. DOI : 10.1080/10789669.2002.10391290.
- [58] L. G. SWAN, V. I. UGURSAL et I. BEAUSOLEIL-MORRISON, « Occupant related household energy consumption in Canada : Estimation using a bottom-up neural-network technique », *Energy and Buildings*, t. 43, n° 2-3, p. 326–337, 2011. DOI : 10.1016/j.enbuild.2010.09.021.

- [59] F. AMARA, K. AGBOSSOU, A. CARDENAS, Y. DUBÉ et S. KELOUWANI, « Comparison and Simulation of Building Thermal Models for Effective Energy Management », n° April, p. 95–112, 2015.
- [60] R. BRAUSE, *Adaptive Modellierung und Simulation*, DR. RÜDIGER BRAUSE, éd. Frankfurt, 2010, p. 1–50.
- [61] P. BESSE, « Apprentissage Statistique », Département Génie Mathématique et Modélisation, rapp. tech., 2006, p. 1–145.
- [62] P. PEREZ, « Modèles et algorithmes pour l'analyse probabiliste des images », thèse de doct., 2004, p. 113.
- [63] D. ZHOU, M. BALANDAT et C. TOMLIN, « A Bayesian Perspective on Residential Demand Response Using Smart Meter Data », 2016. arXiv : 1608.03862.
- [64] M. MOKHELE, « Modelling short term probabilistic electricity demand in South Africa », thèse de doct., University of the Witwatersrand Johannesburg, 2016, p. 1–153.
- [65] M. A. RAHMAN et M. Y. EL-SHARKH, « A Bayesian Model to Enhance Domestic Energy Consumption Forecast », p. 1891–1896, 2012.
- [66] S. EVANS, « Bayesian regression analysis. », thèse de doct., University of Louisville, 2012. DOI : 10.18297/etd/412.
- [67] J. J. LÉGER, L. ; DAUDIN, « Étude d'un modèle de régression non paramétrique : la régression par directions révélatrices », *Revue de Statistique Appliquée*, t. 4, p. 21–48, 1993.
- [68] M. DEISENROTH, « Efficient Reinforcement Learning using Gaussian Processes », thèse de doct., Institut de technologie de Karlsruhe-Allemagne, 2010, p. 1–217.
- [69] H. CARSTENS, X. XIA et S. YADAVALLI, « Bayesian Energy Measurement and Verification Analysis », *Energies*, t. 11, n° 2, p. 380, 2018. DOI : 10.3390/en11020380.
- [70] N. XIAO, J. ZARNIKAU et P. DAMIEN, « Testing functional forms in energy modeling : An application of the Bayesian approach to U.S. electricity demand », *Energy Economics*, t. 29, n° 2, p. 158–166, 2007. DOI : 10.1016/j.eneco.2006.02.005.

- [71] M. SHANN et S. SEUKEN, « An active learning approach to home heating in the smart grid », *IJCAI International Joint Conference on Artificial Intelligence*, p. 2892–2899, 2013.
- [72] S. JEAN, « Reconnaissance d'objets astronomiques hi par déconvolution et par estimation de densité », thèse de doct., 1997, p. 1–100.
- [73] A. SOLOMONEN, « Monte Carlo Methods in Parameter Estimation of Nonlinear Models », *Technology*, p. 1–121, 2006.
- [74] R. B. MILLAR, *Maximum Likelihood Estimation and Inference*. Chichester, UK : John Wiley & Sons, Ltd, 2011. DOI : 10.1002/9780470094846.
- [75] C. M. B. TIPPING et M. E., « Bayesian Regression and Classification », *Advances in Learning Theory : Methods, Models and Applications*, t. 190, n° 1, p. 267–285, 2003. DOI : 10.1002/acs.744.
- [76] M. CARBON, « Estimation non paramétrique de la densité et de la régression-prévision non paramétrique », rapp. tech., p. 1–26.
- [77] É. THIÉBAUT, « Approche inverse en traitement du signal », Université Claude Bernard Lyon I, rapp. tech., 2017, p. 1–39.
- [78] J. HAUTH, « Grey-Box Modelling for Nonlinear Systems », thèse de doct., 2008, p. 357.
- [79] F. FLOUQUET, « Local weather correlations and bias in building parameter estimates from energy-signature models », *Energy and Buildings*, t. 19, n° 2, p. 113–123, 1992. DOI : 10.1016/0378-7788(92)90005-2.
- [80] E. VALOR, V. MENEU et V. CASELLES, « Daily Air Temperature and Electricity Load in Spain », *Journal of Applied Meteorology*, t. 40, n° 8, p. 1413–1421, 2001. DOI : 10.1175/1520-0450(2001)040<1413:DATAEL>2.0.CO;2.
- [81] D. M. LE COMTE et H. E. WARREN, « Modeling the Impact of Summer Temperatures on National Electricity Consumption », *Journal of Applied Meteorology*, t. 20, n° 12, p. 1415–1419, 1981. DOI : 10.1175/1520-0450(1981)020<1415:MTIOST>2.0.CO;2.

