



**HAL**  
open science

# Modélisation de la consommation électrique à partir de grandes masses de données pour la simulation des alternatives énergétiques du futur

Thibaut Barbier

► **To cite this version:**

Thibaut Barbier. Modélisation de la consommation électrique à partir de grandes masses de données pour la simulation des alternatives énergétiques du futur. Energie électrique. Université Paris sciences et lettres, 2017. Français. NNT : 2017PSLEM032 . tel-01774316

**HAL Id: tel-01774316**

**<https://pastel.hal.science/tel-01774316>**

Submitted on 23 Apr 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres  
PSL Research University

Préparée à MINES ParisTech

Modélisation de la consommation électrique à partir de grandes masses de données pour la simulation des alternatives énergétiques du futur

**École doctorale n°521**

SCIENCES ET MÉTIERS DE L'INGÉNIEUR

**Spécialité** ENERGÉTIQUES ET PROCÉDÉS

Soutenue par **Thibaut BARBIER**  
le 22 décembre 2017

Dirigée par **François-Pascal NEIRAC**  
et **Robin GIRARD**

## COMPOSITION DU JURY :

**M. Bruno François**  
Ecole Centrale de Lille, Président

**M. Hamid Ben Ahmed**  
ENS Rennes, Rapporteur

**M. Bernard Multon**  
ENS Rennes, Examineur

**Mme. Elena Magliaro**  
EcoCO2, Examineur

**M. Eric Peirano**  
Technofi, Examineur

**M. François-Pascal Neirac**  
MINES ParisTech, Examineur



# Avant-propos

Cette thèse a été effectuée au centre PERSEE de Mines ParisTech, sous la direction de François-Pascal Neirac. Robin Girard a été maître de thèse. Elle a été financée par un contrat doctoral fléché pour normalien de l'École normale supérieure de Rennes.

Cette thèse se base, pour une partie importante, sur un partenariat avec Enedis, entreprise à mission de service public, et principal gestionnaire du réseau de distribution électrique français. Ce partenariat a été riche d'enseignements, tant sur le cadrage des besoins, que sur la fourniture de cas d'études mettant en application les méthodes et modèles de la thèse, avec l'utilisation de données propres aux gestionnaires du réseau de distribution. Bien que cette thèse aborde surtout le cas de la modélisation de la consommation électrique française, les méthodes sont généralisables à d'autres pays, et des communications scientifiques internationales ont été réalisées en ce sens. Un stage de recherche de trois mois à l'université de Stanford a aussi pu être effectué durant cette thèse, de juin à août 2017. Ce séjour a été financé en partie par le centre PERSEE et la fondation Mines ParisTech. Il a constitué les premiers pas d'une collaboration et de travaux sur des problématiques communes entre le centre PERSEE et le département "Civil and Environmental Engineering" de l'université de Stanford.



# Table des matières

<b>Avant-propos</b>	<b>1</b>
<b>Table des figures</b>	<b>9</b>
<b>Liste des tableaux</b>	<b>13</b>
<b>Liste des notations</b>	<b>14</b>
<b>Remerciements</b>	<b>17</b>
<b>1 Introduction</b>	<b>19</b>
1.1 Contexte et enjeux de la thèse . . . . .	19
1.2 Problématique, objectifs, contributions. . . . .	22
<b>2 État de l’art des modèles de consommation électrique</b>	<b>23</b>
Résumé du chapitre . . . . .	23
2.1 Introduction du chapitre . . . . .	24
2.2 Définitions, hypothèses et grille de lecture . . . . .	25
2.2.1 La consommation électrique . . . . .	25
2.2.2 Classification des facteurs influant sur la consommation électrique . . . . .	33
2.3 Classification des différents modèles . . . . .	35
2.3.1 Les modèles basés sur les consommations agrégées d’une zone . . . . .	35
2.3.2 Les modèles basés sur la consommation par type de clients . . . . .	36
2.3.3 Les modèles basés sur la consommation par usage . . . . .	40
2.3.4 Les modèles basés sur la consommation par type de clients et par usage . . . . .	41
2.3.5 Les modèles basés sur la consommation par client . . . . .	42
2.3.6 Les modèles basés sur la consommation par appareil . . . . .	43
2.3.7 Bilan des différents modèles de la littérature . . . . .	45
2.4 Positionnement des travaux de la thèse . . . . .	46
<b>3 Constitution d’une base de données</b>	<b>47</b>
Résumé du chapitre . . . . .	47
3.1 Introduction du chapitre . . . . .	48
3.2 Présentation des bases de données . . . . .	48
3.2.1 Les bases de données d’Enedis . . . . .	48
3.2.2 Les bases de données de l’INSEE . . . . .	53
3.2.3 La base de données des fichiers fonciers . . . . .	54
3.2.4 Les bases de données du CEREN . . . . .	55

3.2.5	Les bases de données de l'IGN . . . . .	55
3.2.6	Les données sur le parc d'appareils . . . . .	55
3.2.7	Les campagnes de mesures . . . . .	56
3.2.8	Les bases de données sur la météorologie . . . . .	57
3.2.9	Conclusion sur les bases de données . . . . .	57
3.3	Association de la base clients et de la base logement . . . . .	58
3.3.1	Introduction au problème d'association . . . . .	58
3.3.2	Description du problème d'association . . . . .	58
3.3.3	Association par la méthode de référence . . . . .	59
3.3.4	Association par une méthode d'optimisation . . . . .	59
3.3.5	Quantification de l'intérêt de la méthode d'association . . . . .	66
3.4	Conclusion et perspectives du chapitre . . . . .	73
<b>4</b>	<b>Construction d'un modèle de consommation électrique</b>	<b>75</b>
	Résumé du chapitre . . . . .	75
4.1	Introduction du chapitre . . . . .	76
4.1.1	Objectifs . . . . .	76
4.1.2	Étude bibliographique des modèles bottom-up . . . . .	76
4.1.3	Cahier des charges du simulateur de courbes de charge de la thèse . . . . .	77
4.2	Description des modèles . . . . .	78
4.2.1	Secteur résidentiel . . . . .	78
4.2.2	Secteur tertiaire . . . . .	87
4.2.3	Modèle de chauffage et de climatisation . . . . .	89
4.2.4	Bilan des données d'entrée nécessaires pour les modèles et leur influence . . . . .	94
4.3	Structure et implémentation du simulateur . . . . .	95
4.3.1	Partie "cœur de calcul" . . . . .	96
4.3.2	Étapes de la simulation . . . . .	96
4.4	Conclusion et perspectives du chapitre . . . . .	99
<b>5</b>	<b>Validation et calibration du modèle</b>	<b>101</b>
	Résumé du chapitre . . . . .	101
5.1	Introduction du chapitre . . . . .	102
5.2	État de l'art de l'évaluation et la calibration des modèles bottom-up . . . . .	102
5.3	Validation et calibration des consommations en énergie du secteur résidentiel . . . . .	104
5.4	Validation et calibration sur les départs HTA d'Enedis . . . . .	108
5.4.1	Introduction . . . . .	108
5.4.2	Validation et calibration des parties non thermosensibles . . . . .	108
5.4.3	Calibration des parties thermosensibles . . . . .	113
5.5	Conclusion et perspectives du chapitre . . . . .	116
<b>6</b>	<b>Simulations prospectives pour l'aide à la planification</b>	<b>119</b>
	Résumé du chapitre . . . . .	119
6.1	Introduction du chapitre . . . . .	120
6.1.1	Contexte du dimensionnement du réseau de distribution . . . . .	120
6.1.2	Méthode de planification du réseau d'Enedis . . . . .	120
6.2	Utilisation des outils de simulation et scénarisation . . . . .	123
6.2.1	Utilisation du simulateur de courbes de charge . . . . .	123
6.2.2	Utilisation du simulateur de production renouvelable . . . . .	127

6.2.3	Bilan sur la démarche de simulations prospectives . . . . .	128
6.3	Cas d'étude à Brest . . . . .	129
6.3.1	Identification de la zone à simuler et données disponibles . . . . .	129
6.3.2	Association de la base Enedis et INSEE . . . . .	130
6.3.3	Évaluation et calibration du simulateur de courbes de charge sur les mesures	130
6.3.4	Mise en place de scénarios prospectifs . . . . .	131
6.3.5	Résultats des simulations et calculs des indicateurs . . . . .	131
6.4	Cas d'étude sur l'évolution des critères dimensionnant selon le taux de pénétration des énergies renouvelables . . . . .	134
6.4.1	Introduction . . . . .	134
6.4.2	Identification de la zone à simuler et données disponibles . . . . .	135
6.4.3	Association de la base Enedis et INSEE . . . . .	136
6.4.4	Évaluation et calibration du simulateur de courbes de charge sur les mesures	136
6.4.5	Scénarisations . . . . .	137
6.4.6	Critère de dimensionnement et influence du taux de pénétration d'énergies renouvelables . . . . .	137
6.5	Conclusion et perspectives du chapitre . . . . .	140
<b>7</b>	<b>Conclusion et perspectives</b>	<b>141</b>
7.1	Synthèse des travaux effectués . . . . .	141
7.2	Limites des travaux et perspectives notables . . . . .	143
	<b>Bibliographie</b>	<b>154</b>
<b>8</b>	<b>Annexes</b>	<b>155</b>
8.1	Annexes du chapitre 3 sur les bases de données . . . . .	155
8.1.1	Données Enedis . . . . .	155
8.1.2	Algorithme d'association des adresses postales . . . . .	157
8.1.3	Tableau des erreurs de prévision de la consommation annuelle par IRIS . . .	158
8.1.4	Tableau des erreurs de prévision de la thermosensibilité selon les différents algorithmes d'association des bases de données . . . . .	160
8.2	Annexes du chapitre 4 sur le modèle de consommation électrique . . . . .	166
8.2.1	Consommation et puissance appelée des principaux usages résidentiels . . .	166
8.2.2	Implémentation du cœur de calcul C++ . . . . .	168
8.3	Annexes du chapitre 5 sur la validation et la calibration du modèle . . . . .	169
8.3.1	Calcul des consommation en énergie par usage de références pour le secteur résidentiel . . . . .	169
8.3.2	Article de journal sur la méthode pour décomposer les départs HTA en profils élémentaires à partir des données d'Enedis . . . . .	169



# Table des figures

1.1	Illustration de la structure du réseau électrique français . . . . .	20
2.1	Les modèles top-down par rapport aux modèles bottom-up [86] . . . . .	24
2.2	Schéma illustratif de la consommation électrique telle qu'elle est abordée dans la thèse : des sources de puissance et des charges dont les consommations s'additionnent.	25
2.3	Exemple de consommation électrique agrégée à l'échelle de la France pour une semaine de janvier 2017 (RTE) . . . . .	29
2.4	Moyenne hebdomadaire des puissances appelées à température de référence par usage pour la France [127] . . . . .	30
2.5	Moyenne hebdomadaire des puissances appelées à températures de référence par type de clients pour la France [127] . . . . .	30
2.6	Puissances horaires appelées pendant une semaine de janvier pour différents types d'appareils résidentiels en France [127] . . . . .	31
2.7	Exemple de courbe de charge individuelle issue d'un compteur linky (Enedis) . . . . .	31
2.8	Exemple de différentes courbes de charge d'un lave-linge [27] . . . . .	32
2.9	Résumé des différents modèles de prédiction, à différents horizons, avec les principaux facteurs influents [135] . . . . .	35
2.10	Consommation horaire par catégorie de clients, janvier 2010 [34] . . . . .	37
2.11	Méthode de classification des clients [150] . . . . .	37
2.12	Illustration de la regression liant la consommation d'électricité à la température [102]	40
3.1	Illustration d'une zone et des clients desservis par un départ HTA . . . . .	49
3.2	Exemple de consommation électrique d'un départ HTA sur 4 ans, avec différents pas de temps . . . . .	50
3.3	Exemple de tracé des données brutes d'un départ HTA . . . . .	50
3.4	Évolution de la consommation électrique résidentielle par IRIS en fonction de chacune des variables extensives . . . . .	60
3.5	Illustration des différentes étapes de l'algorithme de résolution approché du problème d'association [39] . . . . .	66
3.6	Évolution de la thermosensibilité par départ HTA en fonction de chacune des variables extensives obtenues par la méthode de référence . . . . .	68
3.7	Évolution de la thermosensibilité par départ HTA en fonction de chacune des variables extensives obtenues par la méthode d'optimisation . . . . .	69
4.1	Principaux appareils résidentiels placés par leur puissance pic appelée en fonction de l'énergie moyenne annuelle consommée par foyer équipé . . . . .	79
4.2	Tracé de la courbe de charge d'une machine à laver modélisée par blocs de puissance constante, lancée deux fois dans la même journée . . . . .	80

4.3	Distribution relative des puissances appelées de 85 lave-linges suivis pendant 24 jours [33] . . . . .	81
4.4	Consommation annuelle par habitant de l'appareil lave-linge en fonction de la taille de la famille [33] . . . . .	81
4.5	Profils horaire et hebdomadaire des machines à laver mesurées pendant la campagne de mesures ADEME en 2016 [32] . . . . .	82
4.6	Distribution relative des puissances appelées de 16 réfrigérateurs suivis pendant 22 jours [33] . . . . .	84
4.7	Evolution de la consommation par rapport au nombre de personnes pour le réfrigérateur [33] . . . . .	84
4.8	Histogramme des consommations annuelles des réfrigérateurs [32] . . . . .	85
4.9	Profils horaire et hebdomadaire des réfrigérateurs mesurés pendant la Campagne de mesures ADEME en 2016 [32] . . . . .	85
4.10	Profil journalier de l'éclairage des bureaux [30] . . . . .	88
4.11	Profil journalier des usages de bureautique [31] . . . . .	88
4.12	Courbe de charge d'un jour ouvrable de janvier pour un parc d'un million de véhicules électriques [127] . . . . .	89
4.13	Schéma du modèle thermique utilisé par maison, sous forme d'analogie électrique . . . . .	90
4.14	Différentes positions possibles d'un client dans un bâtiment . . . . .	91
4.15	Équivalence entre le niveau de performance et les valeurs de conductivité thermique . . . . .	92
4.16	Équivalence entre l'année de construction et le niveau de performance . . . . .	92
4.17	Extrait de la réglementation Th-I 2005 : correspondance entre la classe d'inertie, la capacité de l'habitation par unité de surface $C_m/Aniv$ , et $A_m/Aniv$ [23] . . . . .	93
4.18	Récapitulatif des caractéristiques du modèle de chauffage . . . . .	94
4.19	Influence de la consommation de chaque usage en fonction du nombre de personnes . . . . .	95
4.20	Influence du nombre de la consommation de chaque usage en fonction de la surface . . . . .	95
4.21	Illustration des attributs de l'objet "House" . . . . .	96
4.22	Différentes étapes de la simulation . . . . .	97
4.23	Illustration de la manière avec laquelle est calculée l'énergie en sauvegardant chaque lancement d'appareils et en les parcourant avec la grille temporelle . . . . .	98
4.24	Exemple de résultats de simulation de 500 clients résidentiels. Dans le cadre supérieur est représentée la consommation électrique ramenée par client, décomposée selon les grandes catégories d'usages. Dans le cadre inférieur figurent les conditions météorologiques . . . . .	98
5.1	Répartition de la consommation d'électricité d'un ménage français (hors chauffage, eau chaude et cuisson) [7] . . . . .	106
5.2	Diagramme illustrant la méthode de référence pour la simulation de la partie non-thermosensible des courbes de charge de départ HTA d'Enedis [81] . . . . .	109
5.3	Evolution des NRMSE par départ avec la consommation moyenne non thermosensible	111
6.1	Exemple de courbes de charge reconstruites à partir des profils-types des clients et de leurs caractéristiques (Source : Enedis) . . . . .	122
6.2	Exemple illustratif d'une simulation d'un fort taux de pénétration d'énergie renouvelable sur un départ HTA où la production peut être dimensionnante. Simulation d'un départ HTA à Lyon avec 871 clients BT, 11 clients importants, 1,5 MW de puissance installée en photovoltaïque et 1.5 MW en éolien . . . . .	123
6.3	Estimation de la consommation annuelle par ménage équipé selon la technologie [127]	126

6.4	Démarche globale de simulations prospectives développée pendant la thèse . . . . .	128
6.5	Tracé de trois semaines représentatives d'un départ HTA à erreur médiane . . . . .	132
6.6	Tracé de trois semaines représentatives d'un départ HTA à erreur dans le dernier quartile . . . . .	132
6.7	Tracé sur une semaine de la consommation simulée d'un départ HTA avec la décomposition des consommations par postes HTA/BT . . . . .	133
6.8	Fréquence de dépassement de la puissance de référence d'un poste HTA/BT selon l'année météorologique [98] . . . . .	133
6.9	Monotone de puissance dans un départ HTA avec des clients majoritairement résidentiels . . . . .	134
6.10	Tracé des deux semaines médianes des deux départs HTA du cas d'étude de la région de Lyon . . . . .	137
6.11	Évolution du gain d'un critère de dimensionnement selon le taux de pénétration d'énergies renouvelables [124] . . . . .	139
8.1	Illustration de la normalisation des relevés de compteur client (Enedis). $En$ est la consommation normalisée, les points bleus sont des relevés de compteurs, DJ représentent les degrés jours . . . . .	156
8.2	Illustration de l'algorithme d'association des adresses clients et adresses IGN de la BD Adresse . . . . .	157
8.3	Illustration de l'implémentation C++ du cœur de calcul du simulateur de courbes de charge . . . . .	168



# Liste des tableaux

2.1	Différentes classifications des niveaux d'agrégation de la consommation électrique . . .	28
2.2	Notations utilisées pour les facteurs influents . . . . .	34
2.3	Utilisation des données par les modèles basés sur les consommations agrégées d'une zone . . . . .	36
2.4	Utilisation des données par les modèles basés sur les consommations par type de clients utilisant les données de mesure individuelle . . . . .	39
2.5	Utilisation des données par les modèles basés sur les consommations par type de clients utilisant les données de mesure par zone . . . . .	40
2.6	Utilisation des données par les modèles basés sur les consommations par usage . . .	41
2.7	Utilisation des données par les modèles basés sur les consommations par type de clients et par usage . . . . .	42
2.8	Utilisation des données par les modèles basés sur les consommations par client . . .	43
2.9	Utilisation des données par les modèles basés sur les consommations par appareil avec identification non intrusive des appareils . . . . .	44
2.10	Utilisation des données par les modèles bottom-up . . . . .	45
2.11	Utilisation des données par les modèles basés sur les consommations par type de clients et par usage . . . . .	46
3.1	Disponibilité des données de mesure de départ HTA pour chaque année . . . . .	51
3.2	Classification des types d'utilisateur du réseau de distribution par Enedis [117] . . .	51
3.3	Informations disponibles selon les types d'utilisateurs du réseau d'Enedis . . . . .	52
3.4	Différentes campagnes de mesures utilisées pour estimer les appareils présents dans les différentes catégories de clients tertiaires . . . . .	56
3.5	Provenance des données permettant de renseigner les facteurs influant, ainsi que leur échelle de disponibilité . . . . .	58
3.6	Remise à l'échelle nationale des résultats de la figure 3.4 pour comparaison des ordres de grandeurs . . . . .	61
3.7	Amélioration de l'erreur de prédiction de la consommation par IRIS en utilisant différentes variables intensives couplées à la Surface par rapport au modèle de référence (surface par IRIS) . . . . .	63
3.8	Amélioration de l'erreur de prédiction de la consommation par IRIS en utilisant différentes variables intensives couplées à la Surface par rapport au modèle de référence (surface par IRIS) . . . . .	64
3.9	Remise à l'échelle nationale des résultats de la figure 3.7 pour comparaison des ordres de grandeur. . . . .	67

3.10	Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées au nombre de personnes, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par méthode de référence et par méthode d'optimisation . . .	71
3.11	Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées aux différentes variables intensives, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par méthode de référence et par méthode d'optimisation . . . . .	72
4.1	Récapitulatif des variables utilisées dans le modèle par bloc de puissance constante et comment celles-ci sont déterminées en lien avec les bases de données . . . . .	83
4.2	Récapitulatif des variables utilisées dans le modèle par bloc de puissance périodique et comment celles-ci sont déterminées en lien avec les bases de données . . . . .	86
4.3	Différentes campagnes de mesures utilisées pour estimer les appareils présents dans les différentes catégories de clients tertiaires . . . . .	87
4.4	Chiffres et profils retenus pour modéliser les différents usages pour les bureaux . . .	88
5.1	Différents modèles bottom-up de la littérature, données utilisées pour la calibration et méthode d'évaluation . . . . .	104
5.2	Valeurs retenues pour la calibration de chaque usage du résidentiel . . . . .	105
5.3	Surface et nombre moyen de personnes par type de logement [22] . . . . .	106
5.4	Valeurs, méthode retenue pour la calibration des usages et résultats obtenus à l'issue de la calibration . . . . .	107
5.5	Erreurs quadratiques moyennes normalisées (NRMSE) (en %) du modèle de référence [81] pour les 3 différentes zones de simulation sur les 4 années de mesure en fonction du nombre de catégories. Les catégories sont issues de la base d'Enedis. La simulation a été effectuée 100 fois, et les NRMSE ont été reportées ainsi que la variance entre parenthèses . . . . .	110
5.6	Erreur quadratique moyenne normalisée entre la partie thermosensible simulée et celle mesurée. Les différents quantiles sont en colonne, les zones de simulation en ligne. Pour la moyenne et l'écart type, les valeurs entre parenthèses correspondent aux valeurs de référence du tableau 5.5 pour l'année 2012 avec 12 catégories . . . . .	111
5.7	Distribution des erreurs relatives dans la simulation de l'énergie annuelle non thermosensible (%) . . . . .	112
5.8	Résultats de la régression de la puissance souscrite par catégorie de clients en fonction de la thermosensibilité par départ, sur les trois zones de disponibilité des données, pour l'année 2012 (le $R^2$ de la régression est de 0.899) . . . . .	114
6.1	Caractéristiques des 10 départs HTA étudiés . . . . .	129
6.2	Récapitulatif des différentes caractéristiques des logements des 39 IRIS du cas d'étude de la ville de Brest . . . . .	129
6.3	Récapitulatif des différentes caractéristiques des logements des 10 départs HTA du cas d'étude de la ville de Brest . . . . .	130
6.4	Résultats de l'évaluation de la partie non thermosensible des départs HTA du cas d'étude . . . . .	131
6.5	Caractéristiques des 2 départs HTA étudiés . . . . .	135

6.6	Récapitulatif des différentes caractéristiques des logements des 5 IRIS du cas d'étude de la ville de Lyon . . . . .	135
6.7	Récapitulatif des différentes caractéristiques des logements des 2 départs HTA du cas d'étude de la ville de Lyon . . . . .	136
6.8	Résultats de l'évaluation de la partie non thermosensible des deux départs HTA du cas d'étude de la région de Lyon . . . . .	136
8.1	Descriptif des informations disponibles dans la base des clients . . . . .	156
8.2	Amélioration de l'erreur de prédiction de la consommation par IRIS en utilisant différentes variables intensives couplées au nombre de personnes par rapport au modèle de référence (surface par IRIS) . . . . .	158
8.3	Amélioration de l'erreur de prédiction de la consommation par IRIS en utilisant différentes variables intensives couplées au nombre de pièces par rapport au modèle de référence (surface par IRIS) . . . . .	159
8.4	Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées au nombre de personnes, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode d'optimisation en utilisant ou non les données des logements avec chauffage électrique . . . . .	160
8.5	Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées au nombre de personnes, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode de référence en utilisant ou non les données des logements avec chauffage électrique . . . . .	161
8.6	Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées à la surface, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode d'optimisation en utilisant ou non les données des logements avec chauffage électrique . . . . .	162
8.7	Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées à la surface, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode de référence en utilisant ou non les données des logements avec chauffage électrique . . . . .	163
8.8	Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées au nombre de pièces, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode d'optimisation en utilisant ou non les données des logements avec chauffage électrique . . . . .	164
8.9	Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées au nombre de pièces, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode de référence en utilisant ou non les données des logements avec chauffage électrique . . . . .	165
8.10	Principales caractéristiques des usages résidentiels. . . . .	167
8.11	Tableau récapitulatif des données CEREN [22] utilisées . . . . .	169
8.12	Nombre de logements par type [22] . . . . .	169

## Liste des notations

	Symbole	Descriptif	Unité
<b>A</b>			
	<i>a</i>	Appareil donné	-
	<i>Ap</i>	Ensemble d'appareils	-
	<i>Ab</i>	Facteurs sur l'abonnement des clients	-
	<i>ACHL</i>	Période regroupée d'achèvement de la construction de la maison ou de l'immeuble (base INSEE)	-
<b>B</b>			
	BT	Basse tension (<1000V)	-
	BD	Base de données	-
<b>C</b>			
	<i>c</i>	Catégorie d'appareils	-
	<i>C</i>	Ensemble de catégories d'appareils	-
	<i>CATL</i>	Catégorie de logement (base INSEE)	-
	<i>CHFL</i>	Chauffage central du logement (électrique ou non)	-
	<i>CMBL</i>	Combustible principal du logement (électrique ou non)	-
	<i>cdc(i)</i>	Courbe de charge (individuelle)	-
	<i>Co</i>	Facteurs comportementaux	-
<b>D</b>			
	<i>D</i>	Ensemble de zones	-
<b>E</b>			
	ECS	Eau chaude sanitaire	
<b>G</b>			
	GDO	Gestion des ouvrages (Enedis)	-
	GRD	Gestionnaire du réseau de distribution	-
	GRT	Gestionnaire du réseau de transport	-
<b>H</b>			
	<i>h</i>	Horizon temporel	s
	HTA	Haute tension de type A (1000V<HTA<50 000V)	-
	HTB	Haute tension de type B (<50 000V)	-
<b>I</b>			
	<i>i</i>	Nombre entier	-
	<i>IC</i>	Immeuble collectif	-
	<i>INPER</i>	Nombre de personnes du ménage (base INSEE)	-
	<i>ir</i>	Zone d'un IRIS	-
	<i>Ir</i>	Ensemble des IRIS	-
	<i>IRIS</i>	îlot regroupé pour l'information statistique	-
<b>J</b>			
	<i>j</i>	Client donné	-
	<i>J</i>	Ensemble de clients	-
<b>K</b>			
	<i>k</i>	Catégorie de clients	-
	<i>K</i>	Ensemble de catégories de clients	-

	Symbole	Descriptif	Unité
<b>L</b>			
	$L$	Facteurs sur les locaux	-
<b>M</b>			
	$m$	Nombre entier	-
	$Ma$	Facteurs liés aux marchés de l'électricité	-
	$Me$	Facteurs liés à la météorologie	-
	$MI$	Maison individuelle	-
<b>N</b>			
	$n$	Nombre entier	-
	NRMSE	Erreur quadratique moyenne normalisée	-
<b>P</b>			
	$P$	Puissance électrique active	kW
	$Pa$	Facteurs sur le parc d'appareils	-
	$P^{*max}$	Puissance mesurée le jour le plus chargé de l'année	kW
	$P_{ref}$	Puissance de référence	kW
	$Psous$	Puissance souscrite	kVA
	$P_{tmb}$	Puissance estimée à la température minimale de base	kW
<b>S</b>			
	$S$	Facteurs socio-économiques	-
	$S_{conso}$	Scénario de consommation électrique	-
	$S_{prod}$	Scénario de production renouvelable	-
	$STOCD$	Statut d'occupation détaillé du logement (base INSEE)	-
<b>T</b>			
	$t$	Temps	s
	$\Delta t$	Pas de temps	s
	$T$	Température de l'air extérieur	°C
	$Ta$	Facteurs technologiques des appareils	-
	$t_{dep}$	temps de dépassement	s
	$Th^z$	Thermosensibilité d'une zone $z$	kW/°C
	$T_{mb}$	Température minimale de base	°C
	$T_n$	Température normale	°C
	$T_s^z$	Température seuil pour la thermosensibilité d'une zone $z$	°C
	$TYPL$	Type de logement (base INSEE)	-
<b>Z</b>			
	$z$	zone géographique	-
	$\sim$	est fonction de	-
	$\sim_l$	est fonction linéaire de	-
		sachant que	-



# Remerciements

De nombreuses personnes ont participé et ont interagi dans ce projet de doctorat. L'idée ici n'est pas d'en dresser une liste exhaustive, mais plutôt de souligner quelques interactions marquantes.

Un grand merci aux deux rapporteurs de ma thèse, Bruno François et Hamid Ben Ahmed, ainsi qu'aux examinateurs, Bernard Multon, Elena Magliaro, Eric Peirano et les invités David Chassin, très matinal et en direct de Stanford et Nicolas Kong.

Je remercie chaleureusement mon directeur de thèse, François-Pascal et mon maître de thèse Robin, pour tout ce qu'ils m'ont apporté durant ces trois ans de thèse. Cela a été un plaisir de travailler avec vous, et je souhaite un jour que l'on puisse travailler à nouveau ensemble.

Merci à toute l'équipe du centre PERSEE et les collègues de Mines ParisTech. Plus particulièrement mes collègues de bureau, Etta, Lucia, Ghislain, Adrian, votre compagnie a été fort agréable. Merci à Georges pour ses relectures pointues. Un remerciement particulier à Evgeniya pour son aide et son écoute bienveillante et bienvenue en fin de thèse ainsi qu'à Alexis pour tous nos échanges intéressants. Merci particulier à Philippe pour sa gentillesse, sa disponibilité et son aide pertinente, merci à Patrick entre autre pour ses bonnes idées, et sa passion des réparations écologiques.

Merci à l'équipe d'Enedis, et notamment Nicolas, cela a été toujours agréable de travailler avec toi et l'équipe.

Merci à Eman, Nadine, Camille, pour avoir été là pendant mon expérience américaine et avoir contribué à la rendre autant agréable et enrichissante.

Merci à toute ma famille pour leur présence constante et leur aide sur tous les aspects, ma mère, mon père, mon frère, mes cousines, mes grands-parents, Tonton et Tatie. Merci à mes amis, Mélodie, Mathilde, Roméo, Thibaud, Flora, Françoise, Alexandre, Karine, Laurent, Laureen, Julie, Romain, Sélène, Sarah, Erwan, Elsa, Marine, Pedro, Gaëtan, Pierre, Olivier et d'autres qui ont participé à me faire grandir et me découvrir. Merci infini à eux.



# Chapitre 1

## Introduction

### 1.1 Contexte et enjeux de la thèse

Dans les pays développés, l'électricité répond à une grande partie de nos besoins principaux, si bien qu'on la considère comme essentielle dans nos sociétés : il n'y a presque aucun usage qui ne se passe d'électricité.

Rendre disponible l'électricité à l'échelle d'un territoire passe par l'électrification de celui-ci. Cela signifie mettre en relation des moyens de production d'électricité avec les usages finaux, via le réseau électrique. Ce dernier est crucial car il assure non seulement la circulation du flux d'énergie mais permet à la collectivité desservie de disposer de nombreux services. Il permet notamment des économies d'échelle, de profiter de l'effet du foisonnement (c'est-à-dire la diminution des fluctuations aléatoires avec le grand nombre de contributeurs et l'abaissement des puissances maximales moyennes), tout en accroissant la sûreté et la qualité de l'approvisionnement [65].

Le réseau est constitué de différents ouvrages (ou organes, *e.g.* lignes électrique, disjoncteurs, transformateurs), qui ont des coûts d'investissements importants et une longue durée de vie (souvent plus de quarante ans) [51]. Ces ouvrages, de façon optimale pour la collectivité, doivent être installés en adéquation avec le développement et l'évolution du territoire, en minimisant les coûts d'investissement et d'exploitation. C'est l'objet de la planification, majoritairement effectuée par les gestionnaires de réseau, qui consiste dans son principe général à "rechercher l'optimum économique sous contraintes, dans le cadre des politiques définies" [51].