- [82] J. AL-ZAYER et A. A. AL-IBRAHIM, « Modelling the impact of temperature on electricity consumption in the Eastern Province of Saudi Arabia », *Journal of Forecasting*, t. 15, n° 2, p. 97–106, 1996. DOI : 10.1002/(SICI)1099-131X(199603)15:2<97::AID-FOR608>3.0.CO;2-L.
- [83] D. J. SAILOR, « Relating residential and commercial sector electricity loads to climate—evaluating state level sensitivities and vulnerabilities », *Energy*, t. 26, n° 7, p. 645–657, 2001. DOI : 10.1016/S0360-5442(01)00023-8.
- [84] T. PEDERSEN, « Forecasting model of electricity demand in the Nordic countries », rapp. tech., 2014.
- [85] J. CRESPO CUARESMA, J. HLOUSKOVA, S. KOSSMEIER et M. OBERSTEINER, « Forecasting electricity spot-prices using linear univariate time-series models », *Applied Energy*, t. 77, n° 1, p. 87–106, 2004. DOI : 10.1016/S0306-2619(03)00096-5.
- [86] G. A. DARBELLAY et M. SLAMA, « Forecasting the short-term demand for electricity : Do neural networks stand a better chance ? », *International Journal of Forecasting*, t. 16, n° 1, p. 71–83, 2000. DOI : 10.1016/S0169-2070(99)00045-X.
- [87] E. M. ALDRICH, « Advanced Quantitative Methods », rapp. tech., 2014, p. 1–130.
- [88] A. G.S.G.J. S. GIRARD, « Trois méthodes non paramétriques pour l'estimation de courbes de référence - application à l'analyse de propriétés biophysiques de la peau », *Revue de Statistique Appliquée*, t. 4, p. 5–15, 1967.
- [89] C. FENG, H. WANG, N. LU, T. CHEN, H. HE, Y. LU et X. M. TU, « Log-transformation and its implications for data analysis. », *Shanghai archives of psychiatry*, t. 26, n° 2, p. 105–9, 2014. DOI : 10.3969/j.issn.1002-0829.2014.02.009.
- [90] A. I. DOUNIS, « Artificial intelligence for energy conservation in buildings », *Advances in Building Energy Research*, t. 4, n° 1, p. 267–299, 2010. DOI : 10.3763/aber.2009.0408.
- [91] M. KRARTI, « An Overview of Artificial Intelligence-Based Methods for Building Energy Systems », *Journal of Solar Energy Engineering*, t. 125, n° 3, p. 331, 2003. DOI : 10.1115/1.1592186.

- [92] S. A. KALOGIROU, « Artificial neural networks in energy applications in buildings », *International Journal of Low-Carbon Technologies*, t. 1, n° 3, p. 201–216, 2006. DOI : 10.1093/ijlct/1.3.201.
- [93] H. DEMUTH, « Neural Network Toolbox », rapp. tech. 1, 2002, p. 1–846. DOI : 10.1016/j.neunet.2005.10.002. arXiv : 1404.7828.
- [94] M. ROSENBLATT, « Remarks on Some Nonparametric Estimates of a Density Function », *The Annals of Mathematical Statistics*, t. 27, n° 3, p. 832–837, 1956. DOI : 10.1214/aoms/1177728190.
- [95] E. PARZEN, « On Estimation of a Probability Density Function and Mode », *The Annals of Mathematical Statistics*, t. 33, n° 3, p. 1065–1076, 1962. DOI : 10.1214/aoms/1177704472.
- [96] R. J. G. HYNDMAN et ATHANASOPOULOS, « Measuring forecast accuracy », in *Forecasting : principles and practice. OTexts*, 2012. adresse : <http://otexts.com/fpp>.
- [97] R. J. HYNDMAN et Y. KHANDAKAR, « Automatic Time Series Forecasting : The forecast Package for R », *Journal of Statistical Software*, t. 27, n° 3, 2008. DOI : 10.18637/jss.v027.i03.
- [98] R. HYNDMAN, X. ZHANG et M. KING, « Bandwidth selection for multivariate kernel density estimation using mcmc », rapp. tech., 2004, p. 1–27.
- [99] G. HYNDMAN, R.J. AND ATHANASOPOULOS, *Forecasting : Principales & Practice*. 2013, p. 1–138. adresse : <http://otexts.org/fpp/>.
- [100] S. FAN et R. J. HYNDMAN, « Forecasting Electricity Demand in Australian National Electricity Market », *Proc. IEEE PES General Meeting, 24-29 July, San Diego, USA*, n° February, p. 1–4, 2012.
- [101] W. DAI, « Different-based methods in nonparametric regression models », thèse de doct., 2014, p. 1–140.
- [102] E. A. NADARAYA, « On Estimating Regression », *Theory of Probability & Its Applications*, t. 9, n° 1, p. 141–142, 1964. DOI : 10.1137/1109020.

- [103] V. VIALON, « Processus empiriques, estimation non paramétrique et données censurées », thèse de doct., 2006, p. 1–145.
- [104] S. M. SOME, « Estimations non paramétriques par noyaux associés multivariés et applications », thèse de doct., 2015, p. 1–194. arXiv : 1502.01173.
- [105] T. GRENIER, « Apport de l'espace des caractéristiques et des paramètres d'échelle adaptatifs pour le filtrage et la segmentation d'image », thèse de doct., 2005, p. 1–190.
- [106] C. G. ATKESON, A. W. MOORE et S. SCHAAL, « Locally Weighted Learning », in *Artificial Intelligence Review*, 1/5, t. 11, 1997, p. 11–73. DOI : 10 . 1023 / A : 1006559212014.
- [107] S. MALLAT, « L'apprentissage face à la malédiction de la grande dimension », rapp. tech., p. 1–7.
- [108] W. CONTRIBUTORS, *Sankey diagram*, 2018. adresse : [https://en.wikipedia.org/w/index.php?title=Sankey{\\\_}diagram{\&}oldid=828307349](https://en.wikipedia.org/w/index.php?title=Sankey%7B%7Ddiagram%7B%7D&oldid=828307349).
- [109] C. GUZMAN, K. AGBOSSOU et A. CARDENAS, « Modeling of residential centralized and baseboard space heating systems », in *2016 IEEE 25th International Symposium on Industrial Electronics (ISIE)*, IEEE, 2016, p. 726–731. DOI : 10 . 1109 / ISIE . 2016 . 7744979.
- [110] M. SUGIYAMA, I. TAKEUCHI, T. SUZUKI, T. KANAMORI, H. HACHIYA et D. OKANO-HARA, « Least-Squares Conditional Density Estimation », *IEICE transactions on information and systems*, n° 3, p. 583–594, 2010. DOI : 10 . 1587 / transinf . E93 . D . 583.
- [111] S. GHOSAL, « The Dirichlet process, related priors and posterior asymptotics », rapp. tech., 2010, p. 35–79. DOI : 10 . 1017 / CB09780511802478 . 003. adresse : <http://ebooks.cambridge.org/ref/id/CB09780511802478A018>.
- [112] D. W. SCOTT et S. R. SAIN, « Multidimensional Density Estimation », *Handbook of Statistics*, t. 24, n° August 2004, p. 229–261, 2004. DOI : 10 . 1016 / S0169 - 7161(04)24009-3.
- [113] V. S. MOERTINI, « Introduction To Five Data Clustering Algorithms », rapp. tech. 2, 2002, p. 87–96.

- [114] J. KOLODA, A. M. PEINADO et V. SANCHEZ, « On the application of multivariate kernel density estimation to image error concealment », in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2013, p. 1330–1334. DOI : 10.1109/ICASSP.2013.6637867.
- [115] T. HONG, T. D. LAING et P. WANG, « Four Best Practices of Load Forecasting for Electric Cooperatives », *Rural Electric Power Conference (REPC), 2014 IEEE*, p. 1–4, 2014. DOI : 10.1109/REPCon.2014.6842203.
- [116] R. HYNDMAN et G. ATHANASOPOULOS, *Transformation, stationarity, differencing*, 2014. adresse : <https://www.otexts.org/fpp/2/4/>.
- [117] R. J. HYNDMAN et S. FAN, « Density forecasting for long-term peak electricity demand », *IEEE Trans. Power Syst.*, t. 25, n° 2, p. 1142–1153, 2010. DOI : 10.1109/TPWRS.2009.2036017.
- [118] Y. ZHANG, Z. O'NEILL, B. DONG et G. AUGENBROE, « Comparisons of inverse modeling approaches for predicting building energy performance », *Building and Environment*, t. 86, p. 177–190, 2015. DOI : 10.1016/j.buildenv.2014.12.023.
- [119] N. HENAO, « Stratégie pour le démarrage à froid d'une pile à combustible à membrane d'échange de protons dans un véhicule électrique », thèse de doct., Université du Québec à Trois-Rivières, 2013, p. 1–114.
- [120] Y. A. LECUN, L. BOTTOU, G. B. ORR et K. R. MULLER, « Efficient backprop », *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, t. 7700 LECTU, p. 9–48, 2012. DOI : 10.1007/978-3-642-35289-8-3. arXiv : 9809069v1 [arXiv:gr-qc].
- [121] A. SAYED et T. KAILATH, « Recursive Least-Squares Adaptive Filters », in *Digital Signal Processing Fundamentals*, 2009, p. 1–40. DOI : 10.1201/9781420046076-c21.
- [122] R. J. HYNDMAN et A. B. KOEHLER, « Another look at measures of forecast accuracy », *International Journal of Forecasting*, t. 22, n° 4, p. 679–688, 2006. DOI : 10.1016/j.ijforecast.2006.03.001.

- [123] W. TANSEY, K. PICHOTTA et J. G. SCOTT, « Better Conditional Density Estimation for Neural Networks », *ArXiv preprint arXiv :1606.02321*, 2016. arXiv : arXivpreprintarXiv:1606.02321.
- [124] F. BOURGIN, « Comment quantifier l'incertitude prédictive en modélisation hydrologique ? Travail exploratoire sur un grand échantillon de bassins versants », thèse de doct., 2014, p. 1–230.
- [125] R. MERHAB, « Fiabilité des outils de prévision du comportement des systèmes thermiques complexes », thèse de doct., Université de Bordeaux 1, 2013, p. 1–155.
- [126] U. D. HANEBECK et A. LINDQUIST, « Moment-based Dirac Mixture Approximation of Circular Densities », *IFAC Proceedings Volumes*, t. 47, n° 3, p. 5040–5048, 2014. DOI : 10.3182/20140824-6-ZA-1003.02486.
- [127] D. M. BASHTANNYK et R. J. HYNDMAN, « Bandwidth selection for kernel conditional density estimation », *Computational Statistics and Data Analysis*, t. 36, n° 3, p. 279–298, 2001. DOI : 10.1016/S0167-9473(00)00046-3.
- [128] M. OLIVEIRA, R. M. CRUJEIRAS et A. RODRIGUEZ-CASAL, « Nonparametric circular methods for exploring environmental data », *Environmental and Ecological Statistics*, t. 20, n° 1, p. 1–17, 2013. DOI : 10.1007/s10651-012-0203-6.
- [129] R. JACOBS, RODNEY L. AND JONES, « A Bayesian Approach to Adaptive Expectations », rapp. tech. March, 1977.
- [130] C.-H. CHENG, J.-W. LIU et C.-C. HUNG, « Multi-attribute Fuzzy Time Series Based on Adaptive Expectation Model », in *Second International Conference on Innovative Computing, Informatio and Control (ICICIC 2007)*, IEEE, 2007, p. 80–80. DOI : 10.1109/ICICIC.2007.399.
- [131] M. KRISTAN, D. SKOCAJ et A. LEONARDIS, « Online kernel density estimation for interactive learning », *Image and Vision Computing*, t. 28, n° 7, p. 1106–1116, 2010. DOI : 10.1016/j.imavis.2009.09.010.
- [132] M. NERLOVE, « Adaptive Expectations and Cobweb Phenomena », *The Quarterly Journal of Economics*, t. 72, n° 2, p. 227, 1958. DOI : 10.2307/1880597.