Historiquement, le réseau électrique français s'est construit de façon centralisée après la guerre, en même temps que le développement des usages électriques et des centrales de production de taille croissante. Cela a favorisé l'émergence d'un système d'interconnexion du réseau de transport de très grande taille, d'abord en France, puis en Europe. Le réseau de distribution quant à lui achemine l'énergie de l'amont vers l'aval [65]. Ainsi globalement, dans le réseau de distribution français actuel, la demande électrique est imposée en aval par les clients et le flux de puissance provient du réseau de transport, situé en amont. La connaissance et la prévision à court et long terme de la consommation électrique de chaque client est essentielle, car c'est elle qui impose le flux de puissance dans les différents ouvrages du réseau, ainsi que la mobilisation des moyens de production. La figure 1.1 illustre la structure traditionnelle du réseau électrique français : les moyens de production centralisés et les clients haute tension de type B ( $HTB > 50\text{kV}$ ) sont connectés au réseau de transport, opéré par RTE. Le réseau de distribution, connecté au réseau de transport par l'intermédiaire du poste source, assure la desserte des clients haute tension de A ( $1\text{kV} < HTA < 50\text{kV}$ ), et des clients basse tension ( $BT < 1\text{kV}$ ) par l'intermédiaire des postes HTA/BT.

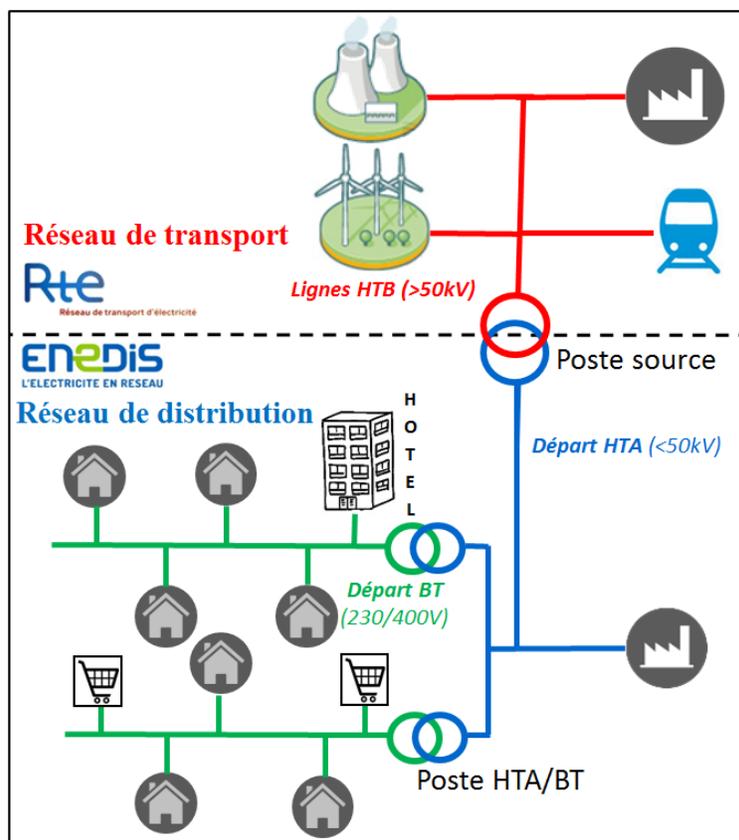


FIGURE 1.1 – Illustration de la structure du réseau électrique français

A l'échelle de la France, la consommation électrique au niveau du réseau de transport est mesurée et bien connue depuis des décennies. Chaque jour, RTE prévoit la consommation nationale pour le lendemain avec des taux d'erreur très faibles [19]. L'exercice de prévision est rendu plus difficile sur le long terme : la consommation électrique future dépend de nombreux aléas difficilement prévisibles. Malgré les difficultés, RTE effectue des modélisations à partir de scénarios les plus probables de la consommation future, en se basant sur les tendances passées, couplés à des scénarios de consommation prenant en compte la météorologie, les politiques énergétiques, la croissance économique et les évolutions technologiques. Ces modèles semblent robustes et adaptés pour le futur malgré les nombreuses évolutions possibles : RTE publie chaque année un bilan prévisionnel de l'équilibre offre/demande et du réseau qui intègre ces problématiques [127].

L'exercice de modélisation de la demande est plus délicat au niveau du réseau de distribution : les consommations électriques sont beaucoup plus sujettes aux variations locales, notamment selon les caractéristiques des clients (résidentiel, commercial, industriel etc.). Ce réseau peut être décomposé en 2 : le réseau haute tension de type A (HTA), avec une tension comprise entre 50kV et 1kV, et le réseau basse tension (BT), avec une tension inférieure à 1kV.

En France, c'est Enedis qui est le principal gestionnaire du réseau de distribution. Le réseau HTA est constitué de 2 247 postes sources en 2014 dans le périmètre Enedis. Chaque poste contient plusieurs départs haute tension de type A (HTA), et leurs mesures de consommation sont réalisées depuis longtemps par Enedis. Ceci permet à des modèles statistiques de prévoir la consommation de façon suffisamment robuste sur le long terme tout en gardant un niveau d'incertitude maîtrisé, en prenant en compte des aléas complexes, comme ceux climatiques, économétriques et démogra-

phiques. Cependant, ces modèles ne sont pas robustes aux changements majeurs car ils se basent majoritairement sur les tendances passées.

Quant au réseau basse tension (BT), avec ses 769 494 postes de transformation HTA/BT dans le périmètre Enedis, il n'est pas encore instrumenté, notamment à cause du coût que représenterait l'installation de mesures sur un nombre aussi important de postes [64]. Ainsi très peu de mesures sont faites, et un autre type de modèle est utilisé pour estimer la demande. Ce modèle utilise les données des relevés de compteur des clients, ainsi que des courbes de charge typiques par catégorie de clients [42]. Ce type de modèle présente l'inconvénient que les courbes typiques ne sont pas robustes à des changements importants dans la façon de consommer l'électricité.

Aujourd'hui, malgré les limites et incertitudes de ces méthodes au niveau des réseaux HTA et BT, le développement du système électrique en France permet la fourniture de l'électricité avec une haute qualité. Un indicateur illustrant cela est le temps de coupure annuel pour les utilisateurs du réseau basse tension : il est de l'ordre d'un peu plus d'une heure par an depuis 2005, d'après la commission de régulation de l'énergie (CRE) [3]. Cependant, cela ne signifie pas que cette qualité sera la même dans le futur : le système électrique est complexe, et repose sur des équilibres de chaque instant, qui ne sont jamais acquis [65]. De plus de nombreux changements sont en cours et à venir avec la transition énergétique, qui impacte grandement le réseau électrique, et particulièrement le réseau de distribution. Cette transition est amenée par de nombreux problèmes préoccupants, qu'ils soient climatiques, économiques, géopolitiques ou sanitaires. Ces changements impactent la demande électrique à tous les niveaux (comportements, moyens de consommation nouveaux ou plus efficaces, nouvelles réglementations), et aussi la structure même de l'équilibre offre/demande : le développement des énergies décentralisées et renouvelables fait que le réseau de distribution devient à la fois consommateur et producteur d'énergie, avec des incertitudes dans les injections et le contrôle de la production qui sont liées à l'intermittence de la ressource renouvelable.

Ces changements modifient considérablement la façon de penser le réseau électrique et particulièrement le réseau de distribution dont les modèles sont moins robustes que celui de transport. Cela nécessite une nouvelle approche de modélisation de la demande basée sur une modélisation fine descendant à l'échelle du client et du moyen de consommation. Car la consommation électrique regroupe une somme considérable de contributeurs, allant du moyen de consommation au comportement humain, couplée à des facteurs extérieurs, comme ceux climatiques par exemple. La paramétrisation de chaque contributeur est essentielle et peut être réalisée en se basant sur la quantité grandissante de données depuis quelques années. Leur disponibilité est un atout considérable. Ces données sont à la fois celles des mesures en puissance de la consommation électrique au niveau des lignes HTA, ce depuis des années, ainsi que des mesures issues du déploiement des compteurs individuels, initié il y a maintenant quelques années, et qui devrait s'étendre rapidement à toute la France et l'Europe d'ici 2020. Mais ce sont aussi les données de recensement des logements, des entreprises qui sont disponibles en accès libre sur le site de l'institut national de la statistique et des études économiques (INSEE). Il s'agit aussi des données de cartographie, des impôts, des données météorologiques. Ces données regroupent une quantité considérable d'informations à l'échelle nationale et sont d'une précision spatiale et temporelle fine. Ces données sont aussi très hétérogènes, ce qui constitue une difficulté pour la modélisation.

## 1.2 Problématique, objectifs, contributions.

Cette thèse part d'un besoin de modélisation plus fine de la consommation électrique qui prend en compte des scénarios futurs incertains à toutes les échelles, cela pour des besoins de planification, d'estimation de potentiels d'économies d'énergie ou de gestion de la demande. La problématique est de tirer parti des grandes quantités de données hétérogènes disponibles. Le but est de renseigner un modèle de consommation électrique capable de simuler la demande de façon fine et de pouvoir mettre en œuvre et tester l'influence de différents scénarios d'alternatives énergétiques de l'échelle du quartier aux régions. L'objectif final est de développer une méthode pour comprendre la consommation électrique actuelle à l'échelle de la France à partir de la grande quantité de données disponibles pour simuler différentes alternatives énergétiques. Cet objectif général est très ambitieux et cette thèse n'a pas pour vocation d'y répondre pleinement, mais plutôt de contribuer à ce grand chantier qui nécessite de nombreux acteurs et plusieurs années de travaux.

Les modèles de consommation électrique sont nombreux, et aucune étude bibliographique n'existe à notre connaissance sur les différents types de modèles et leurs objectifs majeurs. Ainsi une première contribution, au chapitre 2 de la thèse, est d'apporter une grille de lecture générale et de l'appliquer aux modèles de la consommation électrique de la littérature. Cela a aussi permis de positionner les travaux de la thèse. Dans le chapitre 3, un travail de recensement des bases de données disponibles à l'échelle de la France pouvant expliquer la consommation électrique est proposé. Un cas d'étude sur l'association de deux bases de données entre elles est décrit dans le détail, avec l'apport quantitatif et qualitatif de l'association, ouvrant la voie à l'association d'autres bases de données. Un modèle de consommation électrique conciliant les objectifs de planification et d'outil de génération réaliste de courbes de consommation électrique au niveau des appareils a été développé. Le modèle est adapté aux bases de données précédemment trouvées. Sa démarche de conception, de choix des modèles et de leur implémentation est décrite dans le chapitre 4. Une première démarche et des résultats d'évaluation de calibration du modèle sont ensuite présentés au chapitre 5. Enfin, ce modèle fait l'objet, au chapitre 6, d'études méthodologiques de scénarios prospectifs. Deux cas d'étude sont présentés en étudiant l'impact de scénarios sur le réseau électrique et son dimensionnement.

## Chapitre 2

# État de l'art des modèles de consommation électrique

### Résumé du chapitre

*L'introduction nous a permis de souligner les nombreux intérêts de modélisation de la consommation électrique dans le dimensionnement, la planification, le fonctionnement du réseau électrique. Dans ce chapitre, nous présentons les différents types de modèles de la consommation électrique de façon originale. Les études bibliographiques ne nous ont pas permis une classification aisée de tous les modèles de consommation électrique existants. En effet, à notre connaissance, la littérature ne recense pas dans une même étude tous les types de modèles de consommation électrique. Il s'agit notamment des modèles de prévision à court terme, à moyen et long terme, des modèles utilisant les données de mesure individuelles, des modèles d'identification et de prédiction non intrusives, des modèles de la demande électrique à différents niveaux d'agrégation spatiale et temporelle, aussi bien top-down que bottom-up. Ceci peut s'expliquer par les différences dans les données utilisées, l'échelle de simulation des modèles et leurs buts variés. Dans ce chapitre, nous proposons une grille de lecture permettant de classifier et comparer l'ensemble des modèles de consommation électrique existants entre eux. La grille repose sur des catégories d'échelles spatiales et temporelles de la consommation électrique utilisées par les différents modèles. Cette approche permet de poser les définitions, notations et hypothèses utilisées pour la thèse, de faire un premier état des lieux des différents modèles. Ce chapitre positionne l'approche de la thèse qui tend à tirer partie de chaque type de modèles et de données en allant au-delà de la segmentation existante.*

## 2.1 Introduction du chapitre

Nous avons vu dans l’introduction les nombreux intérêts de modélisation de la consommation électrique dans le dimensionnement, la planification, le fonctionnement du réseau électrique. Plus récemment, avec la transition énergétique, des problématiques comme l’estimation des potentiels d’économies d’énergie, la gestion de la demande, l’intégration des énergies renouvelables, sont apparues et sont traitées à l’aide d’outils de modélisation de la charge, ce qui mène aujourd’hui à un nombre de modèles existants conséquent. De façon générale, tous ces modèles cherchent à comprendre certains aspects de la consommation électrique pour différents buts. Ceci fait qu’il apparaît une forte segmentation des modèles dans la littérature, même dans les articles d’étude bibliographique.

Dans ce chapitre nous définissons la consommation électrique de façon adaptée au cadre de la thèse, de façon temporelle, puis spatiale en la classifiant en 6 niveaux d’agrégation. Nous définissons aussi les différents facteurs influents utiles pour résoudre les problématiques des modèles. Ensuite, nous présentons les différents modèles au regard de ces définitions et hypothèses.

Classiquement, quand on se réfère à un modèle de consommation électrique, les auteurs les classifient en deux grands types de modèles : l’approche ascendante, ou top-down, et l’approche descendante, ou bottom-up. Selon Swan et Ugursal [143], cette classification est basée sur le type données d’entrée : des données macroscopiques (parc de logements) ou microscopiques (logement par logement). Un schéma illustratif de Grandjean *et al.* [86] est donné à la Figure 2.1.

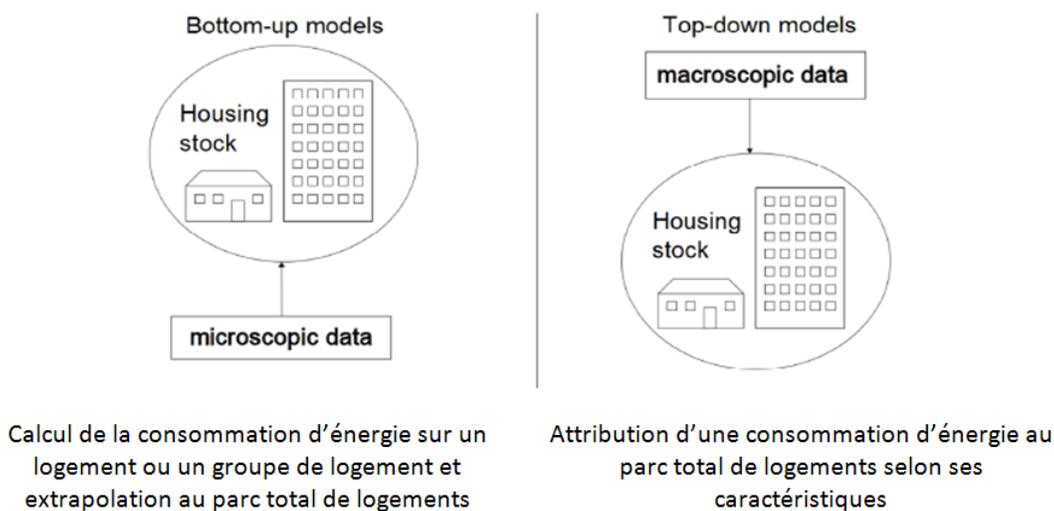


FIGURE 2.1 – Les modèles top-down par rapport aux modèles bottom-up [86]

Cette classification est adaptée pour la modélisation du secteur résidentiel, mais nous a semblé incomplète pour la modélisation électrique de façon générale. En effet, nous avons préféré parler d’échelle élémentaire de simulation permettant ainsi de classer tous les modèles de la littérature sur la consommation électrique (et pas ceux seulement résidentiels) en 6 catégories et non pas en 2. Notons que ce chapitre n’a pas pour vocation d’être exhaustif concernant les différents modèles de la littérature, mais permet de présenter une méthode nouvelle pour les décrire et les comprendre, illustrée avec les principaux modèles de la littérature à notre connaissance.

Finalement nous exprimons notre besoin de modélisation avec cette même grille de lecture et nous mettons en avant nos choix de modélisation au regard de la littérature existante.

## 2.2 Définitions, hypothèses et grille de lecture

### 2.2.1 La consommation électrique

#### Composante temporelle

Nous adoptons dans cette thèse comme définition de la consommation électrique : “une série temporelle représentant la puissance active moyenne en kilowatts ( $kW$ ) sur un certain pas de temps, ceci avec une date de début et sur une certaine durée”. Cette définition correspond aux données de consommation électrique qui sont utilisées dans la plupart des modèles.

Formellement, si l’on considère une date de début  $t_1$  et une date de fin  $t_n$ , séparés par  $n - 1$  intervalles ( $n \in \mathbb{N}^*$ ) de durée égale au pas de temps constant  $\Delta t \in \mathbb{R}_+^*$ , la consommation électrique dans cette période sera :

$$\{P(t_i)\}_{i \in \{1, 2, \dots, n-1\}} \quad (2.1)$$

avec  $P(t_i)$  la puissance électrique moyenne durant l’intervalle  $[t_i, t_{i+1}[$  (de durée  $\Delta t$ ) considéré. **Nous utilisons dans la suite  $P$  pour désigner la consommation électrique, en sous-entendant que nous avons défini une date de début, une date de fin et un pas de temps. Quand il est écrit “consommation”, sans autre précision, on se réfère à la consommation électrique.**

La composante temporelle de la consommation électrique est extrêmement importante, et selon les besoins des différents modèles cela va conditionner le choix du pas de temps, de la durée considérée, et aussi l’horizon auquel on souhaite prévoir la consommation : à court (de l’heure à la journée), moyen (du mois à l’année) ou long terme (5 à 30 ans).

#### Composante spatiale

La consommation électrique, comme toute puissance, est additive.

Dans cette thèse, nous ne prenons pas en compte les pertes sur le réseau électrique. Les pertes sur l’ensemble du réseau de transport et de distribution représentent environ 8% de la consommation intérieure nette, dont 2% sur le réseau de transport [127]. Les 6% de pertes sur le réseau de distribution, se répartissent en deux catégories : les pertes techniques, à hauteur de 3.5% (pertes joules dans les conducteurs et pertes fer dans les transformateurs), et les pertes non techniques à hauteur de 2.5% (*e.g.* imprécision ou défaut de comptage, fraudes, erreurs humaines) [116]. La modélisation des pertes fait l’objet d’études à part entière qui n’ont pas été effectuées dans le cadre de la thèse.

De façon très schématique, mais illustrative des hypothèses de la thèse, le réseau électrique peut être vu comme la mise en parallèle de sources de puissance (les moyens de productions) et de charges consommant cette puissance. Ceci est illustré en figure 2.2, dans un cas à  $n$  moyens de production et  $p$  charges.

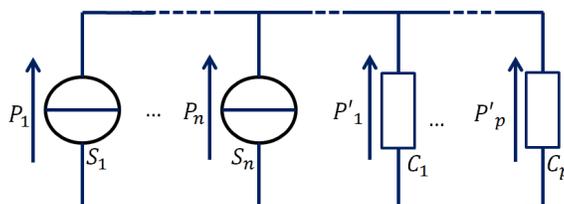


FIGURE 2.2 – Schéma illustratif de la consommation électrique telle qu’elle est abordée dans la thèse : des sources de puissance et des charges dont les consommations s’additionnent.



Si l'on pose le problème de façon formelle, en reprenant les mêmes notations qu'au 2.1 que l'on suivra tout au long de la thèse, la consommation électrique  $P^z$  d'une zone  $z$  sera :

$$P^z = \sum_{j \in J^z} (P^j) \quad (2.2)$$

Avec  $P^j$  la consommation électrique du client  $j$  et  $J^z$  l'ensemble des clients de la zone.

Chaque client  $j$  possède un ensemble d'appareils consommateurs d'électricité  $Ap^j$ . La consommation d'un client s'exprime donc par :

$$P^j = \sum_{a \in Ap^j} (P_a) \quad (2.3)$$

Avec  $P_a$  la consommation de l'appareil  $a$ .

Si l'on reprend les équations 2.2 et 2.3, on a :

$$P^z = \sum_{j \in J^z} \sum_{a \in Ap^j} (P_a) \quad (2.4)$$

Lorsqu'on mesure la consommation électrique des clients, on se rend compte de certaines similarités dans leurs profils de consommation. Par exemple, les profils de consommation des clients qui sont des commerces vont présenter une consommation plus faible le weekend qu'en semaine.

Les différents clients peuvent être regroupés en un ensemble  $K$  de différentes catégories. La consommation de la zone sera alors :

$$P^z = \sum_{k \in K} \sum_{j \in J_k^z} (P^j) \quad (2.5)$$

avec  $J_k^z$  l'ensemble des clients de la zone  $z$  appartenant à la catégorie  $k$ .

Exemple de catégories de clients :

- industrie
- agriculture
- résidentiel
  - tarif heure pleine/heure creuse
  - tarif base
  - etc.
- tertiaire
  - commerce
  - bureau
  - restaurant
  - etc.

Il en est de même pour les appareils, qui présentent des similarités selon leur type. Par exemple, la consommation d'un chauffage électrique sera beaucoup influencée par la température extérieure, l'isolation ainsi que le volume de chauffe, alors que la consommation d'une machine à laver sera dépendante du besoin de lavage de ses usagers.

On peut aussi vouloir regrouper les consommations uniquement par types d'appareils (on appelle cela aussi par usage, ou groupement de types d'appareils). Si on note  $C$  l'ensemble de tous les types

d'appareils,  $c$  un type d'appareils donné,  $Ap_c^z$  l'ensemble des appareils d'un type  $c$  de la zone  $z$ ,  $a$  un appareil donné. On a alors la consommation de la zone qui vaut :

$$P^z = \sum_{c \in C} \sum_{a \in Ap_c^z} (P_a) \quad (2.6)$$

Les types d'appareils peuvent être par exemple :

- éclairage
- chauffage
- climatisation
- eau chaude sanitaire
- production de froid
  - froid positif
  - froid négatif
- appareils dits "bruns" [62]
  - bureautique
  - audiovisuel
  - etc.
- appareils dits "blancs" [62]
  - cuisine
  - lavage
  - etc.

### Grille de lecture

Avec ces définitions et hypothèses sur la consommation électrique sur les aspects temporels et spatiaux, nous proposons dans cette partie une grille de lecture permettant de classifier la consommation électrique par son niveau de foisonnement spatial. Le tableau 2.1 récapitule les différents types d'agrégation selon les deux décompositions qui seront utilisées dans la suite : les clients et les appareils.

	Appareil agrégé	Appareil individuel
Client agrégé, spatial	Consommation électrique d'une zone $z$ : $P^z$	Consommation électrique d'une zone $z$ et d'un type d'appareils $c$ : $P_c^z$
Client agrégé par type	Consommation électrique d'un type de clients $k$ : $P_k^z$	Consommation électrique d'un type d'appareils $c$ et d'un type de clients $k$ : $P_{c,k}^z$
Individuel	Consommation électrique d'un client $j$ : $P^j$	Consommation électrique d'un appareil $a$ : $P_a$

TABLE 2.1 – Différentes classifications des niveaux d'agrégation de la consommation électrique

Chaque type d'agrégation est défini et illustré par les données disponibles dans les paragraphes suivants. Dans cette thèse, différents niveaux d'agrégation spatiale sont utilisés, et ces notations seront utilisées régulièrement pour décrire précisément le niveau spatial considéré.

**Consommation électrique d'une zone  $P^z$ .** Il s'agit de la consommation électrique mesurée sur une certaine zone spatiale. Ce peut être par exemple la consommation électrique de la France, que l'on peut noter  $P^{France}$  comme celle que fournit RTE, illustrée en figure 2.3.

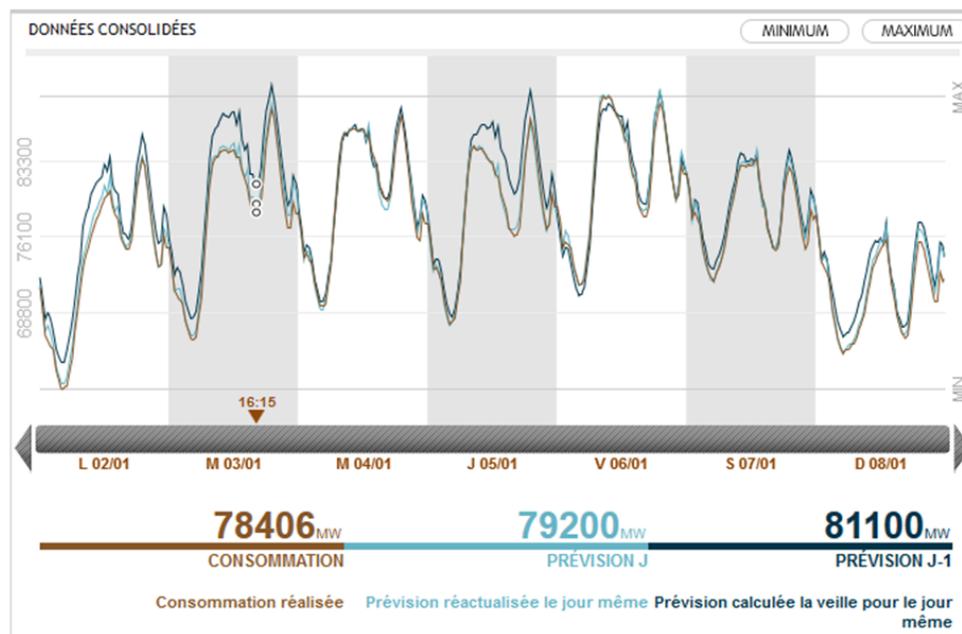


FIGURE 2.3 – Exemple de consommation électrique agrégée à l'échelle de la France pour une semaine de janvier 2017 (RTE)

Les données de mesures de consommation par grandes zones sont historiquement les plus disponibles. En effet, dès les années 1940, Hamilton [91] effectuait déjà des modèles de courbes de charge à partir de mesures agrégées.

**Consommation électrique d'une zone et d'un type d'appareils  $P_c^z$ .** C'est la consommation électrique agrégée d'une zone mais uniquement d'un type d'appareils. Elle est souvent calculée à partir d'un taux de pénétration par type d'appareils dans la zone ou, quand il s'agit d'usages thermosensibles, elle est décomposée à l'aide de données de température. Un exemple est donné en figure 2.4. On a alors :

$$P^{France} = P_{Cuisson}^{France} + P_{ECS}^{France} + P_{EP}^{France} + P_{Autre}^{France} + P_{Climatisation}^{France} + P_{Chauffage}^{France} \quad (2.7)$$

Il peut aussi d'agir de résultats de campagnes de mesures de consommation électrique d'appareils sur une zone et souvent, pour des questions de confidentialité, simplicité de visualisation, ou de stockage des données, la consommation de chaque usage est sommée ou moyennée.

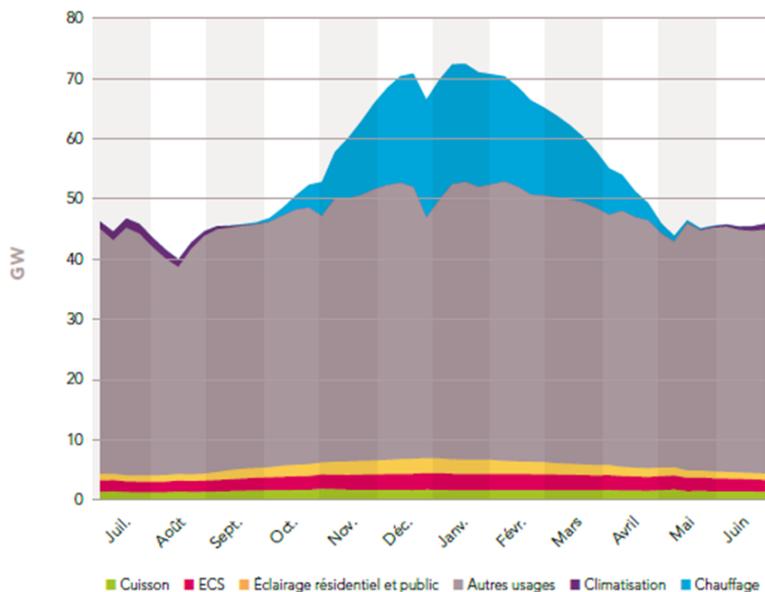


FIGURE 2.4 – Moyenne hebdomadaire des puissances appelées à température de référence par usage pour la France [127]

**Consommation électrique d'un type de clients  $P_k^z$ .** Il s'agit de la consommation électrique de différents clients uniquement du même type. Par exemple, comme c'est illustré en figure 2.5, nous avons la consommation électrique de différents types de clients en France. Sur cette figure, on a :

$$P^{France} = P_{résidentiel}^{France} + P_{tertiaire}^{France} + P_{industrie}^{France} \quad (2.8)$$

Tout comme la consommation électrique d'une zone et d'un type d'appareils, ces valeurs sont soit calculées à partir de la consommation agrégée totale en connaissant la répartition des consommations d'énergie par type de clients, soit reconstruites à partir de mesures ou modèles de clients du même type qui sont ensuite sommés.

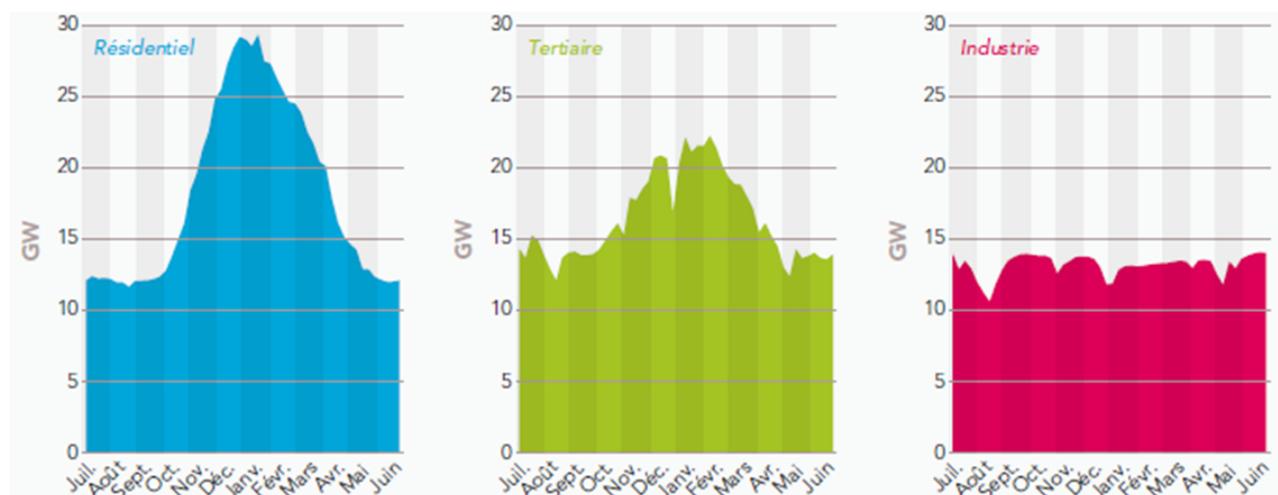


FIGURE 2.5 – Moyenne hebdomadaire des puissances appelées à températures de référence par type de clients pour la France [127]

**Consommation électrique d'un type d'appareils et d'un type de clients  $P_{c,k}^z$ .** C'est la consommation d'un type d'appareils mais seulement pour un certain type de clients. Encore une fois, elle peut être obtenue soit par des mesures et décomposition à partir de taux de pénétration et répartition des consommations, soit en mesurant la consommation électrique d'un certain nombre d'appareils du même type de clients puis en sommant. La figure 2.6 illustre ceci en prenant la consommation électrique de différents types d'appareils dans le secteur résidentiel. Il est représenté ici  $P_{résidentiel,cuisson}^{France}$  et  $P_{résidentiel,ECS}^{France}$

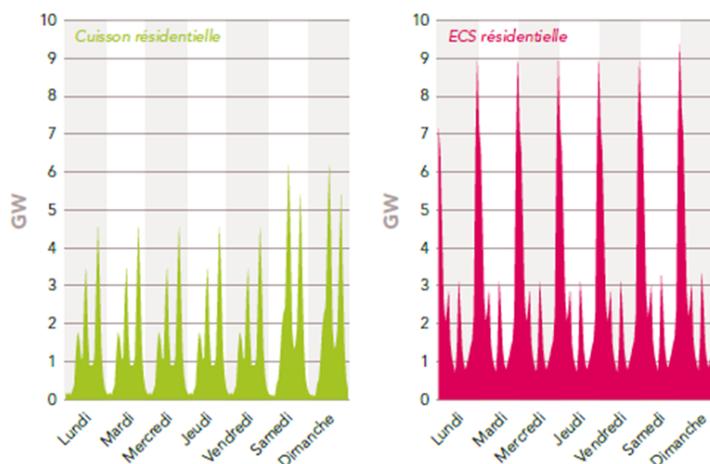


FIGURE 2.6 – Puissances horaires appelées pendant une semaine de janvier pour différents types d'appareils résidentiels en France [127]

**Consommation électrique d'un client  $P^j$ .** La consommation électrique d'un client est celle mesurée à son compteur électrique. Il peut s'agir aussi bien de consommation en énergie, relevée régulièrement pour le fournisseur d'électricité pour la facturation, ou celle en puissance, mesurée par des compteurs dits "évolués" : ces compteurs, tels le compteur linky en France, permettent de mesurer la consommation électrique à un pas de temps inférieur à l'heure. Un exemple de courbe de charge d'un client résidentiel équipé d'un compteur linky est donné en figure 2.7.