- [133] L. S. RESENDE, « Simplified RLS algorithm for adaptive IFIR filtering », in *2000 10th European Signal Processing Conference, Tampere, Finland, 2000*, 2000, p. 1–4.
- [134] C. BENNETT, R. STEWART et J. LU, « Autoregressive with Exogenous Variables and Neural Network Short-Term Load Forecast Models for Residential Low Voltage Distribution Networks », *Energies*, t. 7, n° 5, p. 2938–2960, 2014. DOI : 10.3390/en7052938.
- [135] M. MARCELLINO, J. H. STOCK et M. W. WATSON, « A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series », *Journal of Econometrics*, t. 135, n° 1-2, p. 499–526, 2006. DOI : 10.1016/j.jeconom.2005.07.020.
- [136] G. CHAGNY et A. ROCHE, « Adaptive estimation in the functional nonparametric regression model », *Journal of Multivariate Analysis*, t. 146, p. 105–118, 2016. DOI : 10.1016/j.jmva.2015.07.001.
- [137] C. A. MEDEL et S. C. SALGADO, « Does BIC Estimate and Forecast Better than AIC? », *Revista de Analisis Economico*, t. 28, n° 1, p. 47–64, 2013. DOI : 10.4067/S0718-88702013000100003.
- [138] X. GUYON et J.-f. YAO, « On the Underfitting and Overfitting Sets of Models Chosen by Order Selection Criteria », *Journal of Multivariate Analysis*, t. 70, n° 2, p. 221–249, 1999. DOI : 10.1006/jmva.1999.1828.
- [139] M. LANNE et P. SAIKKONEN, « Non-linear GARCH models for highly persistent volatility », *The Econometrics Journal*, t. 8, n° 2, p. 251–276, 2005. DOI : 10.1111/j.1368-423X.2005.00163.x.
- [140] J. HINMAN et E. HICKEY, « Modeling and Forecasting Short-Term Electricity Load Using Regression Analysis », thèse de doct., 2009, p. 1–51.
- [141] S. SIDDIQUE, « Automation of Energy Demand Forecasting », thèse de doct., 2013. DOI : 10.1017/CB09781107415324.004. arXiv : arXiv:1011.1669v3.
- [142] S. TATSA, « Modélisation et prévision de la consommation horaire d'électricité au Québec Comparaison de méthodes de séries temporelles », rapp. tech., 2013.

- [143] A. HANBANCHONG, « SARIMA Based Network Bandwidth Anomaly Detection », *2012 Ninth International Joint Conference on Computer Science and Soft Engineering (JCSSE)*, p. 104–108, 2012.
- [144] H.-t. YANG, C.-m. HUANG et C.-l. HUANG, « Identification of ARMAX Model for Short Term Load Forecasting : An Evolutionary Programming Approach », p. 325–330, 1995.
- [145] F. P. INCROPERA, D. P. DEWITT, T. L. BERGMAN et A. S. LAVINE, *Fundamentals of Heat and Mass Transfer*. 2007, p. 997. arXiv : 1105-.

## Annexe A - Techniques de prévision

Cette annexe est citée dans la partie « 2.2.1- État de l'art sur les techniques de prévision de la consommation d'électricité ». Nous présentons ici la suite des méthodes des séries temporelles fréquemment utilisées dans la prévision de la demande d'électricité.

### A.1 Séries temporelles et régression

Deux processus linéaires qui sont à la base des séries temporelles, le processus de la moyenne mobile (MA) sur les erreurs et le processus autorégressif (AR) qui est utilisé lorsque la valeur à prévoir est corrélée avec les valeurs précédentes dans la série temporelle [84]. Les modèles MA et AR sont définies comme suit :

— Le processus  $MA(q)$

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (1-1)$$

Le processus MA est construit en calculant la moyenne mobile de l'erreur générée à chaque point de temps. En règle générale, les valeurs moyennes sont pondérées. Le processus s'exprime comme une combinaison linéaire de bruit blanc à l'horizon  $t - q$ . Où  $Y_t$  indique la valeur à prévoir à l'instant  $t$  qui est une moyenne pondérée de l'erreur aux instants précédents du temps, et les valeurs  $\theta$  se sont les coefficients de la moyenne mobile [140], [141]. L'équation 1-2 pourrait être réécrite comme suit :

$$Y_t = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (1-2)$$

Où  $q$  représentant le nombre des erreurs précédentes intégrés dans le modèle,  $q$  est connu comme l'ordre du modèle MA. Notez que, selon [142] la méthode des moyennes mobiles n'est pas efficace pour la modélisation de la demande électrique, elle est généralement appliquée

pour le lissage de la série temporelle.

Le processus autoregressive  $AR(p)$  a la forme suivante :

$$Y_t = \varepsilon_t + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} \quad (1-3)$$

Avec  $Y_t$  indique la valeur à prévoir à l'instant  $t$  et  $Y_{t-i}$  se sont les valeurs enregistrées à l'instant  $t$  pour  $i \in [1, n]$ ,  $n$  est le nombre d'échantillon,  $\varphi_i$  indique les coefficients de  $Y_{t-i}$ , et le terme supplémentaire comprennent la variable aléatoire normale en fonction du temps  $\varepsilon_t$ . L'équation 1-4 pourrait être réécrite comme suit :

$$Y_t = \varepsilon_t + \sum_{i=1}^p \varphi_i Y_{t-i} \quad (1-4)$$

Où  $p$  représente le nombre de valeurs antérieures de la série temporelle qui doit être intégré dans le modèle. Cette variable  $p$  est l'ordre du modèle AR [141]. La combinaison entre les deux processus produit le modèle  $ARMA(p, q)$  suivant :

$$Y_t + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (1-5)$$

Selon [84] l'utilisation des deux composantes permet de mieux capturer la dynamique sous-jacente, qui est le résultat de l'autocorrelation des variations saisonnières et les aspects aléatoires. Il existe aussi la version saisonnières du modèle  $ARMA(p, q)$ , qui permet aux processus AR et MA de s'adapter avec les cycles observés (semaine, mois, saison) [142]. Ces modèles sont notés par  $SARMA(p, q)(P, Q)_s$  ou  $s$  est la cyclicité du processus et les paramètres  $P, Q$  et  $s$  sont spécifiques à la composante saisonnières. Des modèles dynamiques équivalents à ceux existent pour les séries non-stationnaires, pour lesquelles la non-stationnarité est d'origine stochastique [143]. Il s'agit des modèles  $ARIMA(p, d, q)$  et  $SARIMA(p, d, q)(P, D, Q)_s$ , qui sont des modèles adaptés aux séries qui deviennent stationnaires par différenciation d'ordre  $d$  pour la partie non saisonnière et d'ordre  $D$  pour la partie saisonnière [142], [143]. Selon la littérature si le modèle  $ARMA$  est insuffisant pour la prévision de la demande électrique à court terme, il est suggéré d'inclure une variable exogène tels que la température extérieure qui peut potentiellement améliorer les performances de prévision [84]. Nous obtenons alors

des modèles appelés ARX, ou ARMAX. Le processus exogène autorégressive ARX, est un modèle de série temporelle autorégressive avec des paramètres d'entrée exogènes, il peut être exprimé par :

$$Y_t + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} = u_t + \omega_1 u_t + \dots + \omega_s u_{t-s} + \varepsilon_t \quad (1-6)$$

Le processus ARMAX, est une extension du modèle ARX où les paramètres de la moyenne mobile sont ajoutés, il est exprimé comme suit :

$$Y_t + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} = u_t + \omega_1 u_t + \dots + \omega_s u_{t-s} + \varepsilon_t + \theta_1 \varepsilon_t + \dots + \theta_q \varepsilon_{t-q} \quad (1-7)$$

L'équation 1-8 pourrait être réécrite de la façon suivante :

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=0}^b \omega_i u_{t-i} \quad (1-8)$$

Avec  $Y_t$  est la valeur à prévoir à l'instant  $t$  et  $p$  est l'ordre d'autorégression,  $q$  est l'ordre de la moyenne mobile,  $\varphi$  se sont les paramètres autorégressifs et  $\theta$  se sont les paramètres de la moyenne mobile,  $u$  indique les entrées exogènes,  $b$  indique l'ordre d'entrées exogènes,  $\omega_i$  est le paramètre de l'entrée exogène à l'instant  $i$ , et  $\varepsilon_t$  est un bruit blanc [140], [141], [144].

## Annexe B - Les données de cinq maisons

### B.1 Les cinq profils horaires résidentiels réels

Nous présentons les données recueillies de la puissance totale et la température extérieure pour chacune des maison sur l'année 94-96.

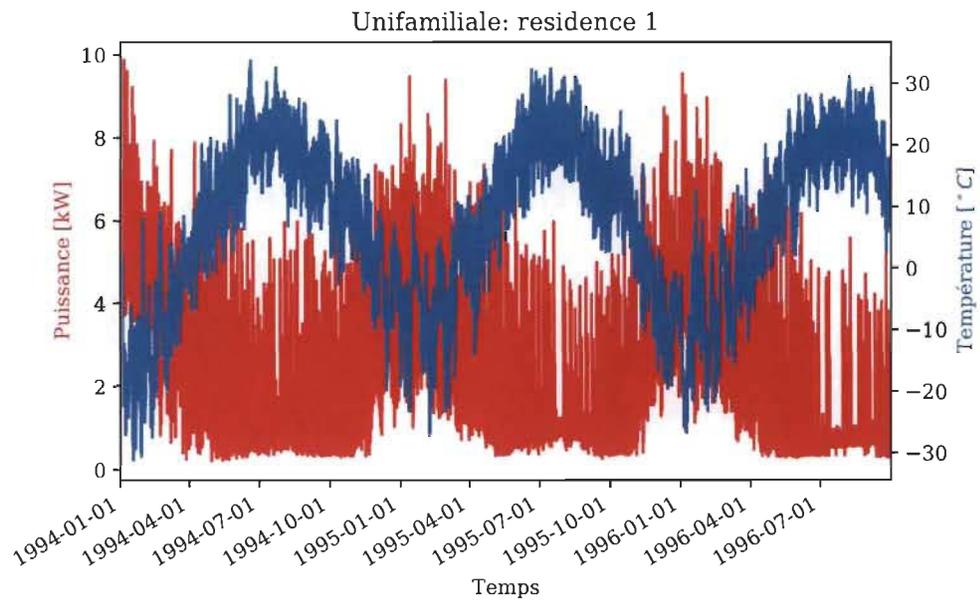


Figure. 2-1 Profils de température et de puissance totale pour la résidence unifamiliale 1