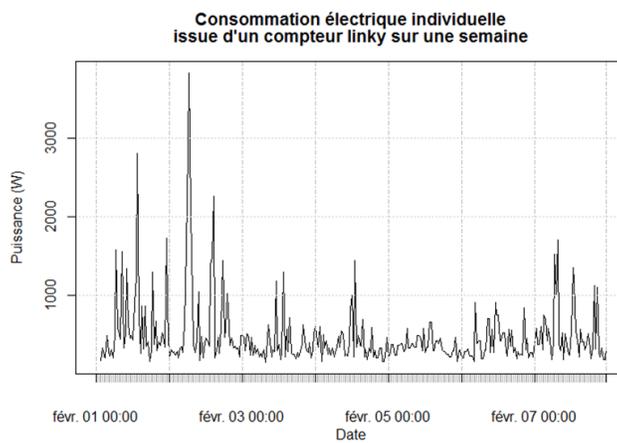


FIGURE 2.7 – Exemple de courbe de charge individuelle issue d'un compteur linky (Enedis)

**Consommation électrique d'un appareil  $P_a$ .** La consommation électrique d'un appareil peut être mesurée en instrumentant ce dernier avec un wattmètre enregistrant les données de puissance. Ce type de mesure a souvent lieu lors de campagnes de mesures où l'on vient mesurer la consommation électrique d'un grand nombre d'appareils. Un exemple de ce type de mesure est donné en figure 2.8. Il s'agit ici de  $P_{lave\_linge}$ .

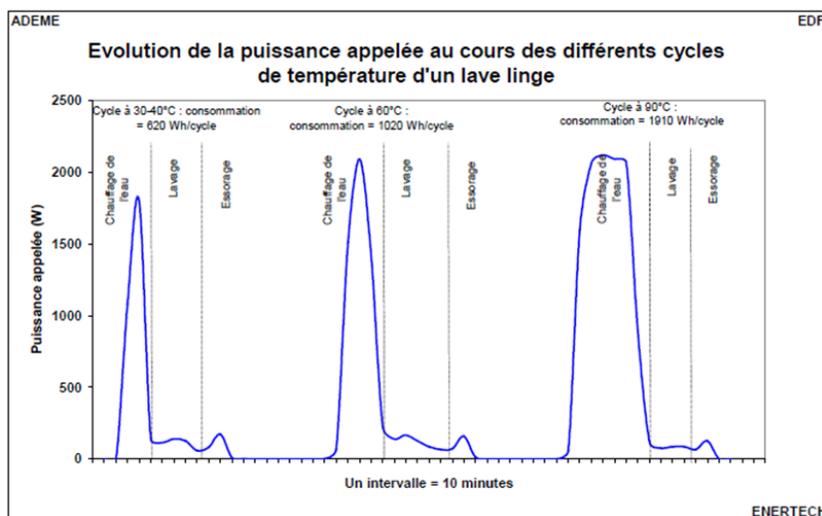


FIGURE 2.8 – Exemple de différentes courbes de charge d'un lave-linge [27]

### Conclusion sur la consommation électrique

Le premier enjeu de la consommation électrique, relatif à sa composante temporelle, est de la prévoir à un horizon défini : court, moyen ou long terme. Pour cela la compréhension des différents contributeurs est essentielle, ce que nous avons appelé sa composante spatiale. Nous avons vu qu'il existe différents niveaux d'agrégation dans la consommation électrique, et que le consommateur élémentaire est l'appareil. Pour comprendre totalement la consommation électrique, l'idéal serait donc de mesurer chaque appareil consommateur d'électricité, mais cela est impossible pour des raisons de complexité et de coût. Notons, que si l'on mesure chaque appareil ou client d'une zone, il est aisé de faire une somme pour connaître la consommation de la zone. Mais si l'on mesure uniquement la consommation d'une zone, il est impossible de trouver exactement comment un appareil ou un client a participé à la consommation électrique mesurée.

Nous avons recensé les différents enjeux des modèles :

1. temporels
  - changement de pas de temps
  - prédiction à un horizon temporel donné
  - prédiction avec différents scénarios (mode prospectif)
2. spatiaux
  - descente d'échelle à partir de données sur une zone
  - prédiction de nouvelles zones
  - généralisation d'une observation locale ou échantillonnée

Pour effectuer toutes ces tâches, des facteurs autres que la consommation électrique doivent être utilisés.

### 2.2.2 Classification des facteurs influant sur la consommation électrique

Dans cette partie nous décrivons de façon qualitative les paramètres influant sur la consommation électrique d'une zone donnée. L'étude bibliographique de Jones *et al.* [94] recense bon nombre de ces paramètres influençant la consommation électrique résidentielle avec leurs études associées. Nous synthétisons ici ces différents paramètres.

**Les appareils.** Comme nous l'avons vu précédemment, la consommation électrique d'une zone est la somme de la consommation de chaque appareil consommateur d'électricité de la zone. Il faut donc d'abord connaître le parc d'appareils de la zone : le nombre d'appareils dans la zone de chaque type, le taux de possession par type d'appareils etc.

Une fois le parc d'appareils connu, il faut connaître les paramètres technologiques de chaque appareil, ou type d'appareils : types de composants, puissance nominale, programmation, isolation d'un appareil thermique, etc.

**Les usagers.** La consommation électrique d'un appareil est fortement influencée par l'interaction usager/appareil. Elle est notamment influencée par la fréquence et la durée d'utilisation des appareils, mais aussi leur réglage (température de consignes des chauffages électriques, type de cycle d'un lave-linge etc.). Ceci est intégré dans les facteurs comportementaux des usagers (par exemple, emploi du temps). Les usagers sont décrits par des informations plus générales, comme leur nombre, l'âge moyen etc, que l'on appelle les facteurs socio-économiques.

**Les conditions extérieures.** Un nombre important d'appareils sont sensibles aux conditions extérieures. C'est le cas notamment des usages thermiques, qui vont chauffer ou refroidir différents systèmes (*e.g.* une pièce, l'intérieur d'un réfrigérateur, l'eau d'un chauffe-eau). Les échanges thermiques qu'ont ces systèmes avec l'extérieur conditionnent les appels de puissance, et donc la consommation électrique. Ces conditions extérieures sont la température extérieure, la température du local, les apports solaires, la température de l'arrivée d'eau etc.

**Les locaux.** Ce dernier aspect est surtout en lien avec la consommation électrique en chauffage ou climatisation. En effet, les caractéristiques physiques du local dans lequel est situé le chauffage ou la climatisation jouent un rôle important dans la consommation électrique : isolation, surface, inertie, présence de vitrage etc.

**Les abonnements du client.** Le client électrique possède un certain nombre d'informations sur son abonnement (puissance souscrite, type d'abonnement, énergie consommée) qui peuvent donner des informations sur les facteurs influant décrits précédemment. Ainsi par exemple, un client avec une puissance souscrite faible (*e.g.* 3kVA) n'a pas d'usages de forte puissance.

**Les marchés de l'électricité.** Ce facteur impacte plus les clients importants, qui achètent et peuvent consommer leur électricité en fonction des prix du marché (*e.g.* [153]), ce qui a donc un impact sur leur consommation. Certains abonnements pour les clients BT ont aussi des prix qui sont variables selon le type de jour (*e.g.* option effacement jour de pointe (EJP) en France) et cela influence leur consommation électrique, car le client fera en sorte de moins consommer les jours où l'électricité est onéreuse (jours dits EJP).

**Résumé des facteurs influents.** Nous avons recensé et classifié les différents facteurs influents sur la consommation électrique qui sont utilisés par les modèles présents dans la littérature, avec leurs notations que nous utiliserons par la suite :

- Les facteurs relatifs aux appareils [94]. Nous les décomposons en deux catégories : les facteurs sur le parc d'appareils notés  $Pa$  (*e.g.* nombre, taux de possession), et les facteurs technologiques notés  $Ta$  (*e.g.* puissance, type de régulation).
- Les facteurs relatifs aux usagers. Nous les avons décomposés par les facteurs comportementaux (*e.g.* emploi du temps, habitudes d'utilisation des appareils, réglage de température de consigne) [151], noté  $Co$ , et les facteurs socio-économiques (*e.g.* nombre d'occupants, revenu) [92] notés  $S$ .
- Les facteurs liés aux conditions extérieures, la météorologie (*e.g.* température, irradiation solaire) [136]. Nous les notons  $Me$ .
- Les facteurs relatifs aux locaux (*e.g.* type de local, année de construction, surface)[94]. Nous les notons  $L$ .
- Les facteurs de l'abonnement du client (*e.g.* puissance souscrite, type de clients) [109, 85]. Nous les notons  $Ab$ .
- Les facteurs liés au marché de l'électricité (*e.g.* prix) [136]. Nous les notons  $Ma$ .

Le tableau 2.2 résume les notations de ces facteurs influents que l'on utilisera dans la suite de la thèse. Nous verrons comment chaque type de modèle utilise ces données, puis par la suite comment nous les avons mises en relation avec les bases de données correspondantes pour développer le modèle de consommation électrique de la thèse.

Symbole de la variable	Signification
$Pa$	facteurs sur le parc d'appareils
$Ta$	facteurs technologiques des appareils
$Co$	facteurs du comportement d'utilisation
$S$	facteurs socio-économiques
$Me$	facteurs liés aux conditions extérieures, la météorologie
$L$	facteurs relatifs aux locaux
$Ab$	facteurs relatifs à l'abonnement du client
$Ma$	facteurs liés au marché de l'électricité

TABLE 2.2 – Notations utilisées pour les facteurs influents

## 2.3 Classification des différents modèles

Avec la grille de lecture sur la classification des niveaux d’agrégation de la consommation électrique vue au 2.2.1 ainsi que celle des facteurs influents vue au 2.2.2, nous allons décrire les différents modèles de consommation électrique de la littérature, par rapport à leur échelle élémentaire spatiale de simulation, en d’autres termes par le niveau minimal d’agrégation spatial (voir table 2.1) que le modèle est capable de fournir en sortie.

### 2.3.1 Les modèles basés sur les consommations agrégées d’une zone

Ce sont les modèles qui ont comme échelle élémentaire de simulation une consommation électrique de la forme  $P^z$ .

Nous allons parler ici des modèles de prévision que l’on appelle dans la littérature ceux de “forecasting”. Ces modèles travaillent uniquement sur la composante temporelle de la consommation électrique. Ils la considèrent comme une série temporelle non linéaire et non stationnaire, faite de la superposition de plusieurs fréquences [135].

Ces modèles se basent sur la consommation électrique agrégée à l’échelle d’une zone (allant de l’échelle du poste source à des régions ou la France), et permettent de faire des prévisions de la charge à court terme (de l’heure à la journée), à moyen terme (du mois à l’année) ou long terme (5 à 30 ans). Pour les fournisseurs d’énergie, ces prévisions sont un point clé car la consommation électrique est la donnée d’entrée la plus importante pour la planification de la production ou l’échange sur les marchés de l’électricité [85].

Ces modèles peuvent être couplés à des données météorologiques ou à l’évolution socio-économique. Ces données supplémentaires améliorent les résultats des prévisions. Comme illustré en figure 2.9, ces données supplémentaires (colonne “Leading impact factors”) dépendent des différents horizons visés par le modèle. L’évaluation de ces modèles se fait en général en calculant l’erreur entre la prévision et l’observation de la consommation agrégée, et une boucle de retour peut être aussi présente pour calibrer le modèle. RTE, le gestionnaire du réseau de transport en France (GRT), utilise par exemple ce type de modèle : “Les prévisions de consommation servent de référence pour s’assurer de l’adéquation de l’offre à la demande d’électricité à différents horizons temporels allant du pluriannuel au temps réel. Elles sont élaborées à partir de données et de prévisions météo, ainsi que des historiques des consommations” [126]. RTE réactualise ses prévisions de la consommation française pour le jour même, en calibrant le modèle par rapport aux observations actualisées.

Prediction type	The main object	Sequence pattern	Leading impact factors	Most commonly applied method
Short-term	Daily, hourly, minutely demand	Relatively stable trend; Significant cyclical fluctuation	Meteorological factors, electricity price	Neural network, SVM
Medium-term	Monthly, weekly demand	Upward trend; Significant cyclical fluctuation	Meteorological factors, economic development	Neural network, regression models, SVM, time series models
Long-term	Annual, seasonal demand	Significant upward trend; No apparently cyclical fluctuation	Economic development, macroeconomic policy	Neural network, SVM, regression models

FIGURE 2.9 – Résumé des différents modèles de prédiction, à différents horizons, avec les principaux facteurs influents [135]

Les modèles de prévision de type top-down présentés par [143] permettent eux aussi des prévisions pour la fourniture d’énergie future, et utilisent des données météorologiques, socio-économiques, les prix de l’énergie, ainsi que des informations sur les parcs de locaux et d’appareils.

On peut écrire les modèles de prévision sous la forme :

$$\{P^z(t_n + h)\} \sim (\{P^z(t_i)\}_{i \in \{1, 2, \dots, n-1\}}, Pa, S, L, Me, Ma) \quad (2.9)$$

Où l'opérateur  $\sim$  signifie "est fonction de". La fonction peut utiliser une méthode par réseaux de neurones, un modèle de régression etc. On rappelle que  $Pa$  sont les facteurs sur le parc d'appareils,  $S$  les facteurs socio-économiques,  $L$  sur les locaux,  $Me$  météorologiques et  $Ma$  des marchés de l'électricité.

Le tableau 2.3 récapitule les données utilisées par ces modèles.

type de variables du modèle	Données de consommation électrique utilisées	Données pour les facteurs influents utilisées
Entrée	Consommation électrique d'une zone $\{P^z(t_i)\}_{i \in \{1, 2, \dots, n-1\}}$	Données sur le parc d'appareils $Pa$ , socio-économiques $S$ , liées aux locaux $L$ , météorologiques $Me$ , liées aux marchés, $Ma$
Sortie	Consommation électrique d'une zone à un horizon $h$ : $P^z(t_n + h)$	-
Évaluation	Comparaison prévision et réalisation	-
Calibration	Comparaison prévision, réalisation et calibration	-

TABLE 2.3 – Utilisation des données par les modèles basés sur les consommations agrégées d'une zone

Ces modèles ont l'avantage d'être précis si la tendance d'évolution des facteurs est linéaire, mais ils ne prennent pas en compte les changements technologiques ou les nouveaux usages. De plus ils ne sont représentatifs que de la consommation de la zone d'étude, et ne permettent pas de descendre à des échelles plus fines ou des sous zones en modélisant l'effet d'agrégation entre les clients.

### 2.3.2 Les modèles basés sur la consommation par type de clients

Ce sont les modèles dont l'échelle élémentaire est la consommation par type de clients  $k$ , notée  $P_k$ .

Les courbes de charges selon les différents types de clients (*e.g.* résidentiels, commerces, industrie, agriculture) peuvent être de nature très différente [101]. La figure 2.10 illustre bien cela : on voit notamment que le secteur résidentiel présente son pic caractéristique le soir, alors que les secteurs industriels, les services privés, et agriculture ont leur pic journalier vers midi. On voit aussi la baisse de consommation électrique le weekend pour les industries et les services privés.

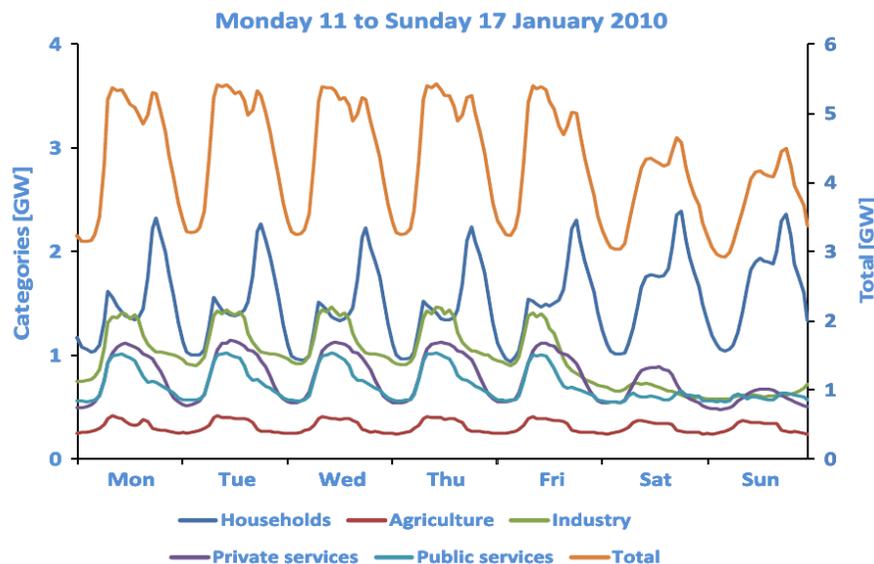


FIGURE 2.10 – Consommation horaire par catégorie de clients, janvier 2010 [34]

Ainsi de nombreux modèles utilisent comme échelle élémentaire la consommation par type de clients, car cette échelle permet de regrouper une grande diversité en un nombre restreint de profils (de l'ordre d'une dizaine [57]).

### Les modèles utilisant les données de consommations individuelles comme entrée

Pour accéder à des consommations par type de clients, de nombreux chercheurs passent par des données de consommations individuelles  $P^j$ . Le déploiement dans toute l'Europe des compteurs évolués rend de plus en plus accessible ce type de données : il est prévu que 80% des clients aient ce type de compteur installé en 2020 [132]. Cependant, ce déploiement massif est freiné par des problèmes de coût et de respect de la vie privée [107] : en 2014, seulement 23% des compteurs évolués prévus pour 2020 étaient installés [59]. Utiliser les données de mesures individuelles n'est pas toujours adapté, notamment dans certains pays où la représentativité n'est pas suffisante, ou si le déploiement est trop récent pour couvrir des périodes de mesure assez longues. Une fois ces données de consommation individuelles disponibles, leur classification en différentes catégories fait l'objet de recherches florissantes (voir les revues [101, 121]). Les chercheurs utilisent des mesures individuelles venant de compteurs évolués et appliquent différentes méthodes de regroupement (en anglais *clustering*) [150]. Une illustration de cela est donnée en figure 2.11.

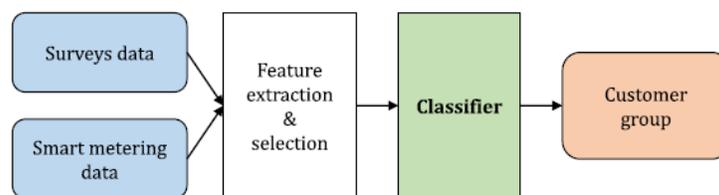


FIGURE 2.11 – Méthode de classification des clients [150]

Ainsi si l'ensemble  $K$  des catégories de clients sont exhaustives, on peut retrouver la consom-

mation de la zone à partir de la consommation de chaque catégorie :

$$P^z = \sum_{k \in K} \sum_{j \in J_k^z} (P^j) = \sum_{k \in K} (P_k^z) \quad (2.10)$$

Tout ceci permet une réduction des dimensions des données utilisées et une meilleure manipulation et interprétation [108]. Avec cette classification, chaque client est associé avec son cluster et son profil de charge correspondant [118]. La contribution de chaque type de clients dans la consommation totale agrégée de la zone peut ensuite être quantifiée [131], en connaissant la liste des clients avec des informations sur leur abonnement. Ceci permet de resimuler une courbe de charge agrégée pour pouvoir comparer les résultats et donc évaluer les performances du modèle.

Les gestionnaires du réseau de distribution (GRD) utilisent ce type de modèle pour planifier leur réseau et anticiper les investissements [109, 131]. C'est par exemple de cette manière que procède le principal gestionnaire du réseau de distribution français Enedis pour effectuer la planification de son réseau basse tension. Son modèle, appelé "Bagheera", est décrit par Ding [64]. Les GRD peuvent aussi vouloir connecter leurs lignes de façon à limiter les problèmes de déséquilibre de puissance et utilisent l'estimation des consommations à l'aide des profils des clients [56]. Des scénarios prospectifs utilisant des projections sur l'énergie consommée pour les différentes catégories peuvent aussi être utilisés pour prévoir la distribution de consommation des différentes catégories, tel que présenté par Andersen *et al.* [35]. Cependant, la classification des clients individuels a plusieurs inconvénients : elle requiert un nombre significatif de données de compteurs évolués, et les changements dans les profils sont difficiles à modéliser (*e.g.* évolution de l'occupation dans les logements) [34, 118]. Pour faire face à ces problèmes, ce sont souvent les modèles utilisant les données agrégées par zone vus au 2.3.1 qui sont utilisés pour planifier des zones à plus grande échelle.

Pour résumer, une fois les groupements en un ensemble  $K$  de catégories effectués, le problème est alors de trouver les profils-types par catégorie  $P_k$  selon les informations sur chaque client et sur la météorologie, pour estimer au mieux la consommation électrique de la zone.

$$P^z \sim (P^z, Me^z, Ab^z) | K \quad (2.11)$$

L'opérateur  $|$  désigne signifie "sachant que",  $K$  signifie l'ensemble des catégories. Pour simuler une nouvelle zone  $z'$ , les modèles utilisent les mêmes types de profils et de données, mais on peut aussi y inclure les paramètres sociaux économiques de la nouvelle zone :

$$P^{z'} \sim (P^{z'}, Me^{z'}, Ab^{z'}, S^{z'}) | K \quad (2.12)$$

Certains auteurs font de même pour prévoir sur des horizons temporels, plutôt au long terme, tel qu'effectué par exemple par Andersen *et al.* [36].

Le tableau 2.4 récapitule les données utilisées par ces modèles.

type de variables du modèle	Données de consommation électrique utilisées	Données pour les facteurs influents utilisées
Entrée	Consommation électrique individuelle $P^j$	Données météorologiques $Me$ , d’abonnement client $Ab$
Sortie	Consommation électrique par type de clients (ou profil du client) $P_k$	-
Évaluation	Consommation électrique consommée par zone $P^z$	Données socio-économiques $S$ , météorologiques $Me$ , d’abonnement client $Ab$
Calibration	Consommation électrique consommée par zone $P^z$	Données socio-économiques $S$ , météorologiques $Me$ , d’abonnement client $Ab$

TABLE 2.4 – Utilisation des données par les modèles basés sur les consommations par type de clients utilisant les données de mesure individuelle

### Les modèles utilisant les données de consommation par zone comme entrées

Au cours de cette thèse, [81], nous avons développé une méthode utilisant les données de consommation par zone comme données d’entrée. Nous prenons comme hypothèse que la demande par zone est une somme de différentes proportions de profils associés à différentes catégories de clients. Une méthode statistique utilisant les données de mesures de départ HTA ainsi que des données de description des clients (venant d’un partenariat avec Enedis) a été proposée et permet de retrouver les profils par catégorie de clients. Ces profils peuvent être utilisés pour comprendre la contribution de chaque catégorie dans la courbe agrégée. Ils permettent aussi de faire des prévisions de la consommation sur des zones non mesurées, en connaissant seulement la proportion en énergie consommée de chaque catégorie. Ce type de modèle permettant une telle génération de profils est à notre connaissance nouveau dans la littérature.

Formellement, ce modèle suppose que la consommation électrique d’une zone est :

$$P^z = \sum_{k \in K} p_k^z P_k + \epsilon^z \quad (2.13)$$

Où  $p_k^z$  est la proportion de la consommation en énergie de la catégorie  $k$  de la zone  $z$ ,  $\epsilon^z$  est un terme résiduel supposé petit. Notons que l’on suppose les différents profils par catégorie  $P_k$  indépendants de la zone. Ainsi l’enjeu dans ces travaux est de retrouver les différents profils par catégorie  $P_k$  avec une méthode d’optimisation. L’apprentissage de ces profils se fait en travaillant sur des centaines de zones, qui correspondent à des centaines de départs HTA répartis dans 3 zones géographiques françaises.

Le tableau 2.5 récapitule les données utilisées par ce type modèle.

type de variables du modèle	Données de consommation électrique utilisées	Données pour les facteurs influents utilisées
Entrée	Consommation électrique de différentes zones $P^z$	Données sur l'abonnement des clients $Ab$
Sortie	Consommation électrique par type de clients $P_k$	-
Évaluation	Comparaison sur la consommation électrique de nouvelles zones	-
Calibration	-	-

TABLE 2.5 – Utilisation des données par les modèles basés sur les consommations par type de clients utilisant les données de mesure par zone

### 2.3.3 Les modèles basés sur la consommation par usage

Un des modèles de ce type est celui décomposant les usages thermosensibles de ceux non thermosensibles à partir de courbes de charge de différentes zones. Ceci permet par exemple aux GRT et GRD de pouvoir décomposer la partie de la consommation qui dépend de la température, et celle qui n'en dépend pas et ainsi de comprendre l'influence de l'aléa météorologique dans le dimensionnement du réseau.

On définit la thermosensibilité comme la dépendance à la température de la consommation électrique. Un cas typique est illustré en figure 2.12. La consommation électrique française en fonction de la température est représentée. On voit qu'en dessous d'un certain seuil de température dit "hiver" la consommation électrique augmente quand la température diminue : ceci est majoritairement dû aux consommations des chauffages électriques des locaux qui augmentent pour garder leur température de consigne. De même, en été, quand la température devient trop haute, les climatisations se mettent en route et la consommation électrique augmente avec la température.

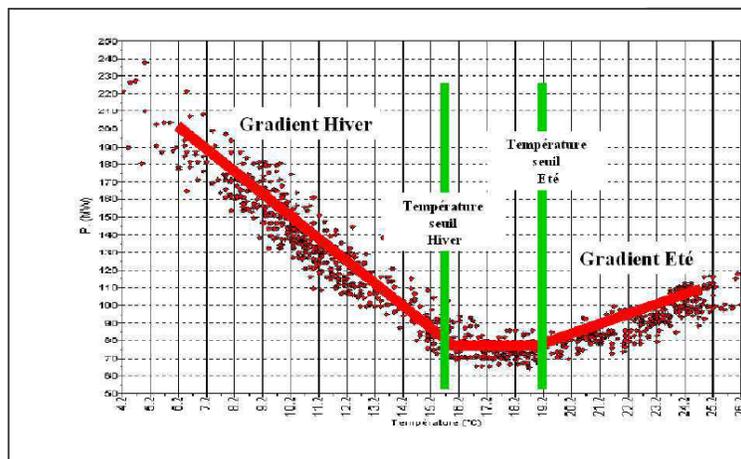


FIGURE 2.12 – Illustration de la régression liant la consommation d'électricité à la température [102]

Le cas de la France est particulièrement parlant puisqu'en hiver, la baisse d'un degré de la température moyenne nationale entraîne une croissance de la puissance appelée de 2,4 GW à 19 h

(2,4GW représente 4,4% de la moyenne de la consommation électrique française en 2015, 54GW (RTE)). A elle seule, la France représente plus de 40% de la thermosensibilité européenne [127].

De façon formelle, ces modèles supposent que :

$$P^z = \sum_{k \in K} P_k \quad (2.14)$$

Le tableau 2.6 récapitule les données utilisées par ce type de modèle.

type de variables du modèle	Données de consommation électrique utilisées	Données pour les facteurs influents utilisées
Entrée	Consommation électrique d'une zone $P^z$	Données météorologique $Me$
Sortie	Consommation électrique par usage (type d'appareils) $P_c^z$	-
Évaluation	-	-
Calibration	-	-

TABLE 2.6 – Utilisation des données par les modèles basés sur les consommations par usage

### 2.3.4 Les modèles basés sur la consommation par type de clients et par usage

Nous présentons ici les modèles du secteur résidentiel qui sont basés sur la consommation par type d'usage du secteur. Ce type de modèle regroupe les modèles présentés par Grandjean *et al.* [86] qui sont qualifiés de modèle top-down de désagrégation déterministe statistique (*Deterministic statistical disaggregation model*). Ce type de modèle consiste à désagréger les profils de consommation électrique avec des informations sur le taux de possession des appareils, et éventuellement des informations socio-économiques permettant des scénarios d'évolution de la demande, comme le modèle de Bartels *et al.* [41]. Ces modèles sont validés et calibrés sur la courbe de charge totale de la zone et du type de clients.

Finalement, ce type de modèle revient à faire l'hypothèse que

$$P_k^z = \sum_{c \in C} P_{c,k}^z \quad (2.15)$$

L'enjeu de ces modèles est de trouver les consommations des différents usages  $P_{c,k}$  :

$$P_{c,k}^z \sim (P_k^z, Pa, S, Me) \quad (2.16)$$

Des modèles similaires décrits par [143] sont basés sur la même technique, et sont classés dans leur étude dans les modèles bottom-up par distribution. La différence est qu'ils sont basés sur les consommations en énergie annuelle (donc des consommations électriques sur un pas de temps de un an) et des données à l'échelle nationale. Ils permettent à partir des taux de possession et des consommations en énergie du secteur résidentiel de retrouver la consommation par usage.

Ces types de modèle ont l'avantage d'être assez simples ; ils utilisent des données agrégées en général disponibles et les modèles sont calibrés sur des données mesurées. Cependant, ils ont l'inconvénient de ne pas pouvoir affecter de courbes de charge aux usages qui sont détenus par la majorité des logements, et il n'y a pas de différenciation dans les courbes de charge unitaire par

appareil selon le même type d'usage [86]. Globalement ils ne permettent pas de pouvoir simuler des changements majeurs de technologie ou des nouveaux usages car ils se basent sur des mesures passées.

Le tableau 2.7 récapitule les données utilisées par ce type modèle.

type de variables du modèle	Données de consommation électrique utilisées	Données pour les facteurs influents utilisées
Entrée	Consommation électrique résidentielle $P_k^z$ ou ensemble de données de consommations individuelles $\sum_{j \in J_k^z} P^j$	Données sur le parc d'appareils $Pa$ , socio-économiques $S$
Sortie	Consommation électrique résidentielle par usage $P_{c,k}^z$	-
Évaluation	Comparaison sur la consommation électrique résidentielle mesurée	-
Calibration	Consommation électrique résidentielle $P_k^z$ ou ensemble de données de consommations individuelles $\sum_{j \in J_k^z} P^j$	Données sur le parc d'appareils $Pa$ , socio-économiques $S$

TABLE 2.7 – Utilisation des données par les modèles basés sur les consommations par type de clients et par usage

### 2.3.5 Les modèles basés sur la consommation par client

Les modèles que nous avons recensés dans cette catégorie utilisent des données issues de compteurs évolués pour effectuer des prévisions de la consommation électrique sur le court terme de clients individuels, dans le but d'estimer la demande souvent dans des contextes de réseau *smart grid* ou *micro grid*. Gajowniczek et Zabkowski [80] recensent bon nombre de ce type de modèle ainsi que leurs objectifs principaux.