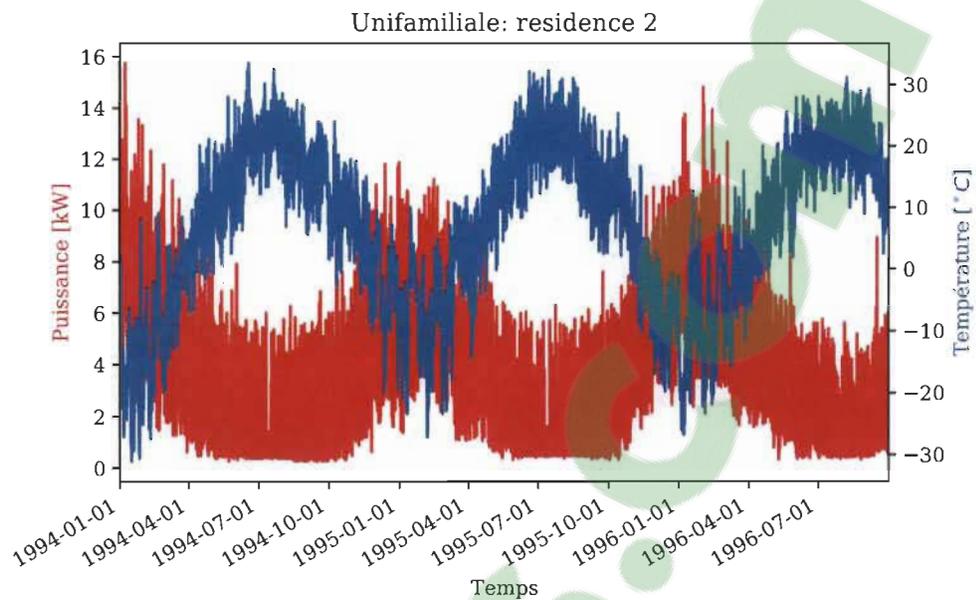


Figure. 2-2 Profils de température et de puissance totale pour la résidence unifamiliale 2

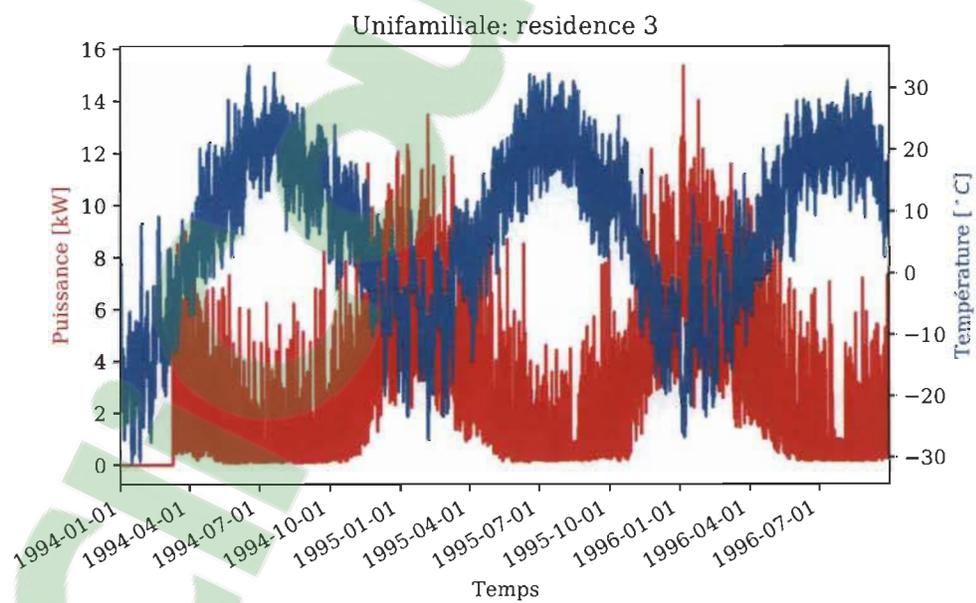


Figure. 2-3 Profils de température et de puissance totale pour la résidence unifamiliale 3

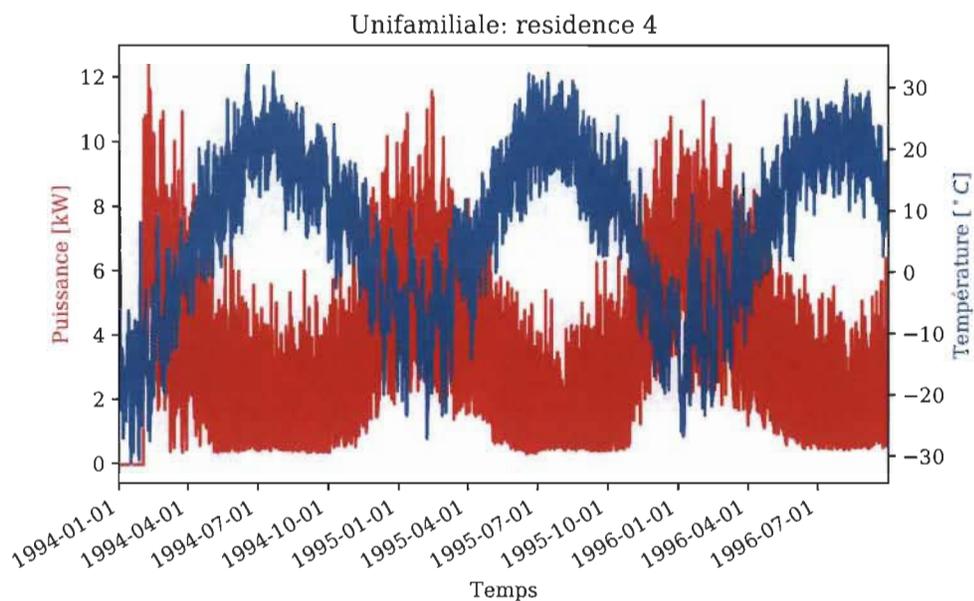


Figure. 2-4 Profils de température et de puissance totale pour la résidence unifamiliale 4

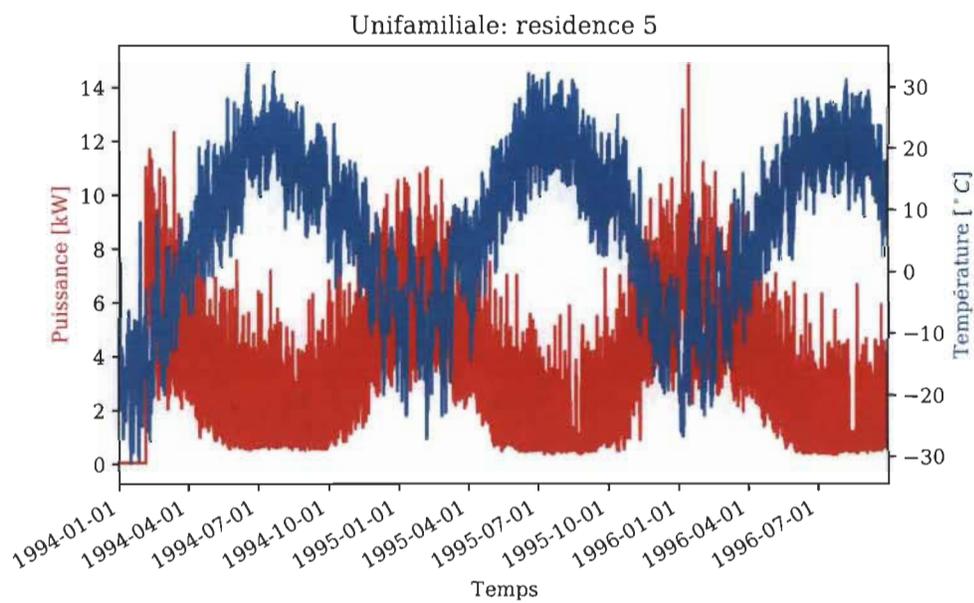


Figure. 2-5 Profils de température et de puissance totale pour la résidence unifamiliale 5

## Annexe C - Notes techniques

### C.1 Coefficient de déperdition d'un bâtiment

Le coefficient de déperdition permet d'évaluer le niveau d'isolation équivalent d'un bâtiment. Il s'agit de la moyenne des déperditions à travers l'ensemble de ses parois. Il caractérise les pertes de chaleur à travers les murs, la toiture, le sol, les portes et fenêtres. Ces pertes de chaleur sont calculées pour une différence de température entre l'extérieur et l'intérieur de 1°C. Il s'exprime en Watt par mètre cube et par degré [145]. De façon simple, la déperdition thermique d'un bâtiment peut s'exprimer de la manière suivante :

$$Q_d = (UA_{mursexternes} + UA_{plafond} + UA_{fenetres} + \dot{m}c_p) \cdot (x^{int} - x^{ext}) + UA_{fondation}(x^{int} - x^{ext}) \approx UA_{biment}(x^{int} - x^{ext}) \quad (3-1)$$

### C.2 Fonctions des probabilités jointes et conditionnelles

Une brève description des notions correspondantes aux fonctions de probabilité jointe et conditionnelle sont présentés dans l'exemple illustré sur la Figure. 3-1. La figure (a) à gauche montre les distributions des fonctions marginales  $f(x)$  et  $f(y)$  ainsi la distribution jointe du couple  $f(x,y)$ .

Dans cet exemple  $X$  et  $Y$  sont considérés des variables aléatoires à deux dimensions, admettons une fonction de densité jointe  $f(x,y)$  et densités marginales  $f_X(x)$  et  $f_Y(y)$ . Trois conditions doivent être vérifiées pour calculer ces densités : 1) la fonction  $f$  est bien continue sur l'intervalle  $[a;b]$ , 2) cette fonction est positive sur cet intervalle,  $a \leq f_X(x) \leq b$ , 3) l'intégrale de la fonction par rapport à  $x$  ou  $y$  doit être égale à 1,  $\int_a^b f_X(x)dx = 1$ .

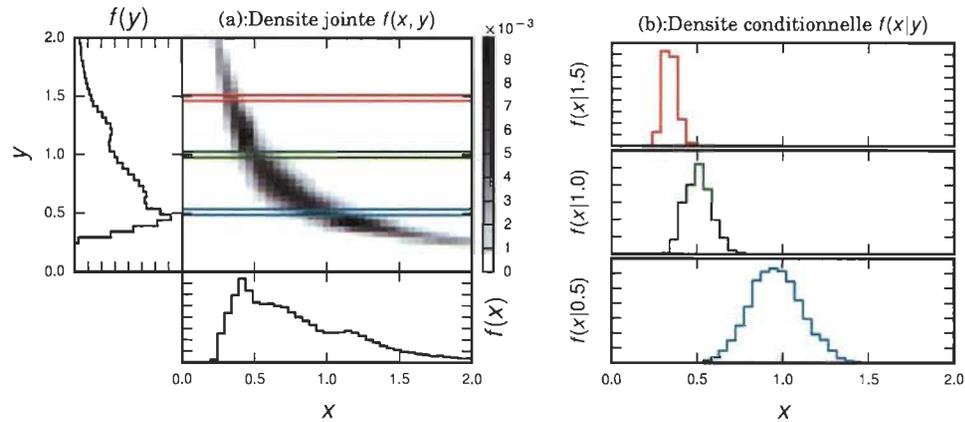


Figure. 3-1 Densité jointe et conditionnelle.