Certains auteurs estiment aussi l'influence de différents niveaux d'agrégation des consommations individuelles, en classifiant les clients de différentes manières, pour des services smart grid [133, 134].

Ces modèles utilisent des données individuelles de consommation électrique en apprentissage sur un modèle statistique puis comparent et calibrent le modèle avec les données sur l'horizon prédit. Ils peuvent éventuellement améliorer leurs prévisions en utilisant des données sur les facteurs technologiques des appareils dans les logements,  $Ta$  [80].

De façon plus formelle, ces modèles peuvent s'écrire sous la forme :

$$\{P^j(t_n + h)\} \sim (\{P^j(t_i)\}_{i \in 1:n-1}, Ta) \quad (2.17)$$

Le tableau 2.8 récapitule les données utilisées par ce type de modèle.

type de variables du modèle	Données de consommation électrique utilisées	Données pour les facteurs influents utilisées
Entrée	Consommation électrique de clients $P^j$	Données technologiques sur les appareils $Ta$
Sortie	Consommation électrique $P^j$	-
Évaluation	Comparaison prévision et réalisation	-
Comparaison prévision, réalisation et calibration	-	-

TABLE 2.8 – Utilisation des données par les modèles basés sur les consommations par client

### 2.3.6 Les modèles basés sur la consommation par appareil

#### Les modèles d’identification non intrusive des appareils

Ces modèles sont appelés en abrégé “NILM” en anglais pour *Non-Intrusive Load Monitoring*. Il s’agit de décomposer la consommation électrique de clients individuels par usage en utilisant des mesures individuelles de clients. Ces dernières agrègent la consommation par appareil. Le principe consiste à avoir un taux d’échantillonnage suffisant (en général la seconde ou moins) pour capturer toutes les signatures énergétiques des différents appareils. Certains modèles tels que décrits dans [61] utilisent des mesures à échantillonnage réduit qui ne caractérisent pas précisément le comportement de tous les occupants et les appareils mais permettent plutôt d’isoler les appareils les plus consommateurs. Ces méthodes permettent ainsi d’estimer la présence ou non d’appareils et leur état chez un client à partir de mesure de consommation individuelle, et cette estimation peut être aidée par des informations sur le parc d’appareils, socio-économiques ou météorologiques. Ces modèles peuvent permettre de mieux prévoir la consommation individuelle à court terme en connaissant l’état des différentes charges.

Formellement, ce type de modèle revient à :

$$P_a \sim (P^j, S, Me) \quad (2.18)$$

Comme tous les modèles qui se basent fortement sur les données passées comme apprentissage, ces modèles ne sont pas performants quand il s’agit de faire des scénarisations prospectives avec des changements technologiques importants. De plus les modèles utilisant des données à fort échantillonnage sont capables de retrouver la quasi totalité des usages mais ce type de données est très rare et la majorité des compteurs évolués déployés massivement ont des pas de temps supérieurs à la dizaine de minutes [75]. Quant aux modèles tels que présentés par Debusschere *et al.* [61] qui utilisent des données de compteurs individuels, ceux-ci ne permettent que l’identification des charges de fortes puissances et ainsi de nombreux appareils ne sont pas pris en compte.

Le tableau 2.9 récapitule les données utilisées par ce type de modèle.

type de variables du modèle	Données de consommation électrique utilisées	Données pour les facteurs influents utilisées
Entrée	Consommation électrique individuelle $P^j$	Données sur le parc d'appareils $Pa$ , socio-économiques $S$ , météorologiques $Me$
Sortie	Consommation électrique par appareil $P_a$	-
Évaluation	Comparaison sur la consommation électrique résidentielle mesurée	-
Calibration	Consommation électrique résidentielle $P_k^z$ ou ensemble de données de consommations individuelles $\sum_{j \in J_k^z} P^j$	Données sur le parc d'appareils $Pa$ , socio-économiques $S$

TABLE 2.9 – Utilisation des données par les modèles basés sur les consommations par appareil avec identification non intrusive des appareils

### Les modèles bottom up résidentiels

Dans les modèles de consommation résidentielle de type bottom-up, la modélisation est effectuée avec un très haut niveau de détails, en établissant des modèles simulant la consommation électrique pour chaque appareil. Les résultats sont en général extrapolés pour obtenir une consommation électrique qui décrit une zone géographique entière. Les données d'entrée requises pour ce type de modèles peuvent être [86] :

- Les propriétés technologiques des appareils ;
- Les données individuelles de consommation des usages domestiques ;
- Les comportements humains ;
- La géométrie et les propriétés thermiques des logements modélisés ;
- Les informations météorologiques ;
- Les factures d'électricité des clients ;

De façon formelle, ces modèles cherchent à modéliser plus ou moins finement des consommations des appareils  $P_a$  de chaque client, selon :

$$P_a \sim (Ta, Co, S, L, Me, Ab) \quad (2.19)$$

Les appareils détenus par le clients  $Ap^j$  sont déterminé avec les informations sur le parc d'appareils selon leur catégories  $C$  et éventuellement les paramètres socio-économiques et des locaux du client  $S^j$  :

$$Ap_j \sim Pa|(C, S^j, L^j) \quad (2.20)$$

Ces modèles ayant le plus petit niveau d'agrégation (consommation par appareil), leur évaluation peut se faire sur tous les types de données spatiales de consommation électrique, mais l'enjeu est de trouver les différentes données qui sont très nombreuses et doivent être renseignées pour chaque client et appareil que l'on souhaite simuler. Bien souvent, pour pallier l'absence de données précises par client, les facteurs influents sont estimés à partir de données plus générales sur la zone.

Comme le souligne Grandjean *et al.* [86], ce type de modèle ne nécessite pas de données de mesure de consommation électrique. De plus, la fonction modélisant un type d'appareils peut représenter un

modèle ayant une signification physique, permettant ainsi une interprétation des variables utilisées et une simulation de nouvelles zones ou nouveaux horizons à partir de scénarios non tendanciel. Ces scénarios devront donc être sur les facteurs influents, qui sont  $Pa$ ,  $Ta$ ,  $Co$ ,  $S$ ,  $L$ ,  $Me$ .

Le tableau 2.10 récapitule les données utilisées par ce type de modèle.

type de variables du modèle	Données de consommation électrique utilisées	Données pour les facteurs influents utilisées
Entrée	-	Données sur les appareils $Pa$ , $Ta$ , les comportements $Co$ , socio-économiques $S$ , sur les locaux $L$ , météorologiques $Me$
Sortie	Consommation électrique par appareil $Pa$	-
Évaluation	Comparaison sur des consommations individuelles $P^j$ ou sur la consommation électrique résidentielle mesurée $P^z_{résidentielle}$	Facteurs influents de la zone ou des clients $S$ , $L$ , $Ta$ , $Pa$ , $Co$ , $Me$
Calibration	Comparaison sur des consommations individuelles $P^j$ ou sur la consommation électrique résidentielle mesurée $P^z_{résidentielle}$	Facteurs influents de la zone ou des clients $S$ , $L$ , $Ta$ , $Pa$ , $Co$ , $Me$

TABLE 2.10 – Utilisation des données par les modèles bottom-up

### 2.3.7 Bilan des différents modèles de la littérature

Dans ce chapitre nous avons défini la consommation électrique ainsi que les hypothèses de la thèse. Nous avons vu les différents niveaux d'agrégation de la consommation électrique d'un point de vue spatial, permettant de classer les modèles de la littérature. Il en ressort que la littérature présente des modèles de consommation électrique à tous les niveaux d'agrégation, mais seuls les modèles basés sur la consommation par appareil de type bottom-up sont capables d'être utilisé pour faire des scénarios qui ne sont pas basé sur des tendances statistiques de la consommation électrique. Cependant, ce type de modèle est surtout développé dans le secteur résidentiel. Le secteur tertiaire, quant à lui, fait l'objet de moins de littérature d'une part, car il y a beaucoup moins de données [46]. D'autre part, c'est un secteur où les consommations sont mieux connues. En effet, le fonctionnement du secteur non résidentiel est plus centralisé et il y a un intérêt propre dans la réduction de l'énergie consommée et un plus grand nombre de réglementations [143].

Il y a un manque de modèle qui, à la fois, ait la possibilité d'effectuer des scénarios non tendanciel et puisse apprendre des données de consommation électrique par zone, en modélisant toutes les catégories de clients. Le challenge vient du manque de données sur les facteurs influençant la consommation électrique de façon aussi précise à de si grandes échelles. Ceci est en train de changer depuis ces dernières années avec les grandes quantité de données qui sont maintenant disponibles à l'échelle nationale. L'autre difficulté vient du fait de devoir évaluer et calibrer un tel modèle sur une grande quantité de données de mesures, ce qui n'est possible en France que si l'on a accès aux mesures de consommation électrique du gestionnaire du réseau de distribution, ce que peu d'acteurs ont à disposition.

## 2.4 Positionnement des travaux de la thèse

La problématique de la thèse est de modéliser la consommation électrique à l'échelle de la France à partir de grandes quantités de données disponibles pour simuler différentes alternatives énergétiques.

Dans cette thèse, l'accès aux données de consommation électrique de nombreuses zones desservies par des départs HTA a été possible grâce à un partenariat avec Enedis. Cela a permis le développement d'un modèle de consommation électrique bottom-up pour la simulation à grande échelle de la consommation de tous les types de clients.

Le modèle étant bottom-up, la consommation électrique des appareils est modélisée de même qu'à l'équation 2.19 :

$$P_a \sim (Ta, Co, S, L, Me, Ab) \quad (2.21)$$

Il tire partie de la grande quantité de données disponibles, et on a vu que certaines données étaient disponibles uniquement par catégorie de clients ou par type d'appareils. Chaque client aura donc un parc d'appareils dépendant de sa catégorie  $k$ , des différentes catégories d'appareils ainsi que de ses facteurs socio-économiques et de ses locaux :

$$Ap_j \sim Pa|(k, C, S^j, L^j) \quad (2.22)$$

Notons que, tout comme les modèles bottom-up résidentiels, nous estimons la possession des appareils du client  $j$  à partir des données sur le parc d'appareils comme vu à l'équation 2.20.

Le tableau 2.11 récapitule les données utilisées par le modèle de la thèse.

type de variables du modèle	Données de consommation électrique utilisées	Données pour les facteurs influents utilisées
Entrée	-	Données sur les appareils $Pa$ , $Ta$ , les comportements $Co$ , socio-économiques $S$ , sur les locaux $L$ , météorologiques $Me$
Sortie	Consommation électrique par appareil $P_a$	-
Évaluation	Comparaison sur des consommations de départ HTA	Facteurs influents de la zone desservie $S$ , $L$ , $Ta$ , $Pa$ , $Co$ , $Me$
Calibration	Calibration sur des consommations par type de clients $P_k$ et sur la consommation électrique mesurée des départs HTA	Facteurs influents de la zone ou des types de clients $S$ , $L$ , $Ta$ , $Pa$ , $Co$ , $Me$

TABLE 2.11 – Utilisation des données par les modèles basés sur les consommations par type de clients et par usage

En disposant des données d'Enedis de consommation par départ HTA, et en les croisant à d'autres bases de données (chapitre 3), il est ainsi possible d'avoir les données d'entrée nécessaires pour un modèle qui simule tous les secteurs d'activité. Nous verrons au chapitre 4 précisément comment la partie résidentielle du modèle développé s'insère dans la littérature des modèles bottom-up résidentiels.

## Chapitre 3

# Constitution d'une base de données

### Résumé du chapitre

*L'opportunité d'avoir une grande quantité de données est aussi un challenge, car celles-ci sont volumineuses et hétérogènes. Dans un premier temps, nous avons recensé les différentes bases existantes nous permettant d'expliquer la consommation électrique française. Les bases de données sont nombreuses, et nous en utilisons une partie à différents niveaux du modèle de consommation électrique. Une particularité de cette thèse par rapport à la littérature est d'utiliser la grande quantité de données de mesures de départs HTA d'Enedis, associée à leur base des clients. Des méthodes de manipulation, d'homogénéisation et d'association de ces données ont été développées. Une méthode d'association entre la base des logements de l'INSEE et les clients d'Enedis a été conçue et son intérêt a été validé. Nous avons ainsi constitué une base de données qui peut fournir en paramètres d'entrée un modèle bottom-up, et qui peut aussi évaluer et calibrer celui-ci.*

### 3.1 Introduction du chapitre

Nous avons vu au chapitre 2, à la section 2.4, que la thèse s'appuyait sur les données de mesure d'Enedis. Cela permet d'accéder à la consommation électrique de nombreuses zones et ainsi à une évaluation et une calibration à grande échelle du modèle développé. Cependant, le modèle étant bottom-up, il est nécessaire d'avoir des informations relatives à chaque client  $j$  et chaque appareil  $a$  des zones simulées (voir équation 2.19). Ce sont donc les facteurs sur le parc d'appareils par client  $Pa_j$  et leur technologie  $Ta_a$ , sur les comportements  $Co^j$ , socio-économiques  $S^j$ , sur les locaux  $L^j$ , et enfin la météorologie  $Me^j$ .

Ce chapitre présente d'abord les bases de données qui vont être utilisées en analysant leur disponibilité, leurs caractéristiques et leur échelle spatiale. Ensuite, un cas d'étude complet présentant l'association des données des logements et des clients est présenté. L'intérêt tant qualitatif que quantitatif de différentes méthodes d'association est développé, puis des perspectives pour les autres bases de données sont présentées. Cette nouvelle base de données issue de l'association permet de renseigner les entrées d'un modèle bottom-up à l'échelle de la France.

### 3.2 Présentation des bases de données

#### 3.2.1 Les bases de données d'Enedis

Enedis, principal GRD du réseau de distribution français, effectue des mesures sur ses lignes, ainsi que le recensement et le comptage des utilisateurs de son réseau depuis des années.

Depuis avril 2015, Enedis met à disposition un certain nombre de jeux de données en ligne [8]. Cependant ces données sont les données de consommation agrégées à la maille Enedis régionale, qui rassemble la consommation de millions de clients avec très peu d'informations sur ces derniers. Ces données sont très difficilement exploitables pour fournir en données d'entrée, évaluer et calibrer un modèle de consommation électrique de type bottom-up.

Pour cette thèse, le travail a été effectué avec des données issues d'un partenariat spécifique avec Enedis, donnant accès à un grand nombre de données de mesures de lignes et d'informations sur ses utilisateurs.

#### La base des mesures des départs HTA d'Enedis

Cette base de données permet d'obtenir les consommations par zone définies au chapitre 2. Les zones correspondent ici à celles desservies par les départs HTA. Ces données sont utiles notamment pour évaluer et calibrer le modèle de consommation électrique.

**Définitions.** Le réseau haute tension de type A (HTA) (anciennement appelé réseau moyenne tension (MT)) est instrumenté au niveau du départ des lignes par Enedis depuis des années. Déjà en 1991, une publication dans la revue "Techniques de l'ingénieur" [51] parle du fait que : "le moyen le plus direct [pour connaître la charge actuelle] consiste à effectuer des mesures sur le réseau. C'est en général ce que l'on fait pour les réseaux MT, les ouvrages à mesurer ne sont pas excessivement nombreux si l'on se limite aux postes sources et aux départs MT".

Un départ HTA est une ligne électrique intégrée au sein d'un ouvrage électrique plus grand qui est le poste source. D'après le site internet d'Enedis, le poste source est un ouvrage électrique industriel, propriété d'Enedis, qui se trouve à la jonction des lignes électriques HTB (réseau de transport) et HTA (réseau de distribution). Le poste source est constitué de plusieurs transformateurs qui abaissent la tension des lignes HTB (>50kV) vers la tension HTA (<50kV). Chaque ligne

qui part du transformateur est appelé départ HTA, et chaque départ HTA assure l'alimentation en électricité des clients d'un secteur précis.

La figure 3.1 illustre les différents types de clients desservis par un départ HTA : les producteurs HTA et BT (*e.g.* centrales photovoltaïques, éoliennes), les clients HTA (*e.g.* hôpitaux, industries), les clients BT (*e.g.* commerces, hôtels, restaurants, clients résidentiels).

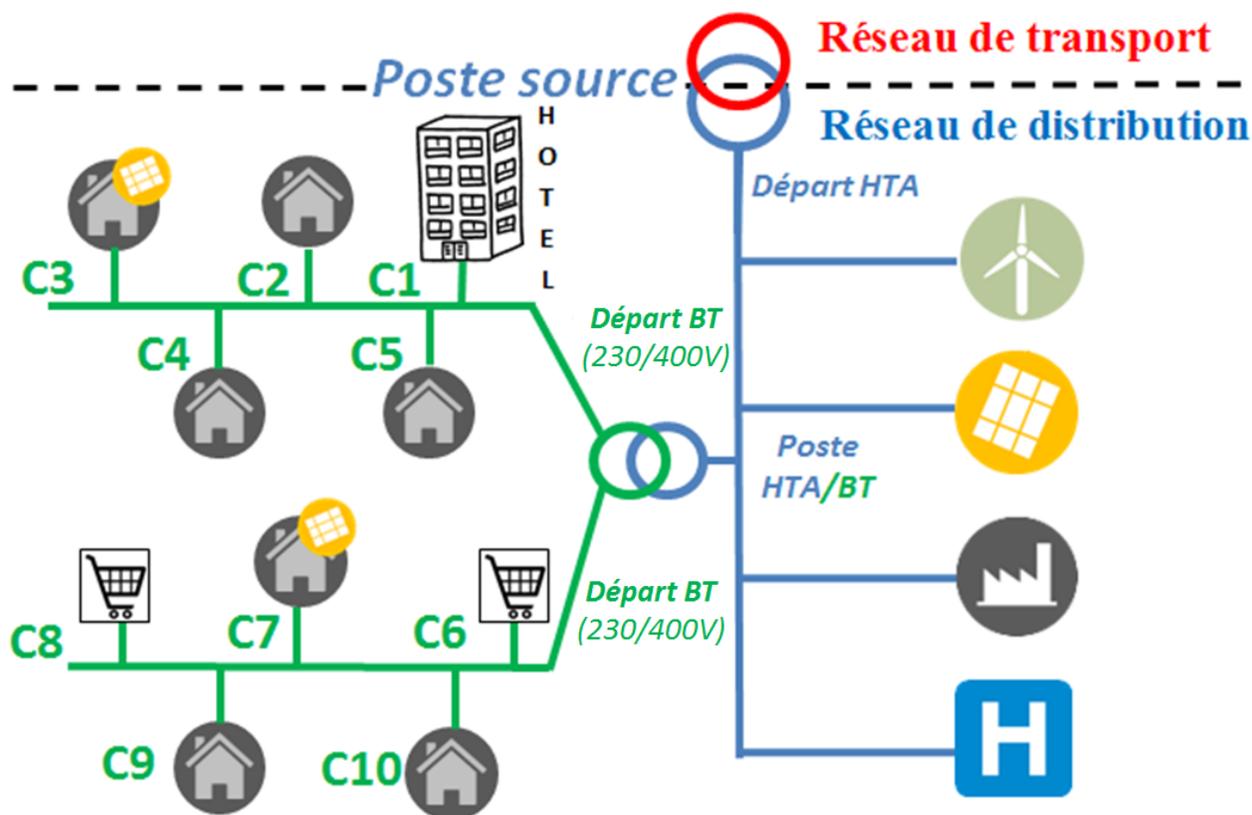


FIGURE 3.1 – Illustration d'une zone et des clients desservis par un départ HTA

**Description des données.** Les données que nous avons pu récupérer pour la thèse, issues du partenariat avec Enedis, sont des mesures sur des départs HTA pour 3 zones en France autour de grandes villes : une autour de Lyon, l'autre autour de Rennes, et la troisième autour de Blois. Au total, c'est près de 3507 mesures de départ HTA qui ont été fournies par Enedis sur 4 années, au pas de temps de 10 minutes (du 1er janvier 2010 au 31 décembre 2013 à 23h50). Cela représente un total de plus de 700 millions de points de mesures de consommation électrique. Chaque départ HTA est identifié un code GDO (Gestion Des Ouvrages) : c'est le nom de son poste source concaténé avec la lettre C suivi de 4 chiffres identifiant le départ. Un exemple de consommation électrique d'un départ HTA sur 4 ans est fourni en figure 3.2.

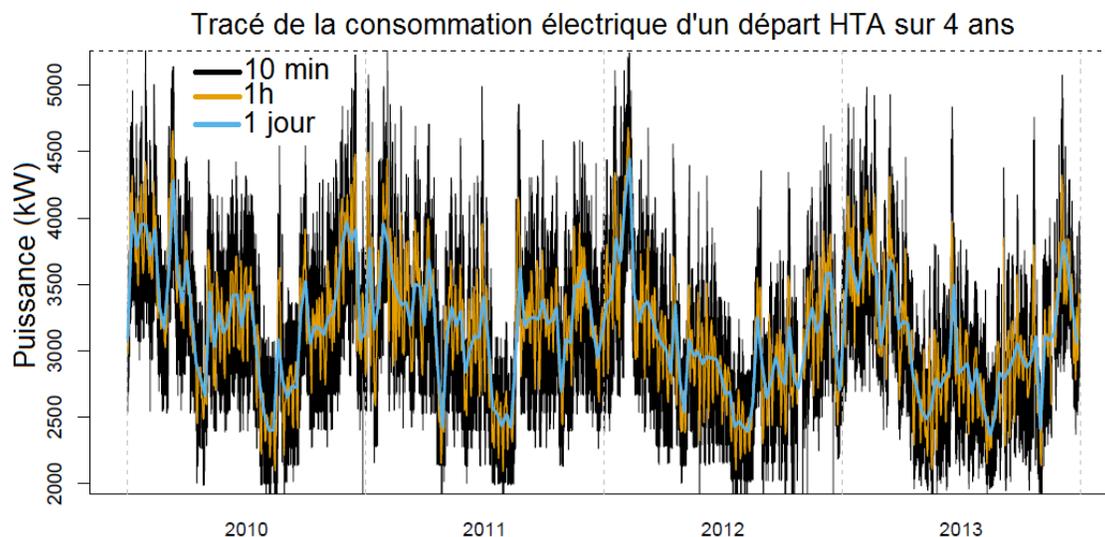


FIGURE 3.2 – Exemple de consommation électrique d'un départ HTA sur 4 ans, avec différents pas de temps

**Détection des erreurs.** Ces données de mesure ont été sujettes à des opérations de détection des erreurs présentes. Notamment, les données peuvent présenter des pics anormalement haut ou des valeurs ponctuellement nulles ou négatives, que nous avons remplacé par un indice “valeur manquante”. Une problématique importante est celle des réorganisations, qui est un problème assez courant chez Enedis. Un départ HTA, pour des raisons de maintenance sur la ligne, est temporairement raccordé à un autre départ [85]. Ceci est visible par exemple en figure 3.3 où l'on voit la première partie de l'année (de janvier à environ mi-mars) une surconsommation par rapport au reste de l'année liée à une réorganisation. Dans les données fournies, un indicateur permet de renseigner si une réorganisation a eu lieu ou non sur le pas de temps considéré. De même que pour les valeurs nulles ou anormalement élevées, nous avons remplacé les données au niveau des réorganisations par un indice “valeur manquante”. De plus, les données sur certains départs n'étaient disponibles que par exemple 1 an sur les 4 ans.

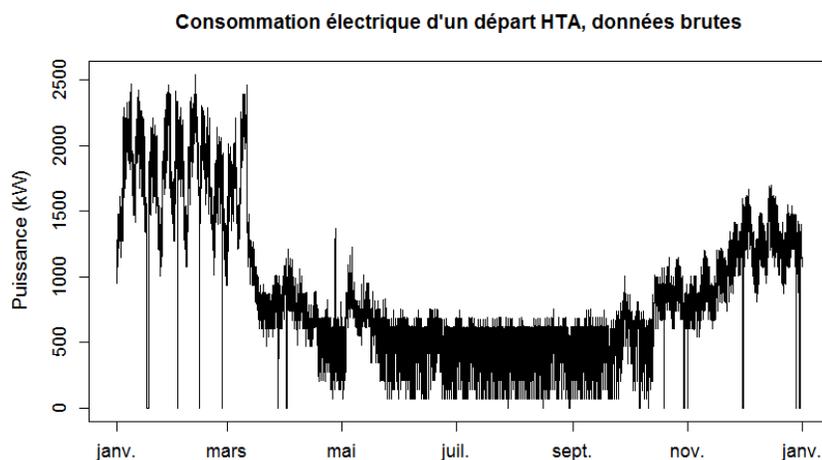


FIGURE 3.3 – Exemple de tracé des données brutes d'un départ HTA

Finalement, après ce contrôle qualité, la disponibilité des données de mesure est de 46% sur les 4 ans pour l'ensemble des 3507 départs HTA fournis. Le tableau 3.1 récapitule la disponibilité des données de mesure pour chaque année

Année	2010	2011	2012	2013
Disponibilité des données	57%	44%	38%	46%

TABLE 3.1 – Disponibilité des données de mesure de départ HTA pour chaque année

### La base des clients du réseau d'Enedis

**Définitions.** Enedis dispose d'une base de données dans laquelle est recensée la liste de leurs clients, avec leurs informations contractuelles, qui est régulièrement mise à jour. En anglais, les auteurs parlent du "Customer Information System" (CIS) [109]. Nous l'appelons dans cette thèse, de même qu'Enedis, la "base des clients" ou "base clients".

Enedis classe les clients de son réseau en deux catégories : les producteurs et les consommateurs. Leur classification est résumée dans le tableau 3.2, qui est issu d'une documentation d'Enedis [117].

Dénomination	Segment utilisateur	Type de tarif (consommateur seulement)	Puissance de raccordement $P_r$	Tension de raccordement
Client courant	C5	Bleu	$P_r < 36\text{kVA}$	BT
Client important	C4	Jaune	$36\text{kVA} < P_r < 250\text{kVA}$	BT
Client HTA	C1, C2, C3	Vert ou émeraude	$250\text{kVA} < P_r$	HTA
Producteur courant	P4		$P_r < 36\text{kVA}$	BT
Producteur important	P3		$36\text{kVA} < P_r < 250\text{kVA}$	BT
Producteur HTA	P1, P2, P3		$250\text{kVA} < P_r$	HTA

TABLE 3.2 – Classification des types d'utilisateur du réseau de distribution par Enedis [117]

**Description des données.** Dans le cadre d'un contrat avec Enedis et le centre PERSEE, Enedis a fourni une liste des clients courants, importants et HTA, ainsi que les producteurs HTA des trois zones décrites au 3.2.1 : Lyon, Rennes et Blois. Ces données correspondent à des extractions sur la base de données d'Enedis. C'est une copie instantanée de l'état des utilisateurs du réseau qui a pu être effectuée début 2014. Il est très complexe pour Enedis de remonter à un instantané de l'état des clients de son réseau dans le passé, et cela n'a pas pu être effectué pour cette thèse.

Sur les différentes données de départs HTA décrites au 3.2.1, nous avons recensé à partir de l'extraction faite par Enedis plus de 2,2 millions de clients courants, 22800 clients importants, 4200 clients HTA, 120 producteurs HTA.

Les informations dont nous disposons par utilisateur du réseau diffèrent selon leur type. Le tableau 3.3 recense les informations pertinentes dont nous disposons par type d'utilisateur.

Un descriptif détaillé des informations relatives aux clients sont disponibles en annexe 8.1.1.

Information	Client courant	Client important	Client HTA	Producteur HTA
Code GDO du départ HTA	×	×	×	×
Code GDO du transformateur HTA/BT	×			
Numéro de rue du branchement	×			
Nom de la rue	×			
Complément d'adresse	×			
Code INSEE	×	×		
Puissance souscrite	×	×	×	×
Puissance de raccordement			×	×
Statut client	×	×	×	×
Consommation annuelle HP	×	×	×	×
Consommation annuelle HC	×	×	×	×
Consommation annuelle en base ou EJP	×	×	×	×
Libellé activité client	×	×		
Libellé Option Tarifaire	×	×		
Code APE			×	

TABLE 3.3 – Informations disponibles selon les types d'utilisateurs du réseau d'Enedis

**Hypothèses et sélection des données.** Nous utilisons dans ces travaux la seule liste des utilisateurs du réseau de 2014 dont nous disposons pour traiter les 4 années de mesures. Ainsi nous supposons que les clients en 2014 étaient les mêmes que 4 ans plus tôt, ce qui est une hypothèse forte car les utilisateurs peuvent très bien déménager, changer d'abonnement, les commerces peuvent fermer etc. Nous verrons dans ce chapitre une méthode qui permet de comparer les informations venant de différents types de données et l'évolution de cette qualité en fonction de l'année.

Les travaux de la thèse n'ont pas été fait en bénéficiant des données concernant les producteurs courants et importants (*e.g.* panneaux solaires sur les toitures) et nous ne les prenons pas en compte dans ces travaux.

Cependant, ces données ont été disponibles pour quelques cas d'étude précis, et il est possible de les intégrer dans des travaux, avec un simulateur de production ENR, comme décrits dans [124, 98].

La modélisation des producteurs HTA est en dehors du cadre de cette thèse. Ainsi nous gardons pour notre étude uniquement les départs HTA sans producteurs HTA.

### Bilan sur les données Enedis

Avec toutes ces hypothèses, l'ensemble des données arrive à 1393 mesures départs HTA constitués de clients courants, importants et HTA. Ce nombre est beaucoup plus faible que les 3507 départs trouvés au 3.2.1 notamment car beaucoup de départs HTA mesurés n'avaient pas de clients dans les données fournies.

Nous appelons la base des mesures des départs HTA d'Enedis l'ensemble des mesures de consommations électriques :

$$\{P^z\}_{z \in D} \quad (3.1)$$

Avec  $D$  l'ensemble des zones desservies par les 1393 départs HTA mesurés pendant 4 ans. Ces me-

sures seront utiles à l'évaluation et la calibration du modèle de consommation électrique développé pendant la thèse et présenté au chapitre 4.

Nous appelons base des clients d'Enedis l'ensemble des clients courants, importants et HTA qui sont connectés à ces 1393 départs HTA sélectionnés.

Cette base des clients permet d'avoir la liste de chaque client  $j$  connecté à chaque départ ainsi que les informations sur l'abonnement de chaque client  $j$ , noté  $Ab^j$ .

Les informations précises contenues dans la variable  $Ab^j$  sont listées dans le tableau 3.3. Au total, c'est plus de 2,2 millions de clients courants avec 14 informations par client, 22800 clients importants avec 14 informations par client et 4200 clients HTA avec 8 informations par client dont nous disposons.

Avec les hypothèses précédemment vues, nous supposons donc dans la suite que :

$$\forall z \in D, \exists J^z / P^z = \sum_{j \in J^z} P^j \quad (3.2)$$

avec  $J^z$  la liste des clients courants, importants et HTA de la base des client d'Enedis dont le code GDO est celui du départ HTA desservant la zone  $z$ ,  $P^z$  est la consommation électrique du départ HTA desservant la zone  $z$ . Cette somme peut aussi se faire par catégorie de clients :

$$P^z = \sum_{k \in K^z} \sum_{j \in J_k^z} P^j \quad (3.3)$$

Cette décomposition en catégories de clients (voir 2.2.1, *e.g.* résidentiels, tertiaires, industries) est très utile car le type d'informations ainsi que les bases de données disponibles est différent selon les catégories de clients.

### 3.2.2 Les bases de données de l'INSEE

L'INSEE publie une très grande quantité de données en accès public sur son site [13].