La fonction de densité jointe  $f(x,y)$  est une probabilité des événements  $X$  et  $Y$  qui se produisent simultanément. Tandis que, la fonction de densité marginale est défini par la probabilité d'occurrence d'un seul événement. Dans le cas continu, nous obtenons la fonction de densité marginale  $f_X(x)$  en sommant les  $f(x,y)$  suivant les lignes et celle de  $f_Y(y)$ , en sommant les  $f(x,y)$  selon les colonnes de la façon suivante :

$$\begin{aligned} f_X(x, \cdot) &= \int f(x,y) dy \\ f_Y(\cdot, y) &= \int f(x,y) dx \end{aligned} \quad (3-2)$$

De même pour le cas des variables aléatoires discrètes, la définition des fonctions densités marginales et jointes consiste de remplacer les intégrales du cas continu par des sommes. Enfin, nous définissons les densités conditionnelles à partir des densités jointes et marginales calculées dans le panneau (a) de gauche. Sur la Figure. 3-1-(b) nous présentons trois distributions des fonctions conditionnelles  $f(x|y)$  pour trois valeurs différentes de  $y$ .

Comme indiqué dans le panneau de gauche, nous présentons trois distributions des fonctions conditionnelles  $f(x|y)$  pour trois valeurs différentes de  $y$ . Ceci est calculé à partir de la densité  $f(x,y)$  montrée dans l'exemple de la Figure. 3-1. La fonction conditionnelle définit la probabilité que l'événement  $x$  se réalise sachant que l'événement  $y$  est réalisé. Ceci va nous faire, restreint l'univers de possible de  $y$ , donc se retrouve dans la situation la probabilité que l'événement  $x$  se réalise sachant l'événement  $y = 0.5$  par exemple.

### C.3 Évaluation de la précision de la régression

Les études présentées dans la littérature proposent des métriques différentes pour mesurer la précision des modèles de régression ainsi que celle de la prévision de la demande d'électricité. Le but avec ces métriques est de mesurer l'erreur et d'obtenir un résumé clair sur sa distribution par rapport à la référence de la consommation électrique observée. Hyndman et Koehler [122] donnent une revue détaillée de différentes mesures de précision en proposant les classer en groupes de métriques. Dans ce classement, le biais a été pris en compte pour nous dire si nous avons tendance à surestimer ou sous-estimer la consommation électrique. De plus, si la valeur prédite de la demande est trop biaisée dans les deux sens, ceci peut dégrader la capacité de prévoir correctement le futur de la consommation. Par ailleurs, la forme de base de mesure de précision des prévisions est de calculer l'erreur  $e_k = y_k - \hat{y}_k$  entre la demande électrique mesurée et prédite. Cette formule a subi avec le temps des dérivées en fonction des analyses préconisées à savoir :

#### C.3.1 Mesures dépendantes de l'échelle

Les deux mesures dépendantes de l'échelle les plus couramment utilisées sont basées sur les erreurs absolues  $AE_k = |y_k - \hat{y}_k|$  et les erreurs quadratiques  $SE_k = (y_k - \hat{y}_k)^2$ . Les deux mesures d'erreurs sont moyennées par une moyenne arithmétique pour donner l'erreur absolue moyenne (MAE, équation (3-3)) et l'erreur quadratique moyenne (RMSE, équation (3-4)). Ces métriques sont utilisées seulement lors de la comparaison entre les séries temporelles des données ayant les mêmes échelles.

$$MAE = \frac{1}{n} \sum_{k=1}^n |y_k - \hat{y}_k| \quad (3-3)$$

$$MSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2} \quad (3-4)$$

En comparant les deux métriques, le MAE mesure la moyenne des erreurs dans l'ensemble de données prédites sans considérer le signe et ceci à cause de la valeur absolue, cette dernière est moins sensible aux valeurs extrêmes de prévision. Le RMSE supprime le signe de l'erreur

par le carré des erreurs en évitant le besoin de la valeur absolue. Il pénalise les erreurs extrêmes de prévision. Par contre, il donne un poids relativement élevé aux erreurs élevés, car elles sont mises au carré avant d'être moyennées, donc il peut être plus approprié que dans certains cas mais difficile de l'interpréter comparé au MAE.

### C.3.2 Erreurs en pourcentage

L'erreur en pourcentage est donné par  $PE_k = 100 \frac{(y_k - \hat{y}_k)}{(y_k)}$ . Les erreurs en pourcentage ont l'avantage d'être indépendantes de l'échelle et sont donc fréquemment utilisées pour comparer les performances prévues entre différents ensembles de données. La mesure la plus couramment utilisée est l'erreur de pourcentage absolu moyen (MAPE, équation (3-5)).

$$MAPE = \frac{1}{n} \sum_{k=1}^n \left| 100 \frac{y_k - \hat{y}_k}{y_k} \right| \quad (3-5)$$

Cette métrique a l'inconvénient d'être infinie ou indéfinie si  $y_k = 0$  pour tout  $k$  sur la période d'analyse, ainsi elle peut avoir une valeur extrême quand  $y_k$  est proche de zéro. Notamment, lorsque l'ensemble des données incluent de très petites valeurs, c'est le cas de la demande intermittente, où il très possible de tomber sur des valeurs nulles dans les mesures de  $y_k$ .

### C.3.3 Erreurs d'échelle

Hyndman et Koehler[122] ont proposé la métrique d'erreur d'échelle absolue moyenne comme alternative à l'utilisation d'erreurs en pourcentage. Elle est définie par (3-6).

$$MASE = \frac{\sum_{k=1}^n |e_k|}{\frac{1}{n-1} \sum_{k=2}^n |y_k - \hat{y}_{k-1}|} \quad (3-6)$$

Où le numérateur correspond à l'erreur de prévision sur une période donnée, définie entre la valeur réelle  $y_k$  et la valeur prédite  $\hat{y}_k$  et le dénominateur représente l'erreur absolue moyenne de la prévision naïve, cette dernière consiste à reprendre la même valeur, c'est-à-dire, la prévision pour l'instant  $k$  correspond à la valeur de données à l'instant  $k - 1$ .