Ces données sont importantes pour préciser les caractéristiques des clients : elles contiennent des informations sur les facteurs comportementaux, socio-économiques et sur les locaux. Cependant ces données, comme toutes les données en "open data" sont anonymisées : tout est mis en place pour ne pas pouvoir retrouver de quelle personne il s'agit et donc la géolocalisation se fait sur des zones assez larges, au minimum la maille IRIS. La définition qui suit est tirée en grande partie du site de l'INSEE. L'IRIS est un sigle qui signifiait au départ en 1999 « Ilots Regroupés pour l'information Statistique ». Depuis l'IRIS est la brique de base en matière de diffusion des données infra communales. Il doit respecter des critères géographiques et démographiques et avoir des contours identifiables sans ambiguïté et stables dans le temps. Les communes d'au moins 10 000 habitants et une forte proportion des communes de 5 000 à 10 000 habitants sont découpées en IRIS. Ce découpage constitue une partition de leur territoire. La France compte environ 16 100 IRIS dont 650 dans les DOM.

Dans cette section, nous présentons succinctement les différentes bases de l'INSEE ainsi que leur lien avec les facteurs influents sur la consommation électrique.

#### La base logement

Ces données sont issues du recensement de la population. Elles sont localisées à la commune et à l'IRIS et décrivent les caractéristiques des logements et celles des ménages qui les occupent. Cette base de données permet de renseigner les facteurs influant la consommation électrique pour la catégorie des clients résidentiels.

Plus exactement, on a par année un jeu de données pour la France comportant la liste de tous les logements. Chaque logement a un certain nombre d'informations dont des informations spatiales (*e.g.* commune, arrondissement, IRIS), socio-économiques (*e.g.* nombre de personnes par tranche d'âge, diplôme le plus élevé obtenu, activité), sur le local (*e.g.* superficie, année de construction, nombre de pièces, type de logement) et aussi une information sur le parc d'appareils (*i.e.* type de chauffage). D'autres informations moins en lien avec la consommation électrique sont aussi disponibles (*e.g.* année d'aménagement, mode d'évacuation des eaux usées, situation quand à l'immigration). La liste de toutes les variables du fichier logement est disponible en [10].

### La base emploi du temps

Cette base de données est issue d'une enquête effectuée entre septembre 2009 et septembre 2010 auprès de 12 000 ménages ordinaires et de leurs membres. Au total, c'est plus de 18 500 ménages à qui il a été demandé de remplir un carnet journalier au pas de temps de 10min. Finalement ces données ont été traitées et on peut accéder à plus de 100 activités différentes possibles. Cette base de données permet d'avoir un panel noté des comportements des ménages à l'échelle de la France. Cependant aucune information sur la localisation des ménages n'est disponible.

### Le répertoire SIRENE

Cette base rassemble des informations économiques et juridiques sur environ 10 millions d'établissements appartenant à tous les secteurs d'activité [20]. Les différents établissements sont localisés par leur adresse complète et on a des informations notamment sur le nom du dirigeant, le code d'activité principale de l'entreprise (APE) ainsi que l'effectif salarial. Ces données contiennent donc des informations socio-économiques importantes à l'échelle de la France pour le secteur tertiaire et industriel.

#### 3.2.3 La base de données des fichiers fonciers

Ce base est communément appelée MAJIC (mise à jour des informations cadastrales) [21]. Cette base de données regroupe de nombreuses informations sur les locaux et les parcelles cadastrales en France [103] :

- Informations sur les propriétaires
- Usage et caractéristiques du local : habitation ou commercial, surface, nombre de pièces, année de construction, numéro de parcelle cadastrale à laquelle il appartient. . .
- Informations sur les copropriétés
- Informations sur les parcelles : commune, adresse de la parcelle si elle existe.

Ainsi cette base de données donne des renseignements sur les locaux de tous les types de clients ; cette base de données n'est pas anonymisée, et l'adresse des locaux ainsi que le nom des personnes est connu, ce à l'échelle de toute la France.

Cette base de données est disponible auprès de la Direction générale des finances publiques (DGFIP). Elle est payante, au tarif fixé par l'arrêté tarifaire du 16 mai 2011 relatif aux conditions de rémunérations des prestations cadastrales rendues par la DGFIP. Elle est accessible aux collectivités et à tout autre entité chargée d'une mission de service public [2].

### 3.2.4 Les bases de données du CEREN

Le Centre d'études et de recherches économiques sur l'énergie (CEREN) "dispose de bases de données détaillées sur les consommations d'énergie dans le résidentiel, le tertiaire et l'industrie et sur les facteurs explicatifs de ces consommations" [1]. Le CEREN utilise à la fois les bases de données publiques et celles accessibles par les opérateurs énergétiques, et peut aussi réaliser des enquêtes à la demande. Il compte 25 collaborateurs permanents, dont l'INSEE, l'ADEME, EDF, Enedis, RTE, le CREDOC et le MEDEF. Entre autres, il contribue à l'établissement du bilan prévisionnel de RTE, à l'enquête logement de l'INSEE. En particulier, le CEREN publie depuis 2013 des données statistiques sur les consommations énergétiques des trois secteurs (résidentiel, tertiaire, et industriel) à l'échelle nationale. Leur rapport de 2015 [22] présente des données sur les consommations énergétiques par secteur, par branche et par usage à l'échelle nationale. Ces informations seront utilisées pour la partie calibration en énergie du simulateur de courbes de charge (section 5.3).

### 3.2.5 Les bases de données de l'IGN

L'Institut national de l'information géographique et forestière (IGN) met à disposition de nombreuses bases de données, et notamment sur les bâtiments (BD TOPO), sur les adresses (BD ADRESSE) et sur les parcelles (BD Parcellaire). La BD TOPO recense des informations sur les locaux, comme leur surface, leur périmètre ainsi que leur hauteur et leur nature, à l'échelle de la France. La base Adresse Premium (BD Adresse Premium) permet de faire les liens entre les trois bases de données précédentes, via des identifiants uniques. La BD Adresse Premium est disponible sur demande pour l'État ainsi que ses établissements publics à caractère administratif, et est payante pour les entreprises, associations et particuliers [9].

### 3.2.6 Les données sur le parc d'appareils

L'INSEE réalise régulièrement des enquêtes sur les équipements des ménages. Les chiffres sont donnés en taux d'équipement par appareils à l'échelle nationale, selon différents critères comme l'âge de la personne de référence ou le type de ménage (*e.g.* famille monoparentale, couple sans enfant) [79]. Ainsi ces données renseignent sur le parc d'appareils  $Pa$  à l'échelle nationale pour le secteur résidentiel.

Pour le secteur tertiaire et industriel, peu d'études ont été réalisées sur les différents appareils et leur consommation d'énergie [46]. En France, l'INSEE a réalisé en 2011 une enquête sur les consommations d'énergie dans le tertiaire (ECET) et en 2016 une enquête sur les consommations d'énergie dans l'industrie (EACEI). Ces études donnent principalement des informations générales sur la répartition des différentes consommations d'énergie par combustible, et la part consacrée au chauffage [147]. Des informations sur les moyens de chauffage plus détaillées par catégorie de commerces et de services de proximité ont été effectuées par l'ADEME et l'ARENE en 2008 [72]. Cependant dans ces études, il n'y a pas d'information sur les taux de possession et répartition par type d'usage par catégorie de clients tertiaires ou industriels. Pour accéder à ces informations, nous avons utilisé les résultats de campagnes de mesures avec un échantillon restreint, et avons pris l'hypothèse d'une répartition des différents usages à l'échelle nationale similaire aux études. Les différentes campagnes de mesures utilisées sont récapitulées dans le tableau 3.4.

catégorie de clients	Étude utilisée
Bureaux	[31], [30],[114], [113]
Commerce	[29]
Hôtel	[83], [69], [68]
Restaurant	[125], [69], [68]

TABLE 3.4 – Différentes campagnes de mesures utilisées pour estimer les appareils présents dans les différentes catégories de clients tertiaires

### 3.2.7 Les campagnes de mesures

Pour avoir des informations sur la puissance et l'utilisation des appareils, il est nécessaire de les mesurer en conditions réelles d'utilisation. Nous recensons ici les différentes campagnes de mesures depuis les années 1990 qui ont été utiles pour le modèle de consommation électrique développé. Ces études permettent de renseigner les informations sur la technologie des appareils,  $Ta$ , les comportements d'utilisation de ceux-ci,  $Co$ . Devant le coût de telles études, elles sont effectuées sur des échantillons voulant être représentatifs de l'échelle nationale, mais il s'agit à chaque fois d'une dizaine ou d'une centaine d'équipements suivis pendant quelques mois voire quelques années.

#### Secteur résidentiel

Le secteur résidentiel est celui qui a été le plus mesuré. 10 campagnes de mesures, dont la plupart en France, ont eu lieu de 1992 à 2006 [70], sur les principaux usages domestiques. Par la suite, les projets "Remodece" et "Remodece +" ont permis de suivre dans 100 logements en France les appareils de lavage et de froid [27], ainsi que les usages domestiques de l'audiovisuel et de l'informatique [28]. Une étude française a aussi suivi la consommation de 20 piscines individuelles durant l'été 2007 [26]. Un suivi de l'ensemble des usages électroménagers de 400 logements pendant un mois, et de 40 logements pendant une année complète a aussi eu lieu en 2009 en Suède [158]. Enfin, une étude française parue en 2016 a enregistré et analysé la consommation des appareils de production de froid, de lavage et de climatisation de 110 logements [32].

D'autres campagnes de mesures ont eu lieu et ont mesuré la consommation des usages pendant des années. Elles sont couplées à des questionnaires pour connaître les caractéristiques des ménages. C'est le cas notamment des données venant de la compagnie Pecan Street [93]. Ces données proviennent de 345 logements (qui ont l'enregistrement complet d'au moins un appareil) et sont majoritairement localisées à Austin, au Texas, en 2016. Ces données sont utiles pour évaluer le modèle utilisé pour le comportement d'utilisation des appareils.

#### Secteur tertiaire et industriel

Concernant les campagnes de mesures dans le secteur tertiaire, comme vu au 3.2.6, peu de campagnes de mesures existent à notre connaissance sur le secteur tertiaire. Nous avons utilisé les campagnes de mesures décrites au tableau 3.4 pour estimer les paramètres sur la technologie et les comportements des appareils dans le secteur tertiaire. Quant au secteur industriel, nous n'avons pas trouvé d'études représentatives des technologies d'appareils utilisés.

### 3.2.8 Les bases de données sur la météorologie

La NASA met à disposition des réanalyses de données de température provenant du satellite ERA, nommé base MERRA (Modern-era retrospective analysis for research and applications) [17]. Ces données sont au pas de temps de 30min et ont une résolution spatiale de 1km carré. Les données d'irradiation utilisées proviennent de la base de données Helioclim 3 [48] qui permet de disposer des données au pas de temps de 15min à la résolution spatiale de 5km × 7km.

### 3.2.9 Conclusion sur les bases de données

Finalement nous avons décrit ici les différentes bases de données. Nous les avons mises en lien avec les variables explicatives de la consommation électrique et avons constaté leur disponibilité, ainsi que l'échelle spatiale de l'information. Le tableau 3.5 récapitule les différentes données trouvées renseignant les facteurs influents la consommation électrique, ainsi que l'échelle de disponibilité des données.

Paramètre influant	Secteur	Provenance des données	Échelle des données
Parc d'appareils $Pa$	Résidentiel	Enquête équipements des ménages, campagnes de mesures	Échantillon représentatif national
	Tertiaire	Enquête ECET, campagnes de mesures	Échantillon représentatif national
	Industriel	Enquête EACEI	Échantillon représentatif national
Technologique $Ta$	Résidentiel	campagnes de mesures	Échantillon représentatif national
	Tertiaire	campagnes de mesures	Échantillon représentatif national
	Industriel	non trouvé	
Comportement $Co$	Résidentiel	campagnes de mesures, enquête emploi du temps	Échantillon représentatif national
	Tertiaire	campagnes de mesures	Échantillon représentatif national
	Industriel	non trouvé	
Socio-économique $S$	Résidentiel	Enquête logement	Logement anonymisé par IRIS
	Tertiaire	Répertoire SIRENE	Entreprise
	Industriel	Répertoire SIRENE	Entreprise
Météorologie $Me$	Résidentiel	Base MERRA et SODA	kilomètre carré
	Tertiaire	Base MERRA et SODA	Kilomètre carré
	Industriel	Base MERRA et SODA	Kilomètre carré

Paramètre influant	Secteur	Provenance des données	Échelle des données
Local $L$	Résidentiel	Enquête logement, base MAJIC et BDTOPO	Logement anonymisé par IRIS, local et entreprise
	Tertiaire	Base MAJIC et BDTOPO	Local et bâtiment
	Industriel	Base MAJIC et BDTOPO	Local et bâtiment
Abonnement $Ab$	Résidentiel	Bases des clients d'Enedis	Client
	Tertiaire	Base des clients d'Enedis	Client
	Industriel	Bases des clients d'Enedis	Client

TABLE 3.5 – Provenance des données permettant de renseigner les facteurs influant, ainsi que leur échelle de disponibilité

### 3.3 Association de la base clients et de la base logement

Dans le cadre de cette thèse, le cas d'étude concernant l'association entre la base des clients d'Enedis et la base des logements de l'INSEE a été effectué. Nous présentons ici la démarche de la méthode d'association ainsi que l'apport qualitatif et quantitatif de la nouvelle base de données ainsi constituée. L'association des bases MAJIC, BDTOPO et SIRENE, et leur utilisation sont autant de perspectives aux présents travaux.

#### 3.3.1 Introduction au problème d'association

Nous avons vu dans la section précédente que les données disponibles étaient présentes à différentes échelles. Dans cette partie nous allons voir comment nous avons associé la base des clients d'Enedis, où l'échelle disponible est celle du client, à la base logement, où l'échelle disponible est le logement anonymisé par IRIS. Nous allons décrire le problème et la méthode d'association des données, puis montrer l'intérêt quantitatif de cette association.

Dans cette partie, nous ne nous intéressons qu'à la catégorie des clients résidentiels, notée  $k = res$ . **Dans cette section uniquement, quand nous nous référons aux clients, il s'agit uniquement des clients résidentiels.** L'association de la base des clients et des logement a fait l'objet d'une communication [39].

#### 3.3.2 Description du problème d'association

La base des clients d'Enedis contient des informations sur les abonnements des clients :  $Ab^j$ . Cette variable contient notamment à quel départ HTA est connecté le client, son adresse etc.

La base logement de l'INSEE contient des informations sur le parc d'appareils, les facteurs socio-économiques et les locaux par logement  $l$  :  $Pa^l, S^l, L^l$ . Le logement est anonymisé mais on connaît son IRIS d'appartenance.

Le problème d'association consiste à faire correspondre les logements  $l$  anonymes à la maille IRIS aux clients  $j$ , afin de pouvoir lier à chaque client (où l'on a déjà des informations sur son abonnement) des informations de la base logement.

Une façon simple de restreindre ce problème est de le poser par maille IRIS :

- On a une liste de logements anonymisés dans l'IRIS ;
- On a une liste de clients géolocalisés et rattachés à un certain départ HTA dans l'IRIS ;
- On réalise une méthode d'association client/logement dans l'IRIS.

Pour ce faire, un premier travail a consisté à associer chaque client à un IRIS. Ceci est possible car on connaît l'adresse des clients. La base BD adresse de l'IGN donne la correspondance adresse-maille IRIS. Un algorithme a été développé pour faire correspondre adresse client-adresse BD adresse, avec une méthode d'association de chaînes de caractères. Cette méthode est détaillée en annexe 8.1.2.

Une fois l'IRIS des clients estimé, deux méthodes d'association client/logement ont été développées :

- une méthode dite de référence, qui associe les clients à des logements moyens par IRIS ;
- une méthode d'association par IRIS basée sur une optimisation de la distance entre la consommation estimée de chaque logement et la consommation mesurée des clients.

### 3.3.3 Association par la méthode de référence

Cette méthode consiste à associer globalement à un départ HTA les logements de l'IRIS sur lequel ses clients desservis sont situés, en pondérant chaque logement par la proportion de clients dans le départ appartenant à cet IRIS. Cette méthode permet d'avoir une liste de logements pondérés par départ, mais n'associe pas de façon unitaire logement et client.

Si on note  $p_{ir \cap z}$  le nombre de clients dans l'IRIS  $ir$  du départ desservant une zone  $z$  et  $n_{ir}$  le nombre de logements dans l'IRIS  $ir$ , cela revient à avoir pour un logement  $l$  d'un IRIS  $ir$  une pondération  $pond_{l,z}$  pour un départ desservant une zone  $z$  :

$$pond_{l,z} = \frac{p_{ir \cap z}}{n_{ir}} \quad (3.4)$$

Pour illustrer cela, considérons un départ  $d$  avec 20 clients. 10 sont situés sur l'IRIS A, 10 sur l'IRIS B. L'IRIS A comporte 10 logements et l'IRIS B 100 logements. Ainsi avec cette méthode de référence le départ  $d$  sera associé aux 10 logements de l'IRIS A pondéré par  $\frac{10}{10} = 1$  et aux 100 logements de l'IRIS B pondéré par  $\frac{10}{100} = 0.1$ .

### 3.3.4 Association par une méthode d'optimisation

#### Estimation de la consommation annuelle des logements.

La méthode consiste d'abord à estimer la consommation électrique annuelle de chaque logement  $l$  à partir des différentes variables qui lui sont associées (facteurs socio-économiques, sur les locaux et les appareils). Pour cela, notre méthode utilise un modèle d'estimation de l'énergie par logement qualifié par Swan et Ugursal [143] de bottom-up utilisant une méthode statistique de régression.

Il s'agit de supposer la consommation électrique annuelle  $P^l$  d'un logement  $l$  comme étant une fonction de ses facteurs influents :

$$P^l \sim (S^l, L^l, Pa^l) \quad (3.5)$$

Avec  $S^l$  le vecteur des informations socio-économiques,  $L^l$  des locaux,  $Pa^l$  du parc d'appareils du logement  $l$ .

On peut classer ces facteurs en deux catégories :

- les variables extensives (*i.e.* qui sont sommées si on a un ensemble de logements). C'est le cas de la surface  $Surf$ , du nombre de personnes  $INPER$  ainsi que du nombre de pièces  $NBPI$ .
- les variables intensives (*i.e.* dont la valeur ne dépend pas de la taille du logement). C'est le cas de la catégorie de logements  $CATL$  (*e.g.* appartement, maison individuelle), du moyen de chauffage central du logement  $CHFL$  (*i.e.* électricité ou non), si le ménage est propriétaire ou non du logement  $STOCD$ , du type de logement  $TYPL$  (résidence principale, secondaire, logement vacant) et enfin du combustible principal du logement  $CMBL$  (*i.e.* électricité ou non), et la période d'achèvement de la construction du logement  $ACHL$ .

Nous disposons de la consommation des clients, ainsi que leur appartenance à un IRIS grâce à la base des clients d'Enedis. Afin de constater le type de dépendance entre les variables extensives et les consommations en énergie des clients, nous avons représenté l'énergie annuelle consommée par client dans chaque IRIS  $ir : \sum_{j \in ir} P^j$  en fonction de chacune des variables extensives dont nous disposons par IRIS à partir de la base logement en figure 3.4. Avec les données dont nous disposons, c'est la consommation de 1530 IRIS qui a été représentée.

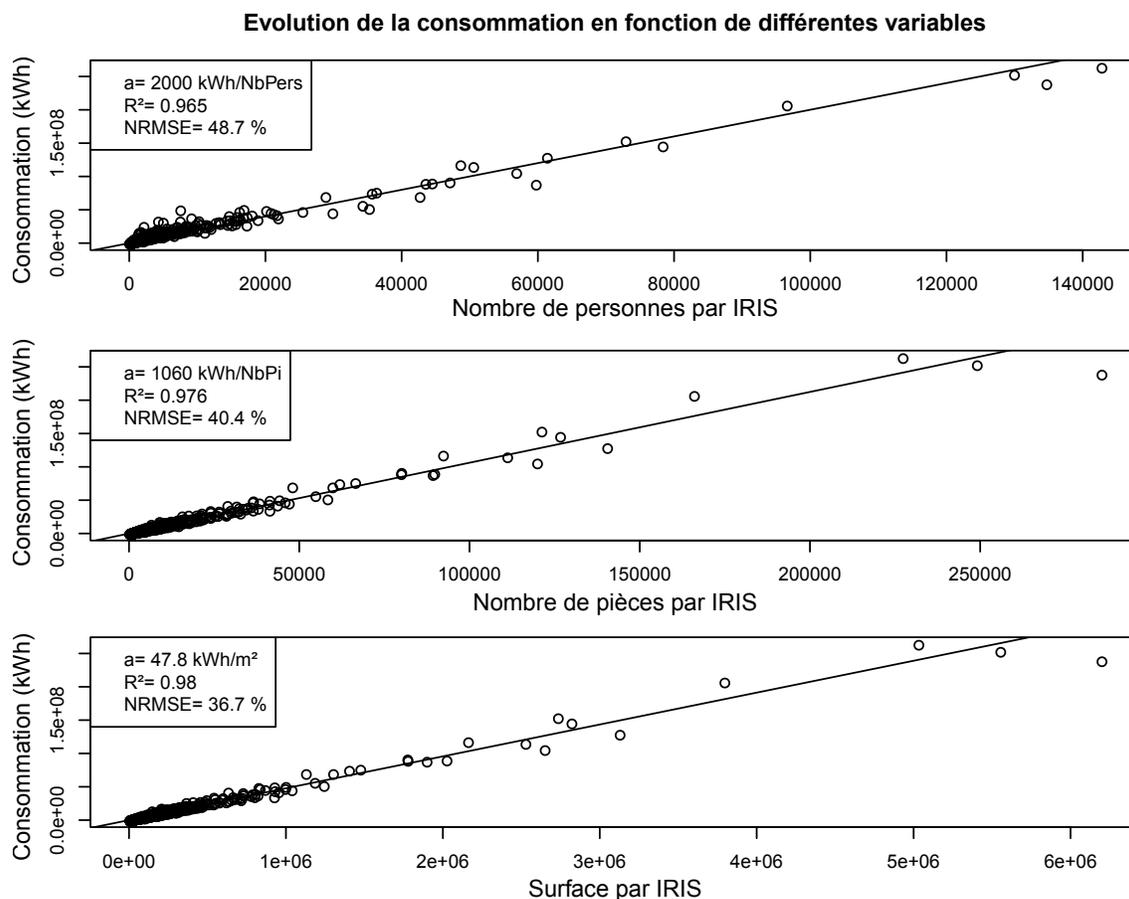


FIGURE 3.4 – Évolution de la consommation électrique résidentielle par IRIS en fonction de chacune des variables extensives

Les résultats illustrés en figure 3.4, montrent des coefficients de regression ( $R^2$ ) élevés pour les variables extensives ( $>0.965$ ).

Les coefficients directeurs des droites de régression nous fournissent des informations intéressantes sur la dépendance de la consommation annuelle à année normale Enedis, avec les variables extensives : 2000kWh par habitant d'un ménage, ou de 1060kWh par pièce, ou encore 47.8kWh/m<sup>2</sup>.

Le tableau 3.6 met à l'échelle nationale ces valeurs avec les chiffres nationaux du recensement en 2013 et les compare avec la consommation annuelle de 2013 du secteur résidentiel issu de [126]. On constate que l'on a tendance à sous-estimer la moyenne nationale d'un peu plus d'une dizaine de pour cent pour les trois variables.

Variable extensive	Nombre de personnes (hab)	Nombre de pièces	Surface (m <sup>2</sup> )
Valeur consommation (kWh/unité de variable extensive)	2000	1060	48
Valeur moyenne par ménage de la variable (INSEE 2013)	2,3	4,1	92
Valeur moyenne ménage x valeur consommation (kWh)	4600	4346	4398
Consommation moyenne ménage 2013 [126] (kWh)	5797	5797	5797
Erreur normalisée (%)	12	15	14

TABLE 3.6 – Remise à l'échelle nationale des résultats de la figure 3.4 pour comparaison des ordres de grandeurs

Ainsi la fonction vue à l'équation 3.5 a été choisie comme étant une fonction linéaire par rapport à ses variables :

$$P^l \sim_l (S^l, L^l, Pa^l) \quad (3.6)$$

Où l'opérateur  $\sim_l$  signifie "est fonction linéaire de".

Pour l'apprentissage des coefficients de la fonction, nous ne disposons pas des consommations des logements, car c'est justement ce que l'on cherche à estimer. Mais nous disposons de la consommation des clients, ainsi que leur appartenance à un IRIS. Ainsi l'apprentissage de la fonction peut se faire par IRIS  $ir$ , et on suppose que :

$$\forall ir \in Ir, \sum_{j \in ir} P^j = \sum_{l \in ir} P^l \quad (3.7)$$

Avec l'équation 3.6 et 3.7, il vient :

$$\forall ir \in Ir, \sum_{j \in ir} P^j \sim_l \sum_{l \in ir} (S^l, L^l, Pa^l) \quad (3.8)$$

$$\forall ir \in Ir, \sum_{j \in ir} P^j \sim_l \left( \sum_{l \in ir} S_l, \sum_{l \in ir} L_l, \sum_{l \in ir} Pa_l \right) \quad (3.9)$$

Ainsi, trouver les coefficients de la fonction linéaire estimant les consommations des logements revient à effectuer une régression linéaire entre la somme de la consommation électrique des clients par IRIS  $ir$  et la somme des des combinaisons des facteurs influents. Avec les données de la thèse

c'est la consommation de 1530 IRIS qui est disponible, ainsi que 1440 combinaisons de variables possibles.

Un vecteur *cat* des différentes combinaisons des variables intensives a été construit puis différentes régressions ont été testées.

Si on note  $P^l_r$  le vecteur des consommations électriques par IRIS,  $Surf^l_r$ ,  $INPER^l_r$ ,  $NBPI^l_r$  respectivement le vecteur des surfaces, nombre de personnes et nombre de pièces par IRIS, les différentes régressions testées sont de la forme :

$$P^l_r \sim_l (Surf^l_r, INPER^l_r, NBPI^l_r) | cat \quad (3.10)$$

La combinaison optimale a été choisie afin de minimiser l'erreur lors de la régression. Le nombre de catégories étant conséquent, un sur-ajustement est possible. Pour éviter le sur-ajustement, la méthode du leave-one-out a été appliquée.

Les résultats en utilisant seulement la surface comme variable extensive sont présentés dans le tableau 3.7. A titre indicatif, les 7 dernières colonnes sont les erreurs en n'utilisant qu'une seule catégorie. La prévision avec la surface seule est la référence.

On constate que l'information contenue dans l'âge du bâtiment *ACHL* est très influente sur la consommation annuelle quand elle est couplée à la surface : on améliore de plus de 54% les erreurs de prédiction par rapport au modèle avec la simple surface. Il vient ensuite l'information contenue dans le combustible principal du logement *CMBL*, très proche derrière la catégorie de logements *CATL* (*e.g.* appartement, maison), le type de chauffage central de la maison *CHFL*, le type de logement *TYPL* (*e.g.* résidence principale ou secondaire) et enfin si le ménage est propriétaire ou non *STOCD*.

Les mêmes résultats montrant l'influence des variables intensives sur les autres variables extensives (nombres de personnes : tableau 8.2; nombre de pièce : tableau 8.3) sont donnés en annexe 8.1.3.

En faisant la régression linéaire en additionnant les variables intensives entres elles, les résultats sont présentés dans le tableau 3.8, toujours en utilisant comme référence les erreurs de prévision du modèle avec uniquement la surface. On voit que l'on arrive à une amélioration importante du modèle initial. Chaque variable extensive est associée différemment avec son jeu de variables intensives pour diminuer l'erreur. On constate aussi que les meilleurs modèles sont ceux avec trois variables intensives associées, au delà il y a un sur-ajustement dans la régression. En effet, si on regarde le tableau complet les modèles à 4, 5 et 6 variables sont classés bien moins performants que ceux présentés ici.

Finalement, au vu des résultats dans le tableau 3.8, nous utilisons le modèle le plus performant pour prédire les consommations par logement, c'est à dire que la fonction  $f$  vue au 3.6 s'exprime par :

$$P^l \sim_l Surf | (CATL \times ACHL \times STOCD) + NbPer | (CATL \times ACHL \times STOCD) + NbPi | (CATL \times TYPL \times CMBL) \quad (3.11)$$

Avec  $P^l$  la consommation électrique annuelle d'un logement  $l$ .

### Association des clients et des logements par une méthode d'optimisation.

Nous disposons dorénavant d'une liste par IRIS de logements avec une consommation électrique annuelle estimée  $P^l$ , et d'une liste de clients avec une consommation annuelle  $P^j$  donnée par la base des clients.

Variables du modèle	Amélioration de l'erreur (%)
Surf (CATLxACHLxCHFLxSTOCDxCMBL)	70,1
Surf (CATLxACHLxSTOCDxTYPLxCMBL)	69,23
Surf (CATLxACHLxCHFLxCMBL)	66,99
Surf (CATLxACHLxSTOCDxCMBL)	66,63
Surf (CATLxACHLxSTOCD)	66,32
Surf (CATLxACHLxCHFLxSTOCD)	62,5
Surf (CATLxACHLxSTOCDxTYPL)	62,16
Surf (ACHLxCHFLxSTOCDxCMBL)	60,92
Surf (CATLxACHL)	60,41
Surf (ACHLxSTOCD)	60,16
Surf (ACHLxSTOCDxTYPL)	58,38
Surf (CATLxACHLxCMBL)	58,16
Surf (ACHLxSTOCDxCMBL)	57,75
Surf (CATLxACHLxCHFL)	57,35
Surf (ACHLxTYPLxCMBL)	56,66
Surf (CATLxACHLxCHFLxSTOCDxTYPL)	56,41
Surf (ACHLxCHFLxSTOCD)	56,19
Surf (ACHLxCHFLxCMBL)	55,95
Surf (ACHLxTYPL)	55,74
Surf (ACHL)	54,55
Surf (CMBL)	23,07
Surf (CATL)	21,07
Surf (CHFL)	20,44
Surf (TYPL)	13,91
Surf (STOCD)	4,23
Surf	0

TABLE 3.7 – Amélioration de l'erreur de prédiction de la consommation par IRIS en utilisant différentes variables intensives couplées à la Surface par rapport au modèle de référence (surface par IRIS)

Variables du modèle	Amélioration de l'erreur (%)
Surf (CATLxACHLxSTOCD) NbPer (CATLxACHLxSTOCD) NbPi (CATLxTYPLxCMBL)	89,66
Surf (CATLxACHLxSTOCD) NbPer (CATLxACHLxSTOCD) NbPi (CATLxCHFLxTYPL)	89,24
Surf (CATLxACHLxSTOCD) NbPer (ACHLxSTOCDxCMBL) NbPi (CATLxTYPLxCMBL)	89,21
Surf (CATLxTYPLxCMBL) NbPer (CATLxACHLxSTOCD) NbPi (CATLxACHLxSTOCD)	89,04
Surf (CATLxTYPLxCMBL) NbPer (ACHLxSTOCDxCMBL) NbPi (CATLxACHLxSTOCD)	88,99
Surf (CATLxCHFLxSTOCD) NbPer (CATLxACHLxSTOCD) NbPi (CATLxACHLxTYPL)	88,89
Surf (CATLxACHLxTYPL) NbPer (ACHLxSTOCDxCMBL) NbPi (CATLxACHLxSTOCD)	88,84
Surf (CATLxCHFLxTYPL) NbPer (CATLxACHLxSTOCD) NbPi (CATLxACHLxSTOCD)	88,77
Surf (CATLxACHLxTYPL) NbPer (ACHLxSTOCDxCMBL) NbPi (CATLxSTOCDxCMBL)	88,57
Surf (CATLxSTOCDxCMBL) NbPer (CATLxACHLxSTOCD) NbPi (CATLxACHLxTYPL)	88,49
Surf (CATLxACHLxSTOCD) NbPer (ACHLxCHFLxSTOCD) NbPi (CATLxACHLxTYPL)	88,46
Surf (CATLxTYPLxCMBL) NbPer (CATLxACHLxSTOCD) NbPi (ACHLxSTOCDxTYPL)	88,41
Surf (CATLxACHLxTYPL) NbPer (CATLxACHLxSTOCD) NbPi (ACHLxSTOCDxCMBL)	88,39
Surf	0

TABLE 3.8 – Amélioration de l'erreur de prédiction de la consommation par IRIS en utilisant différentes variables intensives couplées à la Surface par rapport au modèle de référence (surface par IRIS)

Le problème d'optimisation consiste à associer par IRIS un nombre  $p$  de clients, chaque client étant noté  $j$ , et de  $n$  logements, chaque logement étant noté  $l$ .

La formulation du problème d'optimisation adoptée est la suivante :

$$\min_{(\alpha_{l,j})_{(l,j) \in \{1,2,\dots,n\} \times \{1,2,\dots,p\}}} \left[ \sum_{j=1}^p \left( \sum_{l=1}^n \alpha_{l,j} \times P^l - P^j \right)^2 \right] \quad (3.12)$$

*S.t.* :

$$\alpha_{l,j} \in \{0, 1\} \forall (l, j) \in \{1, 2, \dots, n\} \times \{1, 2, \dots, p\} \quad (3.13)$$

$$\sum_{j=1}^p \alpha_{l,j} \geq 1 \forall l \in \{1, 2, \dots, n\} \quad (3.14)$$

$$\sum_{j=1}^n \alpha_{l,j} = 1 \forall j \in \{1, 2, \dots, p\} \quad (3.15)$$

Avec :

- $(\alpha_{l,j})_{(l,j) \in \{1,2,\dots,n\} \times \{1,2,\dots,p\}}$  l'inconnue d'association :  $\alpha_{l,j} = 1$  si le client  $j$  est associé avec le logement  $l$ , 0 sinon (équation 3.13) ;
- L'équation 3.14 est la contrainte que chaque logement est associé à au moins un client ;
- L'équation 3.15 est la contrainte que chaque client est associé à un unique logement.

Ce problème d'optimisation avec l'équation 3.12 est un problème classique d'une fonction quadratique à minimiser. Cependant, la contrainte exprimée par l'équation 3.13 rend le problème discret. Des méthodes de résolution utilisant des relaxations de la contrainte 3.13 en un intervalle et en ajoutant des contraintes additionnelles pour forcer les solutions à être proches des bornes 0 ou 1 ont été testées sans succès. Tester chaque combinaison de  $(\alpha_{l,j})_{(l,j) \in \{1,2,\dots,n\} \times \{1,2,\dots,p\}}$  mène à des temps de calculs trop longs car le nombre de possibilités est en  $2^{n \times p}$ .

Un algorithme de résolution approché a été développé en implémentant l'algorithme suivant en C++, en supposant le cas  $n > p$  (le cas inverse est implémenté exactement de la même manière en inversant  $P^l$  et  $P^j$ ). Un schéma illustratif de la méthode de résolution approchée est donnée en figure 3.5.

- État initial : tri des  $P^l$  et  $P^j$  ;
- Association aléatoire du jeu de données le plus grand pour avoir la même taille. On obtient alors le même nombre de données à associer  $P^{j \in \{1,2,\dots,p\}}$  et  $\tilde{P}^{j \in \{1,2,\dots,p\}}$  ;
- Minimisation de la distance globale  $\sum_{j=1}^p (\tilde{P}^j - P^j)^2$  en déplaçant le  $\tilde{P}^j$  qui minimise cette distance (résultat de la première étape sur la figure 3.5) ;
- Arrêt quand la distance globale reste la même, ou à partir d'un certain nombre d'itérations quand  $n \times p$  est important (empiriquement  $> 250\,000$ ) (résultat de la seconde étape sur la figure 3.5).

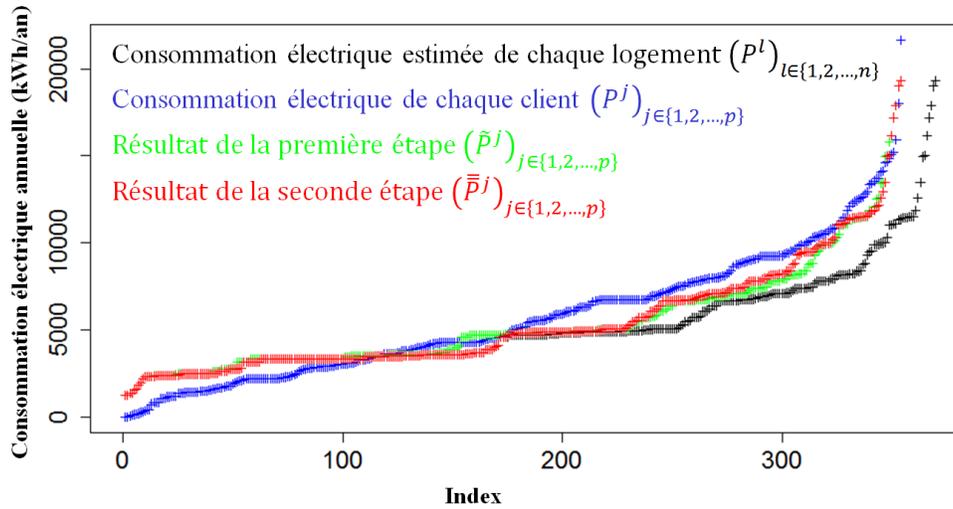


FIGURE 3.5 – Illustration des différentes étapes de l'algorithme de résolution approché du problème d'association [39]

### 3.3.5 Quantification de l'intérêt de la méthode d'association

Les deux méthodes d'association précédemment vues ont permis d'enrichir la base des clients d'Enedis qui disposait uniquement de données d'abonnement des clients avec des informations socio-économiques, sur les locaux et les appareils des clients.

L'objectif de cette partie est de quantifier l'apport de cette association de bases de données, et de pouvoir comparer l'intérêt de l'algorithme d'association par la méthode d'optimisation qui est relativement complexe par rapport à celui par la méthode de référence. Pour ce faire, nous avons utilisé ces nouvelles bases de données pour prédire des indicateurs par départ HTA.

#### Mise en place des indicateurs à prédire

Le premier indicateur qui nous a paru pertinent à prédire à partir de ces nouvelles bases de données est la thermosensibilité des départs HTA. La consommation électrique est fortement influencée par la température extérieure avec tous les usages sensibles à la température, et particulièrement les appareils de chauffage électrique présents chez les clients résidentiels. La consommation électrique de la France est particulièrement thermosensible, puisque qu'elle représente 40% de la thermosensibilité européenne [127].

Cet effet est bien connu et peut être simplement modélisé par une méthode aussi utilisée par RTE en France à l'échelle nationale [102]. Appliquée aux départs HTA, cette méthode consiste en une régression linéaire de la consommation en fonction de la température moyenne de la zone desservie. Dans notre cas, nous limiterons une régression sur tous les points de consommation, alors que RTE calcule une valeur sur chaque heure de la journée.

Si on considère un départ HTA desservant une zone  $z$ , ayant une consommation  $P^z(t)$  à l'instant  $t$ , une température de l'air moyenne de la zone  $T^z(t)$ , le calcul de la thermosensibilité revient à trouver un seuil de température  $Ts^z$  ainsi qu'un coefficient de régression  $Th^z$  par départ tel que :

$$P^z(t) = \begin{cases} P^{z,0}(t) & \text{si } T^z(t) > Ts^z \\ P^{z,0}(t) - Th^z (Ts^z - T^z(t)) & \text{sinon.} \end{cases} \quad (3.16)$$

$P^{z,0}(t)$  est la partie non thermosensible de la consommation électrique du départ HTA. Dans notre cas nous appelons thermosensibilité d'un départ HTA le coefficient de régression  $Th^z$  qui s'exprime en  $\text{kW}/^\circ\text{C}$ .

En croisant les informations venant de la base des clients associée aux foyers ainsi que celles des mesures des départs HTA, c'est 707 départs HTA pour l'année 2012 sur lesquels on a pu calculer puis prédire la thermosensibilité.

### Prédiction de la thermosensibilité

De même qu'au 3.3.4 pour les consommation annuelles, on représente ici la thermosensibilité par départ en fonction des facteurs extensifs dont nous disposons avec la méthode d'association de référence sur la figure 3.6, et la méthode d'optimisation sur la figure 3.7. Dans ce cas la thermosensibilité a été calculée en prenant une température seuil fixe à  $15^\circ\text{C}$ .

On remarque que le nuage de points est assez éparse. La thermosensibilité est un indicateur plus complexe que la consommation annuelle a priori car il dépend surtout des usages thermosensibles, et des surfaces qui sont chauffées à l'électricité. Or les variables extensives seules, comme c'est le cas ici, n'ont pas ces informations. Certains départs présentent probablement un manque de données clients (les points formant une ligne sur la gauche qui sont présents quelles que soient les variables).

Tout comme pour la consommation annuelle, les coefficients directeurs des droites de régression nous fournissent des informations intéressantes sur la dépendance de la thermosensibilité avec les variables extensives :  $26\text{W}/^\circ\text{C}$  par habitant d'un ménage, ou de  $12,8\text{W}/^\circ\text{C}$  par pièce, ou encore  $0,565\text{W}/^\circ\text{C}$  par  $\text{m}^2$ .

Le tableau 3.9 met à l'échelle nationale les valeurs pour la méthode d'optimisation avec les chiffres nationaux du recensement en 2013 et les compare avec la thermosensibilité moyenne de 2013 du secteur résidentiel issu de [126]. On constate que l'on a tendance à sous-estimer la moyenne nationale de quelques pour cent quand on utilise le nombre de personnes, de l'ordre de la vingtaine de pour cent avec les deux autres variables.

Variable extensive	Nombre de personnes (hab)	Nombre de pièces	Surface ( $\text{m}^2$ )
Valeur thermosensibilité ( $\text{kW}/^\circ\text{C}$ / unité de variable extensive)	0,026	0,0128	0,000565
Valeur moyenne par ménage de la variable (INSEE 2013)	2,3	4,1	92
Valeur moyenne ménage x valeur consommation x nb ménages ( $\text{GW}/^\circ\text{C}$ )	1,62	1,43	1,42
Thermosensibilité moyenne résidentielle 2013 [126] ( $\text{GW}/^\circ\text{C}$ )	1,68	1,68	1,68
Erreur (%)	5,1	25	26

TABLE 3.9 – Remise à l'échelle nationale des résultats de la figure 3.7 pour comparaison des ordres de grandeur.

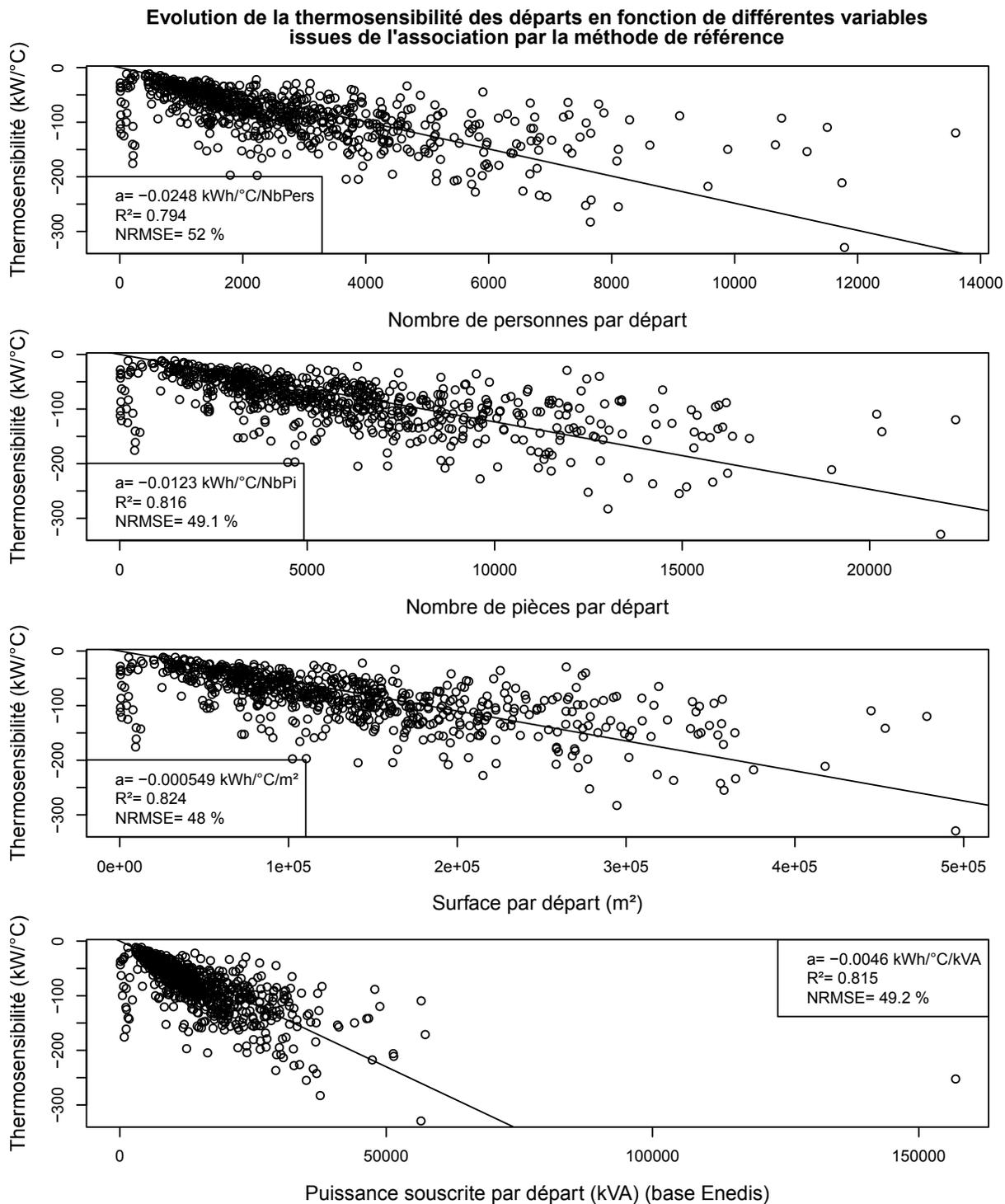


FIGURE 3.6 – Évolution de la thermosensibilité par départ HTA en fonction de chacune des variables extensives obtenues par la méthode de référence

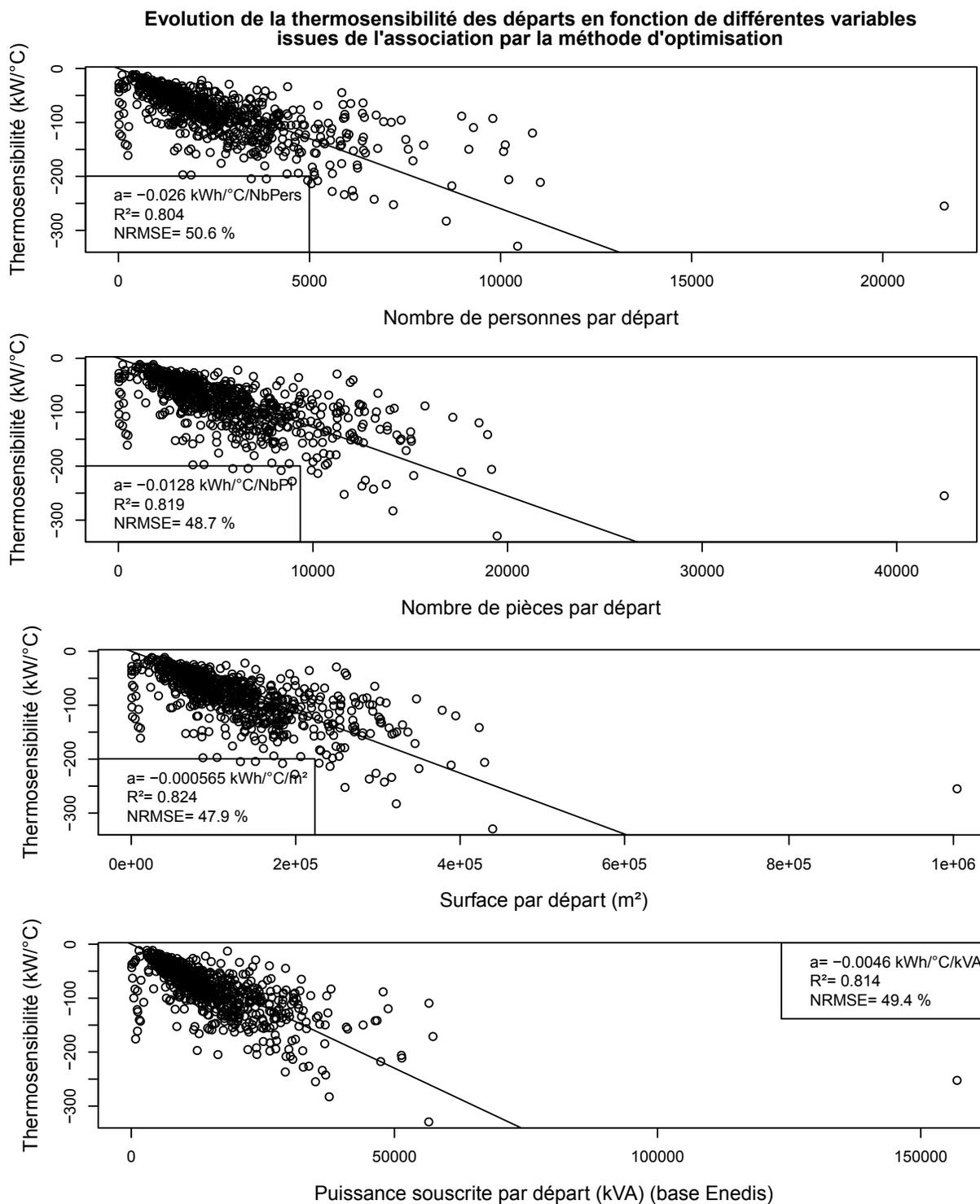


FIGURE 3.7 – Évolution de la thermosensibilité par départ HTA en fonction de chacune des variables extensives obtenues par la méthode d'optimisation

C'est la puissance souscrite qui est l'indicateur extensif qui prédit le mieux la thermosensibilité dans le cas de référence ( $R^2=0.449$ ). Ceci s'explique très probablement par le fait que la thermosensibilité d'un départ HTA somme celle du secteur résidentiel, tertiaire, et industriel, et l'association avec la base logement ne permet d'ajouter que des informations sur la partie résidentielle; la somme des puissances souscrites par départ, quant à elle, prend en compte tous les types de clients. On constate que l'algorithme d'optimisation améliore le  $R^2$  de la régression de l'ordre de quelques dixièmes par rapport à la méthode de référence, ce qui fait passer le  $R^2$  obtenu avec le nombre de personnes devant celui obtenu par la puissance souscrite pour l'association par méthode d'optimisation.

Si on prend comme variable pour prédire la thermosensibilité non pas la somme des puissances souscrites par départ mais la puissance souscrite par catégorie de clients (*e.g.* bleu domestique, bleu agricole, bleu hôtel restaurant), la régression sera alors :

$$Th^D \sim_l (Psous^D)|K \quad (3.17)$$

Avec  $Th^D$  le vecteur contenant la thermosensibilité de chaque départ,  $Psous^D$  le vecteur contenant la puissance souscrite de chaque départ et  $K$  l'ensemble de toutes les catégories de clients. Le  $R^2$  avec cette régression est de 0.504. Ce sera l'erreur de ce modèle qui servira de référence pour prédire la thermosensibilité. Toute amélioration à partir de ce chiffre sera un apport de la base logement pour prédire la thermosensibilité des départs HTA.

Le tableau 3.10 représente les résultats de l'amélioration des erreurs pour la variable extensive  $NbPer$  en fonction des différentes variables intensives. On voit qu'on améliore déjà de 30% dans le cas de la méthode de référence et de 40% dans le cas de la méthode d'optimisation les erreurs de prédiction de la thermosensibilité par rapport à la base Enedis seule.

La méthode d'optimisation apporte plus de précision dans la nouvelle base de données puisqu'elle améliore les erreurs de prévision de l'ordre de 10% par rapport à la méthode de référence.

Les mêmes tableaux pour la surface (table 8.7 et 8.6) ainsi que le nombre de pièces (table 8.7 et 8.6) sont fournies en annexe. On peut noter que pour la méthode de référence, le fait de prendre comme variable surface et le nombre de pièces dégrade les performances de la prédiction. Si on prend uniquement les surfaces et les pièces conditionnées par la variable intensive  $CMBL$  et  $CHLF$  égales à "chauffage électrique", on a alors une amélioration des performances pour la méthode de référence (table 8.5, 8.7, 8.9). Ce n'est pas le cas pour les résultats avec l'algorithme par méthode d'optimisation (table 8.4, 8.6, 8.8).

Une première explication vient du fait que la variable intensive  $CMBL$  signifie "combustible principal du logement", la variable  $CHFL$  signifie "chauffage central du logement". Cependant cela ne veut pas dire que les logements qui n'ont pas le champ "chauffage électrique" dans ces variables n'ont pas d'usage thermosensibles. Ils peuvent par exemple disposer de chauffages électriques d'appoint. D'autres usages, tel le chauffe-eau, sont aussi un peu thermosensibles notamment car la température d'arrivée d'eau est différente selon la saison, ainsi que la température de la pièce qui va jouer sur les déperditions. Cependant, il semble que ces informations de second ordre ne soient pas significatives dans les résultats par la méthode de référence et l'ajout de ces variable a plutôt tendance à dégrader les erreurs. Mais l'information semble prise en compte dans la méthode d'optimisation puisqu'on améliore de quelques points le pourcentage d'erreur en prenant les paramètres extensifs qui ne sont pourtant pas "chauffage électrique".

Variables du modèle optim	Amélioration de l'erreur optim (%)	Variables du modèle ref	Amélioration de l'erreur ref (%)	Apport de la méthode d'optim
NbPer (ACHLxSTOCD xTYPLxCMBL) Psous K	40,35	NbPer (STOCDxCMBL) Psous K	29,51	10,84
NbPer (ACHLxCHFL) Psous K	40,08	NbPer (CATLxSTOCD xCMBL) Psous K	29,51	10,57
NbPer (ACHLxCHFL xTYPL) Psous K	39,98	NbPer (CHFLxSTOCD) Psous K	29,23	10,75
NbPer (ACHLxTYPL xCMBL) Psous K	39,85	NbPer (CATLxCHFL xSTOCD) Psous K	29,23	10,62
NbPer (ACHLxSTOCD) Psous K	39,59	NbPer (CHFLxSTOCD xCMBL) Psous K	28,86	10,73
NbPer (ACHLxCMBL) Psous K	39,54	NbPer (CATLxCHFL xSTOCDxCMBL) Psous K	28,86	10,68
NbPer (ACHLxCHFL xSTOCDxTYPL) Psous K	39,32	NbPer (ACHLxSTOCD xCMBL) Psous K	28,39	10,93
NbPer (ACHLxCHFL xCMBL) Psous K	39,23	NbPer (CATLxACHL xSTOCDxCMBL) Psous K	28,39	10,84
NbPer (ACHL) Psous K	39,14	NbPer (CHFLxTYPL) Psous K	27,76	11,38
NbPer (ACHLxCHFL xTYPLxCMBL) Psous K	38,77	NbPer (CATLxCHFL xTYPL) Psous K	27,76	11,01
NbPer (ACHLxCHFL xSTOCD) Psous K	38,66	NbPer (CHFLxTYPL xCMBL) Psous K	27,76	10,9
NbPer (CHFLxSTOCD xTYPLxCMBL) Psous K	38,48	NbPer (CATLxCHFL xTYPLxCMBL) Psous K	27,76	10,72
NbPer (CHFLxTYPL xCMBL) Psous K	37,6	NbPer (ACHLxCHFL xSTOCD) Psous K	27,75	9,85
NbPer (ACHLxSTOCD xCMBL) Psous K	37,27	NbPer (CATLxACHL xCHFLxSTOCD) Psous K	27,75	9,52
Psous K	0			

TABLE 3.10 – Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées au nombre de personnes, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par méthode de référence et par méthode d'optimisation

Finalement, en croisant toutes les variables intensives, les résultats sont présentés en table 3.11. La base de données INSEE associée avec la méthode d'optimisation améliore au mieux de 43,4% l'erreur de prédiction de la thermosensibilité. La référence utilisée est la prédiction en utilisant la puissance souscrite seule. Quant à la méthode de référence, elle améliore de 36,4%. Finalement, l'association par méthode d'optimisation améliore de 7% les erreurs de prédiction de la thermosensibilité par rapport à la méthode d'association de référence. Cette méthode présente donc bien un intérêt quantitatif. Il y a aussi un intérêt qualitatif du fait que chaque client se retrouve associé à un logement et vice-versa.

Variables du modèle optim	Amélioration de l'erreur optim (%)	Variables du modèle ref	Amélioration de l'erreur ref (%)	Apport de la méthode d'optim
Surf (TYPLxCMBL) NbPer (CHFLxTYPL) NbPi (ACHLxCHFL) Psous K	43,41	Surf (STOCD) NbPer (TYPL) NbPi (STOCD) Psous K	36,38	7,03
Surf (TYPL) NbPer (ACHL) NbPi (CHFL) Psous K	43,27	Surf (STOCD) NbPer (STOCD) NbPi (STOCD) Psous K	35,95	7,32
Surf (TYPL) NbPer (CHFL) NbPi (ACHL) Psous K	43,12	Surf (CATLxSTOCD) NbPer (CATLxTYPL) NbPi (CATLxSTOCD) Psous K	35,95	7,17
Surf (TYPL) NbPer (CMBL) NbPi (ACHL) Psous K	43,03	Surf (STOCD) NbPer (CATL) NbPi (STOCD) Psous K	35,76	7,27
Surf (TYPLxCMBL) NbPer (CHFLxTYPL) NbPi (ACHLxCMBL) Psous K	42,96	Surf (CATLxSTOCD) Nb- Per (CATLxSTOCD) NbPi (CATLxSTOCD) Psous K	35,64	7,32
Surf (CHFL) NbPer (ACHL) NbPi (TYPL) Psous K	42,95	Surf (CATLxSTOCD) Nb- Per (STOCDxTYPL) NbPi (CATLxSTOCD) Psous K	35,41	7,54
Psous K	0			

TABLE 3.11 – Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées aux différentes variables intensives, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par méthode de référence et par méthode d'optimisation

### 3.4 Conclusion et perspectives du chapitre

Dans ce chapitre, nous avons vu les différentes bases de données disponibles qui permettent d'expliquer la consommation électrique. Nous les avons classées selon les différents facteurs influant la consommation, leur échelle d'information et leur disponibilité.

Nous avons ensuite développé une méthode permettant d'associer les clients aux logements qui sont anonymisés. Nous avons montré l'intérêt de la méthode d'association en l'illustrant par l'amélioration des erreurs de prévision de la thermosensibilité des départements HTA. Nous avons aussi pu mettre en évidence l'intérêt de coupler certaines variables intensives aux variables extensives pour prédire efficacement avec un modèle linéaire la consommation annuelle, de même que la thermosensibilité. Cette méthode d'association et d'évaluation des bases de données est une première étape pour son application à l'association d'autres bases de données entre elles, pour voir par exemple l'intérêt quantitatif d'associer à cette nouvelle base la base des impôts fonciers (MAJIC) ou du répertoire des entreprises (SIRENE). De plus, des études sur la prévision d'autres indicateurs des courbes de charge (*e.g.* autres définitions de la thermosensibilité, consommation moyenne par département, consommation estivale) peuvent être effectuées. De façon plus large, la base des clients d'Enedis, associée à d'autres bases de données, peut donner lieu à des études pour mieux comprendre les paramètres les plus influents sur divers indicateurs de la consommation électrique. On peut étudier de façon plus précise des paramètres descriptifs des bâtiments à la maille IRIS (*e.g.* nombre d'étage, taille du bâtiment, année de construction, place des logements dans les bâtiments) qui pourraient expliquer l'énergie consommée relevée par Enedis sur cet IRIS. Cette approche statistique donnerait notamment des pistes sur des potentiels d'économies d'énergie. Ce type d'approche serait alors complémentaire aux approches consistant à effectuer des sondages et enquêtes avec des questionnaires et sur les factures d'énergie.



## Chapitre 4

# Construction d'un modèle de consommation électrique

### Résumé du chapitre

*La modélisation de type bottom-up de la consommation électrique est un sujet d'étude depuis des dizaines d'années. Dans le cadre de la thèse, l'objectif est de développer un modèle de consommation électrique, appelé "simulateur de courbes de charge", pouvant simuler différentes alternatives énergétiques, en particulier pour aider au dimensionnement du réseau d'Enedis. Le modèle se base sur les grandes bases de données développées au chapitre précédent. Son cahier des charges a été développé conjointement avec Enedis. L'étude bibliographique a permis de prendre connaissance des différents modèles utilisés par les auteurs. Nous présentons ensuite les modèles utilisés par le simulateur, qui ont pour objectif d'être les plus simples possible, en particulier pour pouvoir simuler relativement rapidement des régions entières. En même temps, les modèles développés sont adaptatifs aux évolutions dans les données disponibles et aux besoins de précision des simulations. Nous présentons ensuite l'implémentation du modèle que nous avons effectué de façon générique et modulaire. Enfin, nous récapitulons les étapes de la simulation et les possibilités de l'outil qui va être utilisé dans les chapitres suivants.*

## 4.1 Introduction du chapitre

Dans ce chapitre, nous présentons le modèle de simulation de la consommation électrique qui a été conçu et implémenté durant cette thèse en utilisant les grandes bases de données disponibles vues au chapitre 3. Certaines parties de ce chapitre ont fait l'objet de communications qui présentent le modèle ([38], [73]).

### 4.1.1 Objectifs

Le modèle de consommation électrique, aussi appelé "simulateur de courbes de charge", a pour but de simuler différentes alternatives énergétiques du futur. Ce terme large rend ainsi indispensable la simulation de type bottom-up client par client et appareil par appareil. Ce type de modèle peut servir notamment à évaluer les potentiels des scénarios de gestion de la demande comme le font par exemple les modèles présentés en [99], [50] [112], [44]. Les simulateurs de courbes de charge peuvent aussi servir à des fins d'évaluation de l'impact de différentes stratégies "bas carbone" dans le réseau de distribution (*e.g.* [141], [122]), ou à des fins de planification à l'échelle du réseau de distribution (*e.g.* [62], [63]). Dans le cas de la thèse, au sein du centre PERSEE, l'outil est utilisé comme générateur de consommations réalistes par client et par usage, pour toute étude le nécessitant. Un exemple d'étude utilisant l'outil de simulation est l'estimation de la capacité d'accueil de production renouvelable dans un réseau de distribution actif [89]. Le simulateur de courbes de charges de la thèse a aussi été conçu et développé conjointement avec Enedis, principal gestionnaire du réseau de distribution Français, comme outil d'aide à la planification de son réseau. La version du simulateur implémentée chez Enedis a été nommée "MOSAIC" [73]. L'outil sert notamment à Enedis à prévoir la consommation électrique dans deux types de situations [98] :

- L'évolution d'une zone existante. Si la zone est déjà connectée au réseau de distribution, le but de l'étude est d'étudier l'ampleur des modifications nécessaires sur le réseau par le changement de la demande de la zone.
- Création d'une nouvelle zone. Le réseau est alors construit de A à Z. Le but est alors de définir les infrastructures nécessaires pour connecter la zone au réseau (*e.g.* poste source, départ HTA).

Ainsi le développement de l'outil de simulation dans cette thèse a dû concilier les besoins d'outil d'aide à la planification pour Enedis, et ceux plus larges propres à la recherche et la simulation des alternatives énergétiques du futur.

### 4.1.2 Étude bibliographique des modèles bottom-up

Comme défini par Kavgic *et al.* [96], les modèles bottom-up sont construits à partir des données. Cela mène à des bases de données conséquentes pour pouvoir décrire précisément chaque appareil [138].

Le simulateur de courbes de charge développé pendant la thèse a pour objectif général de simuler de façon bottom-up, c'est à dire client par client, appareil par appareil, la consommation électrique d'un grand nombre de bâtiments à l'échelle de la France. On a pu voir au chapitre précédent la grande quantité de données disponibles et nécessaires pour simuler la consommation électrique française. L'association des bases de données entre elles pour renseigner un modèle bottom-up est aussi très spécifique. En ce sens, aucun modèle de la littérature à notre connaissance n'était conçu de façon assez adaptée à cette grande quantité de données. L'outil développé doit avoir des modèles simples, modulaires et génériques. En effet, les modèles doivent être simples pour que le

temps de calcul soit le plus court possible, et que les résultats puissent être fournis aisément à l'échelle de la France, ce pour pouvoir être évalués avec la grande quantité de données de mesures fournies par Enedis. Le modèle doit aussi être modulaire et générique dans son implémentation pour pouvoir ajouter aisément des nouveaux usages ou des nouveaux modèles plus avancés si le besoin se présentait.

Dans la littérature, beaucoup de modèles ne prennent qu'une ou deux bases de données en entrée, comme la majorité de ceux présentés par Grandjean *et al.* [86] : dans leur tableau récapitulatif, la colonne "input of the model" a rarement plus de deux éléments. Parmi ceux présentant des entrées plus élaborées, le modèle de Stokes [141] utilise de nombreuses courbes de charge élémentaires journalières difficiles à obtenir à grande échelle. Le modèle de Richardson *et al.* [122] est adapté à la simulation à l'échelle de villes, tout comme celui de Fischer *et al.* [76], mais les modèles d'occupation implémentés sont relativement complexes pour une première version de l'outil de simulation. Les modèles ayant pour but la planification urbaine [123], [120], [78], utilisent de nombreuses bases de données, mais ils sont plus adaptés à la simulation à l'échelle du quartier [44]. Les modèles Energis [82], SYNCITY [97] et Artelys Crystal City [115], sont des modèles à des fins de planification des infrastructures comme le réseau de transport ou les centrales de production électrique. Ils utilisent des bases de données importantes à différentes échelles mais ils ne sont pas basés sur des modèles bottom-up par usage spécifique à proprement parler [44]. Les modèles de Shimoda *et al.*, [137], Good *et al.* [84] et Berthou *et al.* [44] permettent la génération de la consommation électrique et d'autres types d'énergie et leurs modèles physiques des usages sont renseignés de façon stochastiques, permettant une description aisée d'un parc à grande échelle. Cependant leurs modèles des appareils sont parfois un peu complexes. Ils intègrent notamment des scénarios d'occupation complexes rendant plus difficile la scénarisation du comportement d'utilisation des usages et le temps de calcul peut aussi être affecté. Le modèle GridLAB-D développé par le département américain de l'énergie au *Pacific Northwest National Laboratory* (PNNL) [100], [55], est un modèle open source basé sur un environnement de simulation multi-agent [130]. Ce type de modèle est plus spécifique aux problématiques du réseau américain : il utilise notamment des modèles sur les différents états des appareils ainsi qu'une modélisation de type *ZIP* [129] associée. Cela permet de connaître le comportement en puissance active et réactive des appareils en fonction de la tension. Le modèle de Dickert et Schegner [62], [63], conçu pour des fins de planification du réseau de distribution allemand présente des modèles plus simplifiés qui sont proches de ceux implémentés par le simulateur de courbes de charge de la thèse.

### 4.1.3 Cahier des charges du simulateur de courbes de charge de la thèse

Nous listons ci-dessous les principales spécificités de l'outil de simulation développé.

**Simuler de façon rapide et appareil par appareil, client par client la consommation électrique d'un grand nombre de bâtiments avec un pas de temps variable.** L'objectif principal du simulateur de courbes de charge est de simuler la consommation électrique d'un grand nombre de clients en un temps relativement court en simulant la consommation de chaque appareil à un pas de temps variable pour pouvoir s'adapter aux études et aux données de comparaison disponibles. Un temps de calcul de l'ordre de la minute pour un millier de clients sur un an au pas de temps de 10 minutes avec un ordinateur de bureau classique (simulation sur un cœur, processeur de 3GHz, 8Go de RAM) nous a paru à la fois réaliste et compatible pour une utilisation efficace.

**Utiliser les différentes sources d'information provenant des bases de données précédemment décrites.** On a vu au chapitre 3 que les données peuvent être disponibles à différentes échelles. Pour répondre à cette problématique, si, lors de la simulation d'une zone, l'information nécessaire à la simulation n'est pas disponible, l'outil utilisera des valeurs par défaut renseignées par une base de données à une échelle moins fine [98].

**Compromis complexité et temps de calcul des modèles.** Les modèles devront être un compromis entre complexité des modèles physiques et rapidité des modèles statistiques, pour que les résultats de la simulation puissent être évalués aisément avec les données de mesure disponibles. Tout ajout d'un modèle plus complexe devra prouver son amélioration dans la qualité des résultats par rapport au temps de calcul ajouté.

**Programmation générique et modulaire.** L'implémentation devra être générique et modulaire afin de pouvoir aisément ajouter un nouvel usage ou de modifier le modèle d'un usage existant. L'outil étant utilisé par Enedis, ce dernier devra aussi avoir facilement le contrôle des différents paramètres par défaut.

## 4.2 Description des modèles

### 4.2.1 Secteur résidentiel

#### Choix des appareils à modéliser

Pour choisir les appareils à modéliser dans le secteur résidentiel, nous avons choisi deux critères, qui sont aussi choisis par Dickert et Schegner [62] :

- la puissance pic appelée, qui pourrait éventuellement participer à la pointe de consommation électrique de nombreux utilisateurs se servant de ce type d'appareils en même temps.
- l'énergie annuelle consommée, qui correspond à la quantité d'énergie à amener, et à la facture du client.

La figure 4.1 place les principaux appareils résidentiels dans un diagramme représentant la puissance pic appelée en fonction de l'énergie moyenne annuelle consommée par foyer équipé. Les cercles verts sur la figure entourent les usages retenus pour une première version du simulateur de courbes de charge, et correspondent aux usages à forte énergie consommée et/ou forte puissance appelée. Les chiffres sont obtenus à partir de campagnes de mesures en France, et sont détaillés en annexe 8.2.1 au tableau 8.10. Notons que pour plus de lisibilité, les usages importants qui sont l'éclairage, le chauffe-eau électrique, la piscine extérieure, le chauffage électrique et la climatisation ne sont pas représentés, mais seront aussi modélisés.

Finalement, la première version du simulateur de courbes de charge implémente les usages résidentiels suivants : machine à laver, sèche-linge, lave-vaisselle, ordinateur, éclairage, piscine, four, plaques de cuisson, réfrigérateur, congélateur, chauffe-eau électrique, chauffage électrique, climatisation. Afin de simuler les usages restants, nous avons ajouté deux autres usages : électrodomestique et autres usages.

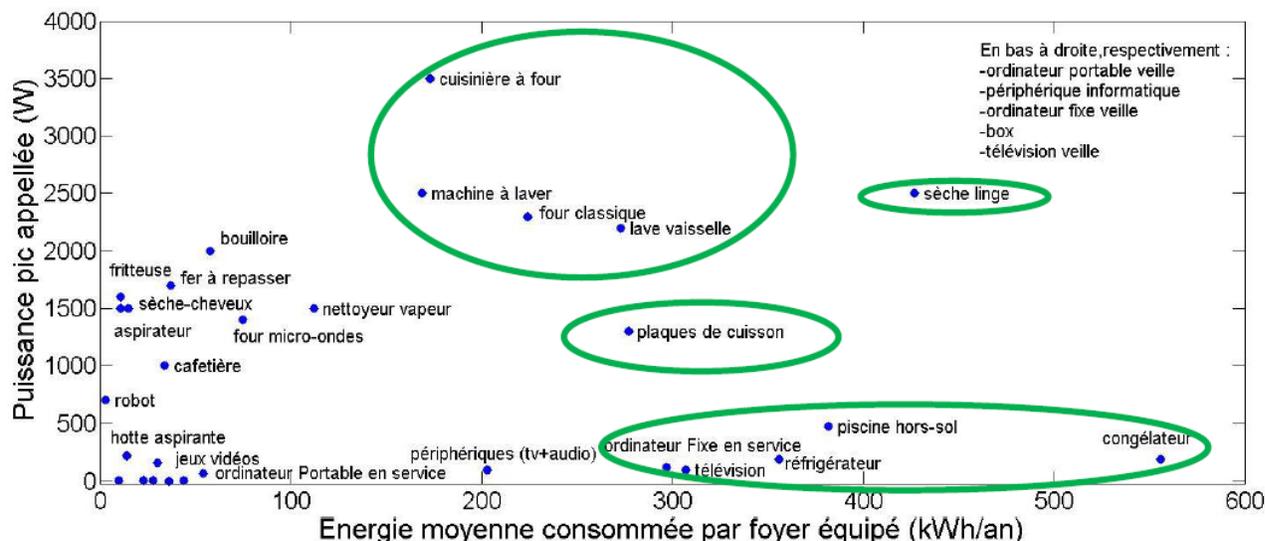


FIGURE 4.1 – Principaux appareils résidentiels placés par leur puissance pic appelée en fonction de l'énergie moyenne annuelle consommée par foyer équipé

### Détails de modèles

Comme nous l'avons vu, nous avons eu pour objectif les modèles les plus simples possible. Nous avons 3 catégories d'usages résidentiels, selon leurs modélisations : par blocs de puissance constante, par blocs de puissance périodique et enfin les usages thermiques. Ces derniers font l'objet d'une sous section au même niveau que les modèles résidentiels et tertiaires car ils sont modélisés de la même manière dans les deux cas. Notons que la modélisation des appareils effectuée dans cette thèse, comme la plupart des modèles de consommation bottom-up, prennent l'hypothèse d'une consommation par usage indépendante de la tension ou la fréquence d'alimentation des appareils. Si par exemple l'utilisation de l'outil de simulation de courbes de charge s'oriente vers des estimations fines du potentiel de gestion de la demande, mettant en œuvre des stratégies de modification de la tension, cette hypothèse pourrait devenir limitante. Dans ce cas, les modèles des appareils implémentés pourront être modifiés pour prendre ces paramètres en considération. Une modélisation de type ZIP par appareil, comme par exemple décrite par Schneider *et al.* [129], pourrait alors être effectuée. De même, nous ne modélisons que la puissance active consommée, mais toute étude nécessitant aussi la puissance réactive pourrait être menée en raffinant aussi les modèles d'appareils.

**Les usages modélisés par blocs de puissance constante.** Ces usages sont lancés un certain nombre de fois et leur cycle élémentaire est modélisé par une succession de puissance constante, tout comme le modèle de Dickert et Schegner [63]. Un exemple est donné à titre illustratif en figure 4.2. On voit deux lancements d'une même machine à laver sur un jour. La durée des cycles est différente, pour modéliser les différents programmes. Cependant les puissances instantanées des différents blocs (chauffage de l'eau, lavage et essorage) sont identiques pour les deux cycles, car il s'agit du même moteur et de la même résistance de chauffage.

Les usages suivant ont été choisis pour être modélisés avec ce type de modèle : machine à laver, sèche-linge, lave-vaisselle, ordinateur, éclairage, piscine, four, plaques de cuisson, chauffe-eau électrique, électrodomestique et autres usages.

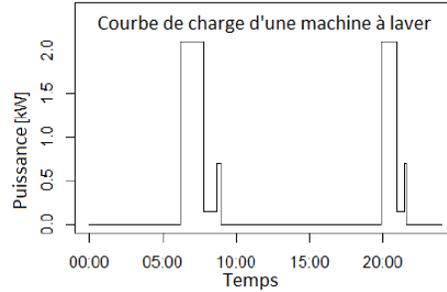


FIGURE 4.2 – Tracé de la courbe de charge d'une machine à laver modélisée par blocs de puissance constante, lancée deux fois dans la même journée

De façon plus formelle, la consommation des appareils modélisés par blocs de puissance constante d'un client  $j$  s'exprime à un instant  $t$  par :

$$P^j(t) = \sum_{a=1}^{A^j} \sum_{n=1}^{N_a^j} \sum_{m=1}^{M_a} P_{a,m}^j \mathbb{1}_{d_{a,n,m}}(t - T_{a,n,m}) \quad (4.1)$$

Avec  $A^j$  le nombre d'appareils du client  $j$ ,  $M_a$  le nombre de bloc modélisant l'appareil  $a$ ,  $N_a^j$  son nombre de lancements.

On a ici trois variables dépendant de l'appareil, du client et du bloc de puissance considéré : la puissance du bloc  $P_{a,m}^j$ , la durée du bloc  $d_{a,n,m}^j$ , et la date de lancement du bloc  $T_{a,n,m}^j$ .

Pour déterminer toutes ces variables, ce sont les données provenant des campagnes de mesures qui nous ont renseignées. Dans l'idée de faire le modèle le plus simple possible, nous avons cherché à retrouver les distributions des campagnes de mesures en tirant les valeurs au hasard dans ces répartitions.

Typiquement pour la machine à laver, pour la distribution des puissances, c'est du graphe figure 4.3 que nous avons pris les valeurs par défaut pour les distributions de puissance. On voit bien sur cette figure les 3 pics, correspondant aux trois étapes majeures d'un cycle. Notre hypothèse a été de faire un tirage dans la distribution de la figure 4.3 pour la puissance de lavage dans celle centrée autour d'environ 150W, pour la puissance d'essorage celle autour de 650W, et enfin de chauffage autour de 2100W.

Nous avons fait de même pour la durée, et ce pour chaque appareil. Les sources des données pour chaque type d'appareils sont décrites en [40]. A noter une spécificité pour l'éclairage dont la puissance installée dépend de la surface. On installe d'abord par maison un nombre de points lumineux selon la surface, selon une répartition issue de campagnes de mesures [66], puis une puissance par point lumineux et la multiplication des deux donne la puissance totale installée pour le logement donné.

Concernant le nombre de lancements pendant la période de temps à considérer, nous avons fait dépendre celui-ci du nombre de personnes, en suivant toujours ce qui a été observé dans la littérature des campagnes de mesures française. La figure 4.4 est un graphe typique nous servant de données d'entrée pour renseigner la relation entre le nombre de lancements et le nombre de personnes. A noter sur cette figure que malgré le très faible coefficient de corrélation, c'est la consommation par habitant en fonction du nombre d'habitants qui est représentée. Malgré la forte dispersion, la tendance est d'avoir en moyenne une consommation par habitant qui diminue sensiblement avec le nombre d'habitants. Ainsi dans le simulateur, pour suivre les résultats de cette étude, nous avons paramétré la consommation du lave-linge par foyer comme étant "légèrement moins que proportionnelle" au nombre d'habitants, suivant la loi de cette même figure.

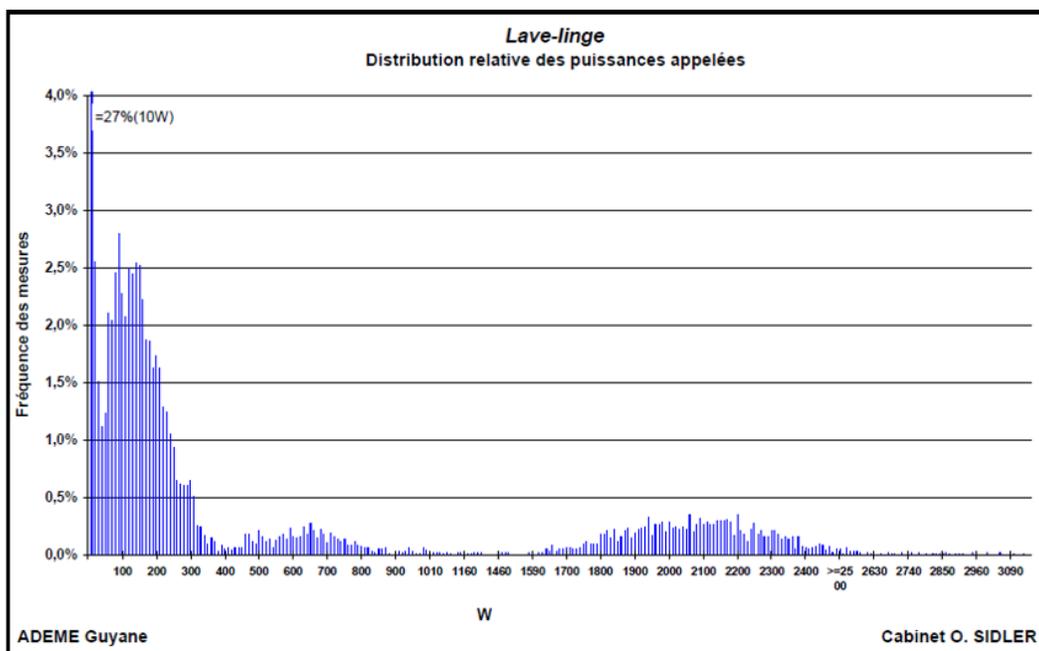


FIGURE 4.3 – Distribution relative des puissances appelées de 85 lave-linges suivis pendant 24 jours [33]

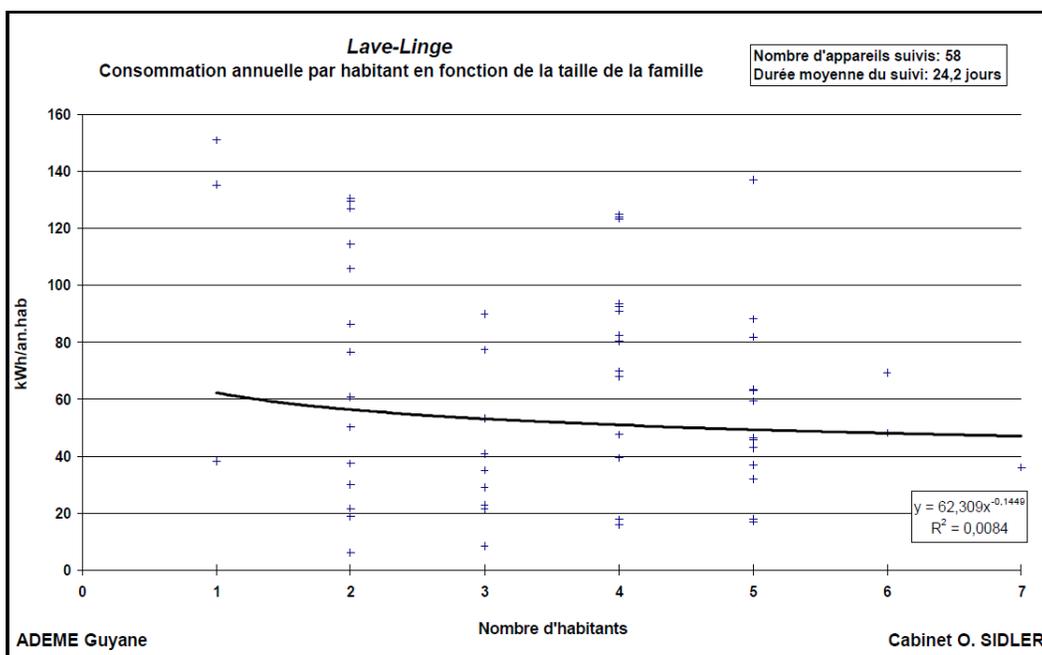


FIGURE 4.4 – Consommation annuelle par habitant de l'appareil lave-linge en fonction de la taille de la famille [33]

Pour les dates de départ, leur tirage est effectué en construisant une densité de probabilité issue du croisement de profils journaliers, hebdomadaires et mensuels observés, comme illustré à la figure 4.5.

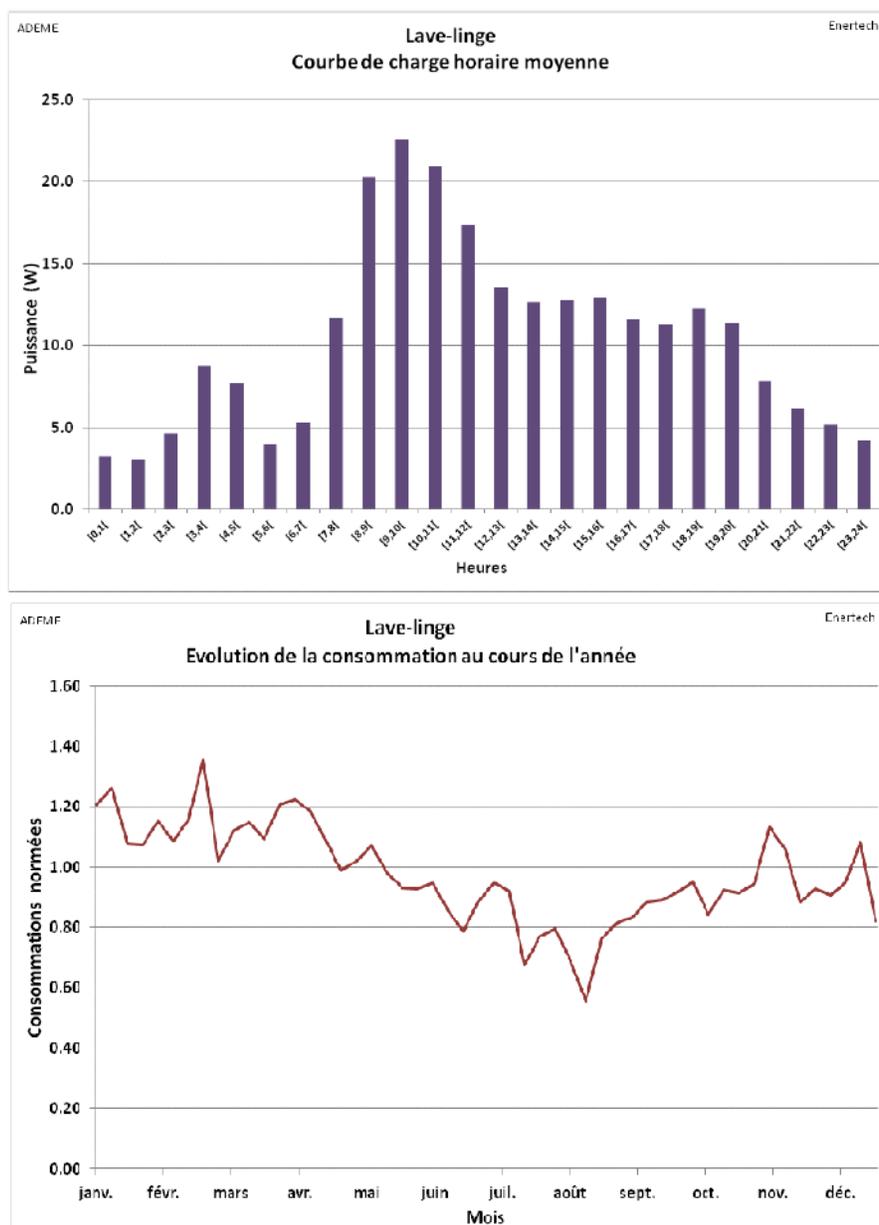


FIGURE 4.5 – Profils horaire et hebdomadaire des machines à laver mesurées pendant la campagne de mesures ADEME en 2016 [32]

Concernant la présence ou non d'un appareil chez le client, ce sont les statistiques nationales sur chaque appareil qui vont conditionner la loi du tirage de l'appareil chez le client. Ces chiffres sont disponibles auprès de l'INSEE comme on l'a vu au 3.2.6.

Finalement, le tableau 4.1 récapitule comment chaque variable du modèle par bloc de puissance constante est renseignée.

Variable	Méthode de détermination des variables par client	Modèle utilisé à chaque lancement	Base de données utilisée
Nombre de bloc de puissance	Selon l'appareil, par rapport à sa courbe unitaire classique	Même nombre de bloc à chaque lancement	Campagne de mesures [33]
Puissance de chaque bloc	Tirage par client selon la répartition des puissances mesurées. Selon la surface pour l'éclairage	Même puissance à chaque lancement	Campagne de mesures [33]
Durée de chaque bloc	Attribution par client d'une répartition selon les campagnes de mesures	Tirage d'une durée dans la répartition à chaque lancement	Campagne de mesures [52], [28], [28],[32]
Date de lancement	Croisement des profils journaliers, hebdomadaires, mensuels	Tirage de la date selon une loi de probabilité issue des profils	Campagne de mesures [52], [28], [28],[32]
Nombre de lancement	Loi selon le nombre de personnes	-	Campagne de mesures [33]
Possession ou non de l'appareil	Tirage selon les statistiques nationales	-	Taux de possession des appareils (voir 3.2.6)

TABLE 4.1 – Récapitulatif des variables utilisées dans le modèle par bloc de puissance constante et comment celles-ci sont déterminées en lien avec les bases de données

**Les usages modélisés par blocs de puissance périodique.** Ce modèle correspond aux usages qui ont un comportement périodique. Le réfrigérateur et le congélateur sont modélisés ainsi. Les paramètres qui définissent ce modèle sont les suivants : puissance, durée, période et rapport cyclique.

La puissance installée par usage par blocs de puissance périodique est renseignée comme pour le modèle précédent, avec des mesures faites sur des appareils. Le tirage des puissances en fonctionnement est effectué dans la distribution mesurée. Un exemple de distribution est illustré pour le réfrigérateur en figure 4.6. Ces puissances sont ensuite pondérées par le nombre de personnes pour obtenir une loi de consommation proche de celle mesurée, comme l'illustre la figure 4.7. Ceci traduit le fait qu'une famille plus nombreuse aura tendance à avoir un appareil plus gros, donc avec une puissance installée plus importante.

En ce qui concerne les fréquences des cycles, le tirage est aussi effectué dans les répartitions des fréquences mesurées, qui varie de façon uniforme dans un petit intervalle à chaque ouverture de l'appareil.

Pour le rapport cyclique, la répartition de la valeur moyenne d'un logement à l'autre est déterminée afin d'avoir un parc d'appareils simulé dont les consommations annuelles sont les mêmes que celles mesurées (*e.g.* pour le réfrigérateur en figure 4.8). Le rapport cyclique est aussi variable au sein d'un même logement selon l'heure, afin d'obtenir une courbe de charge horaire proche de celle mesurée, comme illustré en figure 4.9.

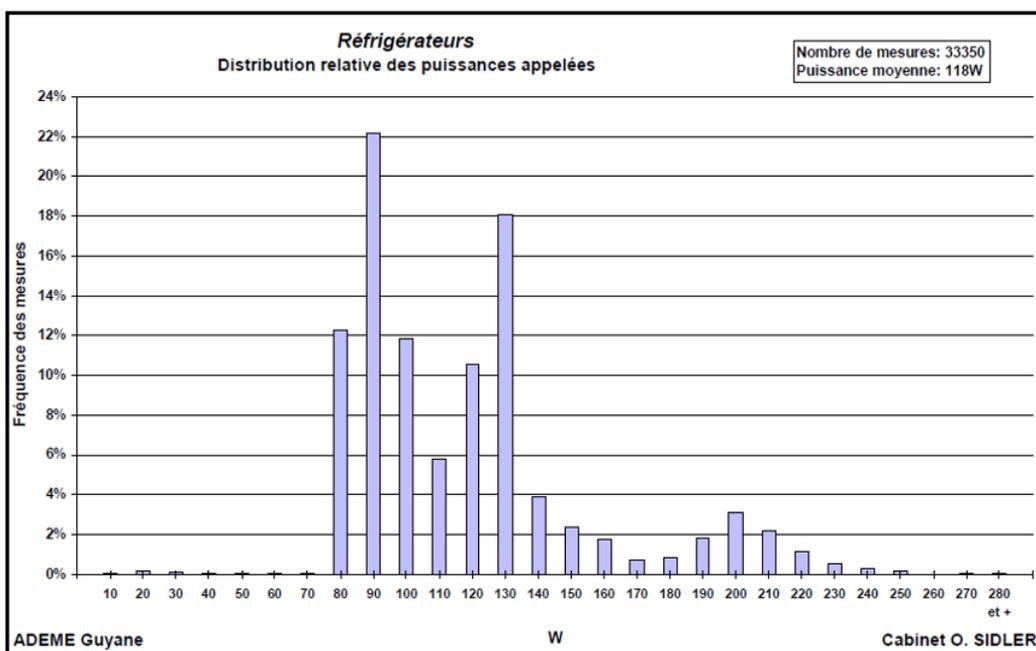


FIGURE 4.6 – Distribution relative des puissances appelées de 16 réfrigérateurs suivis pendant 22 jours [33]

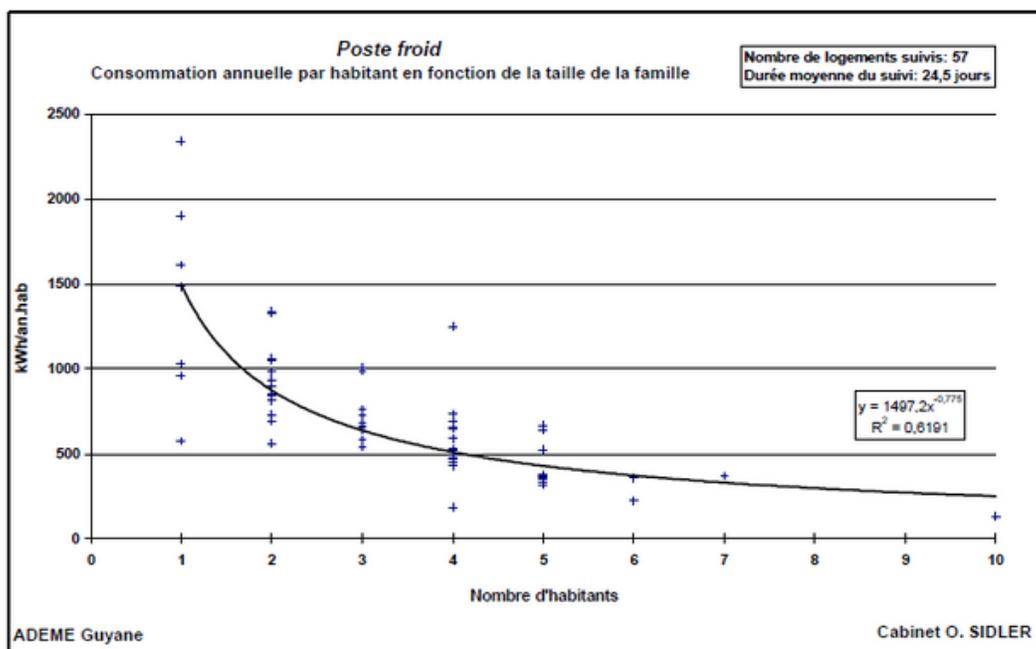


FIGURE 4.7 – Evolution de la consommation par rapport au nombre de personnes pour le réfrigérateur [33]

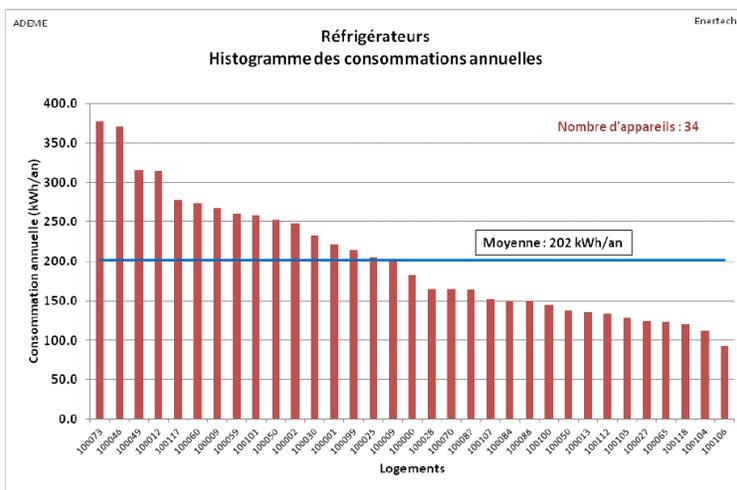


FIGURE 4.8 – Histogramme des consommations annuelles des réfrigérateurs [32]

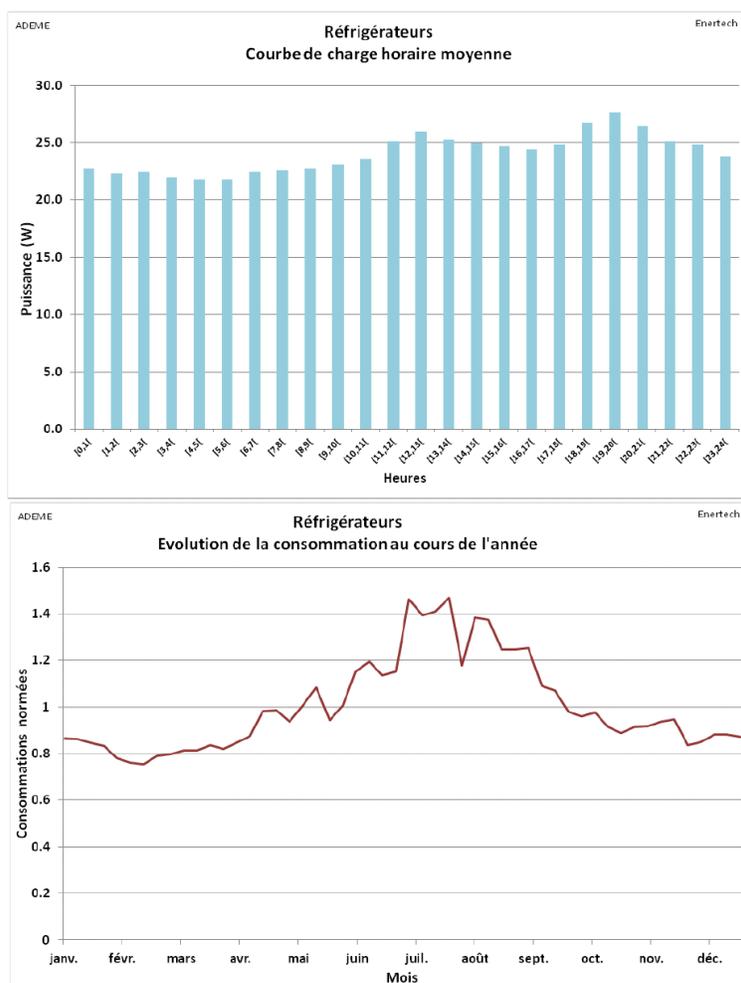


FIGURE 4.9 – Profils horaire et hebdomadaire des réfrigérateurs mesurés pendant la Campagne de mesures ADEME en 2016 [32]

Finalement, le tableau 4.2 récapitule comment chaque variable du modèle par bloc de puissance constante est renseignée.

Variable	Méthode de détermination des variables par client	Modèle utilisés à chaque lancement	Base de données utilisée
Puissance en fonctionnement et en veille	Tirage par client selon la répartition des puissances mesurées, et selon le nombre de personnes	Même puissance à chaque lancement	Campagne de mesures [33]
Période de chaque cycle	Attribution par client d'une valeur selon les campagnes de mesures	Fluctation dans un petit intervalle uniforme à chaque ouverture de l'appareil	Campagne de mesures [32]
Rapport cyclique	Attribution par client d'une valeur moyenne pour obtenir les répartitions des consommations annuelles selon les campagnes de mesures	Fluctations pour obtenir les courbes de charges moyennes	Campagne de mesures [32]
Possession ou non de l'appareil	Tirage selon les statistiques nationales	-	Taux de possession des appareils (voir 3.2.6)

TABLE 4.2 – Récapitulatif des variables utilisées dans le modèle par bloc de puissance périodique et comment celles-ci sont déterminées en lien avec les bases de données

### 4.2.2 Secteur tertiaire

Les études par usage dans ce secteur sont assez peu abondantes, comme nous l'avons vu au 3.2.6. Nous avons modélisé 7 types de clients tertiaires, correspondant aux études que nous avons trouvées sur le sujet.

#### L'éclairage public

Enedis a accès au nom de ses clients. Certains d'entre eux sont des "éclairages publics" et il nous a paru simple de proposer un premier modèle pour simuler la courbe de charge de ces derniers. Le modèle consiste à lancer les éclairages publics quand la luminosité devient inférieure à un seuil. Un tel modèle a été implémenté en utilisant la variable irradiation solaire de la base MERRA. Pour estimer la puissance installée  $P_{installée}^{EP}$ , cela se fait sur l'année correspondante au relevé de compteur Enedis sur la consommation annuelle  $E^{EP}(année)$ . Avec le modèle précédant on peut estimer la durée de fonctionnement sur l'année mesurée notée  $D^{EP}(année)$  et on a alors :

$$P_{installée}^{EP} = \frac{E^{EP}(année)}{D^{EP}(année)} \quad (4.2)$$

#### Bureaux, commerces, hôtels, restaurant

Nous avons modélisé les grandes catégories mesurées par les campagnes de mesures, comme vu au tableau 3.4 dans la partie 3.2.6.

La démarche que nous avons adoptée est d'abord de regarder dans les études quels sont les usages les plus consommateurs d'électricité et de les classer en grande catégories d'usages. Le tableau 4.3 recense les différents usages modélisés par l'outil de simulation pour le secteur tertiaire dans sa version de juillet 2017 (v0.1.15).

Type	Grandes catégories d'usages modélisés (% dans la consommation totale)	Etude considérée
Tout type	Chauffage, Climatisation	
Bureau	Éclairage (24), ventilation (29), ordinateur (5,9), Constant bureautique (7,3), Périphériques (12,2), autres usages (21,6)	[31]
Commerce alimentaire	Éclairage (27), froid constant (43), froid variable (13), autres usages (17)	[29]
Commerce non alimentaire	Éclairage (61), autres usages (39)	[29]
Restaurant	Éclairage (11), ventilation (7), lavage (19), froid (47), cuisson (12)	[125]
Hôtel	Éclairage (20), ventilation (27), ECS (17), Cuisine (11) autres usages (25)	[69]

TABLE 4.3 – Différentes campagnes de mesures utilisées pour estimer les appareils présents dans les différentes catégories de clients tertiaires

Ensuite, la démarche de modélisation choisie est d'abord d'installer une certaine puissance de chaque usage par client. Nous avons fait dépendre les puissances installées de façon linéaire à la

surface du bâtiment simulé, avec les chiffres pris dans les campagnes de mesures de puissance installée par unité de surface. Ensuite, ces puissances sont lancées de façon à suivre un profil issu de Campagne de mesures.

Un exemple des hypothèses choisies pour les bureaux est donné dans le tableau 4.4.

Usage	Consommation annuelle dans l'étude [31] (kWh/an)	Pourcentage consommation annuelle (%)	Consommation surfacique (kWh/m <sup>2</sup> /an)	Profil utilisé
Éclairage	16380	24	6,3	Profil éclairage bureau (figure 4.10)
Ventilation	19793	29	7,6	Constant
Ordinateur	4006	6	1,5	Profil bureautique (figure 4.11)
Fixe bureautique	5002	7	1,9	Constant
Périphérique bureautique	8345	12	3,2	Profil bureautique (figure 4.11)
Autres usages	14725	22	5,6	Constant
Total	68250	100	26,1	

TABLE 4.4 – Chiffres et profils retenus pour modéliser les différents usages pour les bureaux

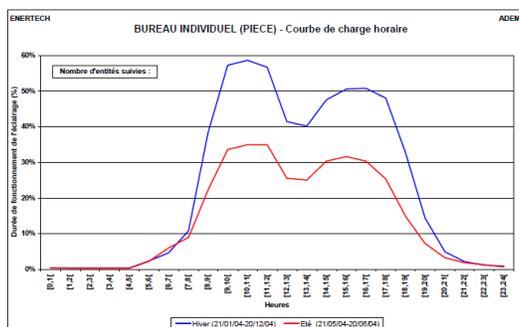


FIGURE 4.10 – Profil journalier de l'éclairage des bureaux [30]

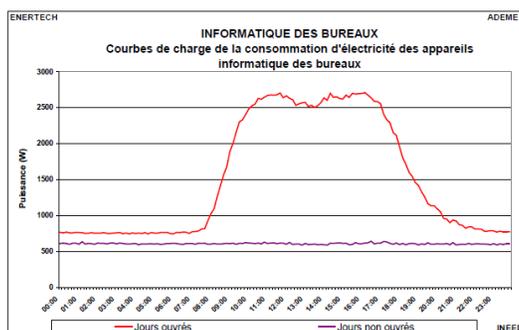


FIGURE 4.11 – Profil journalier des usages de bureautique [31]

Nous avons effectué de même pour les autres types de tertiaire. Les détails des différents modèles sont disponibles en [40]. Notons que pour introduire de la diversité dans les profils, particulièrement dans les commerces, nous avons effectué différents tirages dans les horaires d'ouverture et de fermeture, ainsi qu'introduit une probabilité d'être fermé ou non le midi ou le samedi et le dimanche.

### Véhicules électriques

Un premier modèle simplifié de véhicule électrique a été développé. Celui-ci assigne une puissance installée par défaut de 3kW, avec une durée variable à chaque lancement et un nombre de lancement dépendant du nombre de personnes. Ces valeurs par défaut ont été rentrées à partir de mesures de profils extraites des données de la compagnie Pecan Street [93]. 87 mesures de véhicules électriques sur un an au pas de temps 15 minutes ont été utilisées pour connaître la répartition du nombre de lancement et des durées. Concernant les profils pour les dates de lancement, les données PECAN street ont un biais car certaines heures étaient privilégiées par des incitations tarifaires, et les comportements au Texas sont sans doute différentes de ceux en France. Ainsi nous avons préféré utiliser dans un premier temps les profils déterminés par RTE dans son bilan prévisionnel [127] comme illustré en figure 4.12. Selon les études effectuées avec le simulateur, on choisira les profils correspondants. Selon RTE [127] :

- “Profil « charge naturelle » : les véhicules sont rechargés en fonction du besoin, sans contrainte réglementaire ni incitation tarifaire pour décaler les heures de recharge ; dans ce cas, l'impact sur la pointe journalière est particulièrement marqué ;
- Profil « signal tarifaire » : les véhicules sont rechargés pendant les heures creuses à la réception d'un signal tarifaire par simple incitation tarifaire ;
- Profil « BMS » : par rapport au scénario “signal tarifaire”, une partie des recharges est effectuée en mode BMS (Battery Management System), avec un report encore plus marqué sur les heures nocturnes.”

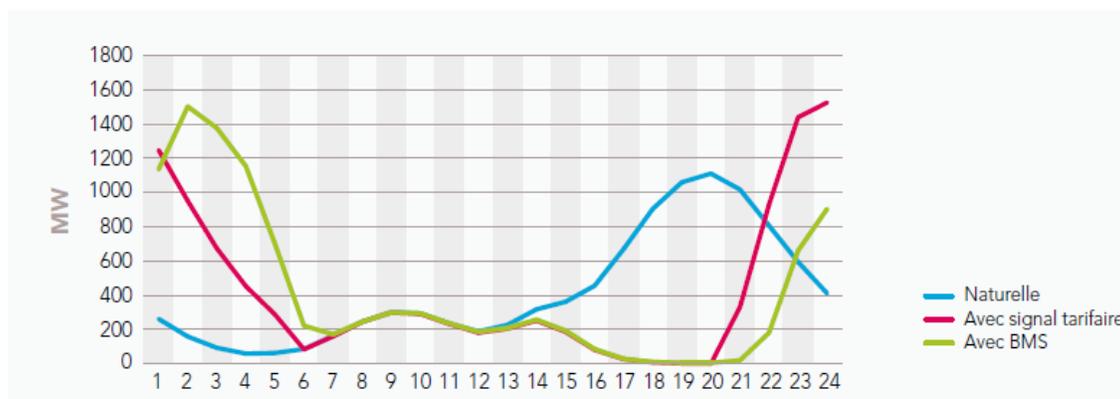


FIGURE 4.12 – Courbe de charge d'un jour ouvrable de janvier pour un parc d'un million de véhicules électriques [127]

#### 4.2.3 Modèle de chauffage et de climatisation

La modélisation de ces deux usages est commune au secteur résidentiel et tertiaire. Le modèle de chauffage et de climatisation nécessite un modèle thermique du bâtiment, ainsi qu'un modèle de régulation.

**Modèle thermique des bâtiments.** De nombreux modèles thermiques des bâtiments existent [43]. Notre choix s'est orienté vers le plus simple existant, et le modèle mono-zone par maison est adapté pour remplir les objectifs d'un simulateur de consommation électrique bottom-up [44]. Un modèle simple mono-zone existant est le modèle à 4 résistances thermiques équivalentes et 2 capacités thermiques équivalentes, encore appelé R4C2 [45]. Notre premier modèle travail n'étant pas d'avoir un modèle simulant précisément le confort des occupants et les stratégies d'effacement, et afin de réduire encore le nombre de paramètres à renseigner, nous prenons l'hypothèse d'un modèle R1C1, qui correspondrait à un modèle R4C2 décrit dans [45] dont la capacité de l'air intérieur et du mobilier a été supposée négligeable. Le modèle thermique choisi schématisé sous la forme d'analogie électrique est représenté en figure 4.13.

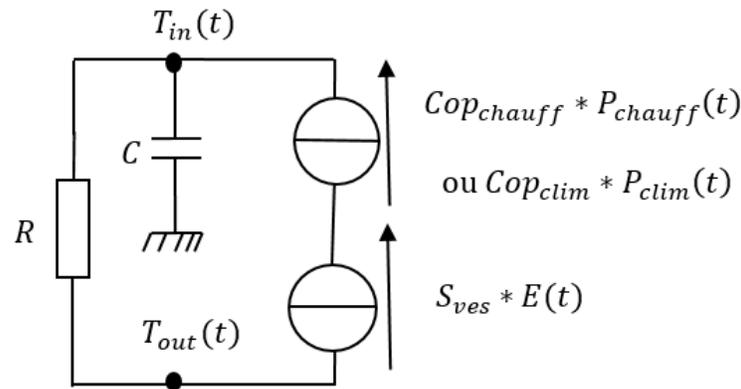


FIGURE 4.13 – Schéma du modèle thermique utilisé par maison, sous forme d'analogie électrique

Sur cette figure,  $R$  ( $^{\circ}C/W$ ) représente la résistance thermique modélisant les déperditions entre l'intérieur et l'extérieur du bâtiment.  $C$  est la capacité thermique ( $J/^{\circ}C$ ), modélisant la chaleur stockée dans le bâtiment.  $P$  est la puissance électrique ( $W$ ) consommée par le système de chauffage, qu'il s'agisse de convecteurs ou d'une pompe à chaleur.  $Cop$  est le coefficient de performance du moyen de chauffage.  $E$  est l'irradiation solaire sur plan horizontal orienté sud ( $W/m^2$ ).  $S_{ves}$  la surface de vitrage équivalent sud ( $m^2$ ).  $T_{in}$  est la température intérieure ( $^{\circ}C$ ) moyenne du bâtiment,  $T_{out}$  la température extérieure ( $^{\circ}C$ ). L'équation différentielle (issue de les lois de Fourier) régissant l'évolution des différentes variables est la suivante :

$$\frac{dT_{in}}{dt}(t) + \frac{1}{RC}[T_{in}(t) - T_{out}(t)] = \frac{1}{C}[Cop \times P(t) + S_{ves}E(t)] \quad (4.3)$$

A noter que dans ce modèle nous ne prenons pas en compte les apports internes liés aux occupants et aux appareils.

**Estimation du paramètre  $R$**  On modélise chaque client par un ensemble de conductance en parallèle :  $U$  en  $W/m^2/^{\circ}C$ , représentant les pertes thermiques par : les murs, la toiture, le plancher, les baies vitrées, le renouvellement d'air. Ces conductances étant en parallèle, il suffit de les ajouter pour obtenir la conductance globale du client.

Plusieurs positions sont possibles pour un client au sein d'un bâtiment. Nous en avons recensé 10 :

- les 6 premières sont celles illustrées à la figure 4.14 ;
- 7 : tout l'étage du rez de chaussée ;

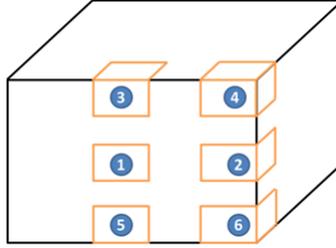


FIGURE 4.14 – Différentes positions possibles d'un client dans un bâtiment

- 8 : tout un étage intermédiaire ;
- 9 : tout le dernier étage ;
- 10 : l'ensemble du bâtiment.

On suppose que si le client est dans un appartement, alors les murs ou plancher qu'il partage avec son voisin auront une conductivité infinie (*i.e.* cela revient à supposer que le voisin a atteint lui aussi la température de consigne et que la température du mur/plancher les séparant est homogène).

Les équations pour le calcul des  $R$  (ou son inverse  $K$ ) sont alors, par exemple, dans le cas où le client est en position 10 (ensemble du bâtiment) :

$$K_{10}(\text{année}) = \frac{1}{R_{10}} = U_{mur}(\text{année}) \times (Pe \times H - 4S_{ves}) + U_{plancher}(\text{année}) \times Surf + U_{toiture}(\text{année}) \times Surf + U_{baie}(\text{année}) \times 4S_{ves} + A_{ra}(\text{année}) \times V \quad (4.4)$$

En position 6, ce serait :

$$K_6(\text{année}) = \frac{1}{R_6} = U_{mur}(\text{année}) \times 2\left(\frac{Pe}{4} \times H - S_{ves}\right) + U_{plancher}(\text{année}) \times Surf + U_{baie}(\text{année}) \times 2S_{ves} + A_{ra}(\text{année}) \times V \quad (4.5)$$

Avec :

- année : année de construction ou de dernière rénovation
- $Pe$  : périmètre
- $H$  : hauteur sous plafond
- $Surf$  : surface habitable
- $S_{ves}$  : surface de vitrage par pan de mur extérieur équivalent sud
- $V$  : volume des pièces

Les conductances sont données dans la figure suivante. Elles sont issues de méthodes 3CL DPE de l'Anah (figure 4.15). L'équivalence entre niveau de performance et l'année de construction ou de dernière rénovation est donnée en figure 4.16.

**Estimation du paramètre C** Pour déterminer C, la réglementation thermique de 2005 propose une partie s'intéressant à l'inertie [23] qui donne la capacité d'une habitation en fonction de sa classe d'inertie. Ceci est illustré en figure 4.17. Dans cette figure,  $A_{niv}$  est la surface habitable du niveau considéré,  $Am$ , la surface des murs déduite des ouvertures. Ainsi dans notre cas on prendra

	Niveau de performance							
	0	1	2	3	4	5	6	7
$U_{\text{mur}} \text{ (W/m}^2\text{/}^\circ\text{C)}$	2,5	1	0,45	0,4	0,36	0,3	0,2	0,15
$U_{\text{toiture}} \text{ (W/m}^2\text{/}^\circ\text{C)}$	2,5	0,4	0,25	0,23	0,2	0,17	0,15	0,1
$U_{\text{plancher}} \text{ (W/m}^2\text{/}^\circ\text{C)}$	2	0,9	0,6	0,4	0,3	0,25	0,20	0,15
$U_{\text{baïe}} \text{ (W/m}^2\text{/}^\circ\text{C)}$	4	2,6	2,1	1,6	1,2	1	0,9	0,8
$A_{\text{ra}} \text{ (W/m}^3\text{/}^\circ\text{C)}$	0,3	0,21	0,2	0,15	0,12	0,10	0,09	0,08

FIGURE 4.15 – Équivalence entre le niveau de performance et les valeurs de conductivité thermique

Année de construction	Niveau de performance
Avant 1975	0
Entre 1975 et 1988	1
Entre 1989 et 2000	2
Entre 2001 et 2005	3
Entre 2006 et 2012	4
Bâtiment en RT2012	5
BBC -20%	6
Bâtiment Passif	7

FIGURE 4.16 – Équivalence entre l'année de construction et le niveau de performance

Inq	Cm/A <sub>nv</sub> (kJ/m²K)									
	80 à 109	110 à 164	165 à 219	220 à 259	260 à 309	310 à 329	330 à 349	350 à 369	370 à 499	500 et plus
1,0 à 1,4	t. lég.	t. lég.	t. lég.	t. lég.	léger	léger	léger	léger	léger	léger
1,5 à 1,9	t. lég.	t. lég.	léger	moy.						
2,0 à 2,4	t. lég.	léger	moy.	moy.	moy.	moy.	moy.	L	L	L
2,5 à 2,9	t. lég.	léger	moy.	moy.	L	L	L	L	L	L
3,0 à 3,4	t. lég.	léger	moy.	L	L	L	L	L	L	TL
3,5 à 3,9	t. lég.	léger	moy.	L	L	L	L	L	TL	TL
4,0 à 4,4	t. lég.	léger	moy.	L	L	L	L	TL	TL	TL
4,5 à 4,9	t. lég.	léger	moy.	L	L	L	TL	TL	TL	TL
5,0 et plus	t. lég.	léger	moy.	L	L	TL	TL	TL	TL	TL

Classe d'inertie : t. lég. = très légère-; moy. = moyenne-; L = lourde-; TL = très lourde

FIGURE 4.17 – Extrait de la réglementation Th-I 2005 : correspondance entre la classe d'inertie, la capacité de l'habitation par unité de surface Cm/Aniv, et Am/Aniv[23]

$A_{niv} = Surf$ , la surface du logement, et  $Am = S_{mur} - 4S_{ves}$ . Ainsi simplement avec ce tableau, en supposant une classe d'inertie donnée (de très légère à très lourde), on en déduit une capacité thermique équivalente par logement  $C = Cm$ .

Chaque logement se verra attribuer dans un premier temps la valeur par défaut d'une inertie moyenne.

**Estimation des paramètres météorologiques.** Il s'agit des variables  $T_{out}$  et  $E$ . Ces dernières sont renseignées par client à l'aide de la base de données sur la météorologie vue au 3.2.8.

**Estimation des autres paramètres.** Ci-dessous sont listés les autres paramètres pour modéliser le comportement thermique du bâti lié au client, avec la manière avec laquelle les valeurs sont déterminées.

L'année de construction ou de dernière rénovation peut être renseignée par la base logement de l'INSEE, ou encore par la base MAJIC (impôts fonciers).

Le périmètre du logement  $Pe$  peut être renseigné par la base BDTOPO de l'IGN. En l'absence de valeurs, celle-ci peut être déterminée en supposant le bâtiment du client carré. On aura alors  $Pe = 4\sqrt{Surf}$ .

La hauteur sous plafond  $H$  elle aussi peut être estimée via la base BDTOPO de l'IGN, qui donne la hauteur des bâtiments. En l'absence d'information, nous prenons comme valeur 3m.

La surface habitable peut être renseignée par la base logement de l'INSEE, la base MAJIC (impôts fonciers) ou encore la base BDTOPO de l'IGN.

La surface vitrée sud équivalente  $S_{ves}$  est notamment définie dans la réglementation thermique de 2005 [24], en prenant en compte les différentes orientations. Dans notre cas, nous prendrons une valeur empirique de 1/16 de la surface des murs comme valeur par défaut.

Le volume des pièces  $V$  est calculé comme étant le produit de la surface habitable par la hauteur.

**Modélisation de la régulation du chauffage et de la climatisation.** Le cas de la modélisation de la régulation du chauffage est développé ci-dessous. Celui de la climatisation est similaire sauf que la puissance électrique utilisée sert à extraire de la chaleur du local, à l'inverse du chauffage qui en apporte.

Chaque client possédant un chauffage électrique a une certaine température de consigne (prise par défaut à la valeur  $19^{\circ}\text{C}$ ) pouvant prendre un profil différent selon le moment de la journée et de l'année. La puissance installée de chauffage est prise par défaut comme pouvant assurer le maintien de la température de consigne avec un écart de  $30^{\circ}\text{C}$  avec l'extérieur. L'asservissement de la puissance du chauffage électrique est assuré par un gain proportionnel, choisi par défaut de façon à ce que le chauffage soit à pleine puissance pour une différence de température entre l'intérieur et la consigne de  $1^{\circ}\text{C}$ . La température extérieure et l'irradiation solaire sont des entrées du modèle, provenant de données météorologiques, et la température intérieure est calculée en résolvant l'équation 4.3. Pour éviter les surchauffes en cas d'apports solaires trop importants, un seuil de température intérieure maximale est fixé à  $30^{\circ}\text{C}$  par défaut.

#### 4.2.4 Bilan des données d'entrée nécessaires pour les modèles et leur influence

Nous avons vu la description des différents modèles, et les différents paramètres d'entrée de ceux-ci. La plupart de ces paramètres est fixée à des valeurs par défaut prises dans la littérature, mais certaines informations sont nécessaires pour la simulation.

Concernant le chauffage électrique, la figure 4.18 récapitule les données nécessaires à sa simulation ainsi que leur influence sur la consommation de l'usage.

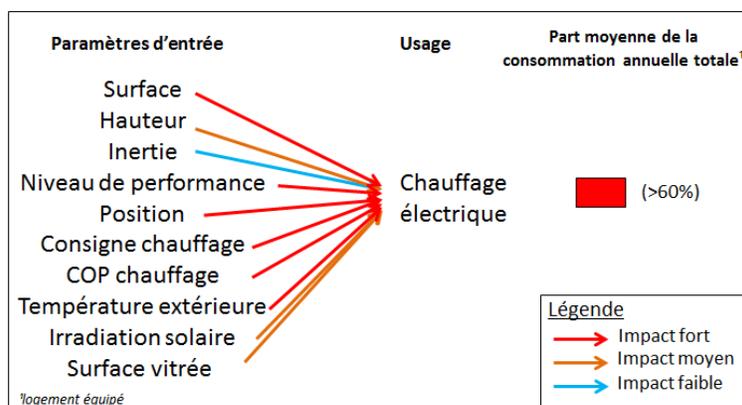


FIGURE 4.18 – Récapitulatif des caractéristiques du modèle de chauffage

Pour les usages du secteur résidentiel autres que le chauffage, ceux-ci sont dépendants de trois paramètres d'entrée principaux : la surface, le nombre de personnes, et le type de résidentiel (*i.e.* maison individuelle, immeuble collectif ou résidence secondaire). La dépendance de la consommation de chaque usage en fonction du nombre de personnes est donnée en figure 4.19. Notons qu'un nombre de personnes de 0 signifie que le logement est une résidence secondaire. La dépendance de la consommation de chaque usage en fonction de la surface est donnée en 4.20. A noter que ces résultats sont effectués avec le simulateur de courbes de charge, après la calibration avec les données CEREN, vu à la section 5.3. Les résultats sont obtenus avec la simulation de 1000 clients résidentiels sur une zone de Brest.

Concernant les modèles du tertiaires, hors chauffage et climatisation, les données nécessaires sont le type de tertiaire (*e.g.* bureau, commerce) et la surface.

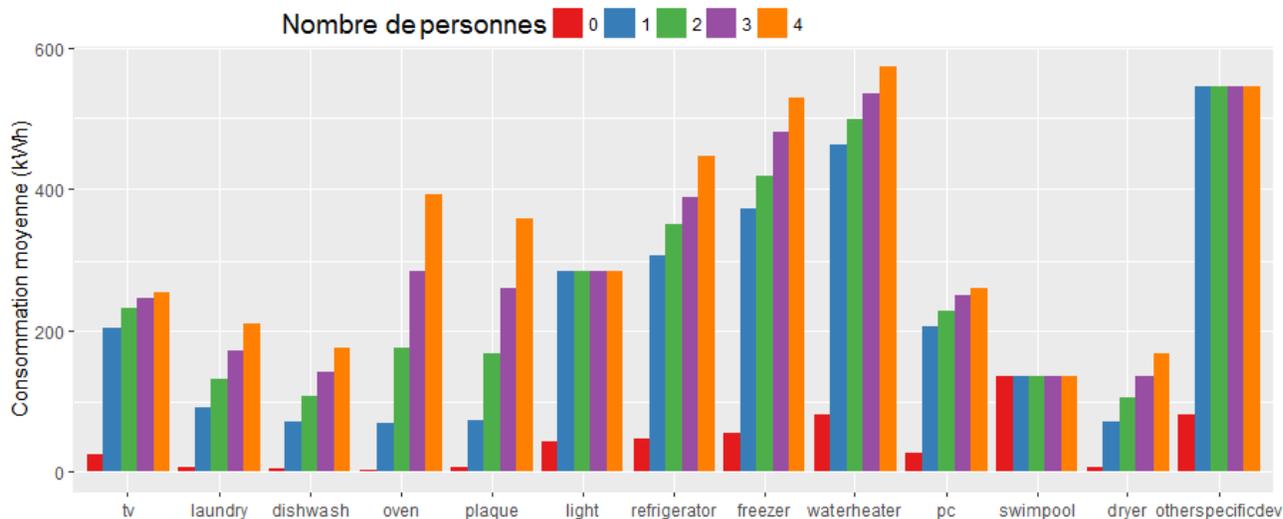


FIGURE 4.19 – Influence de la consommation de chaque usage en fonction du nombre de personnes

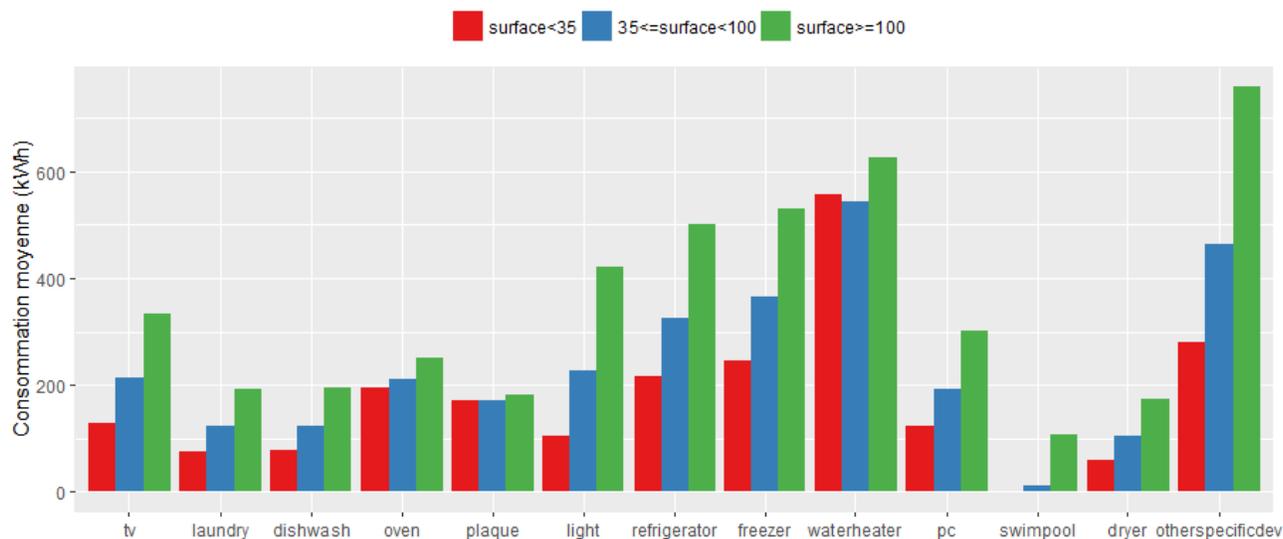


FIGURE 4.20 – Influence du nombre de la consommation de chaque usage en fonction de la surface

### 4.3 Structure et implémentation du simulateur

Les modèles précédemment présentés ont été implémentés en langage C++ pour la partie “cœur de calcul”, afin d’avoir des modèles les plus rapides possible. Ce cœur de calcul a été interfacé avec le langage R, permettant une manipulation aisée des grandes bases de données et leur traitement statistique. Ce langage permet aussi de créer aisément des packages documentés et d’assurer le suivi de version, très pratique pour les livraisons de l’outil de simulation pour Enedis.

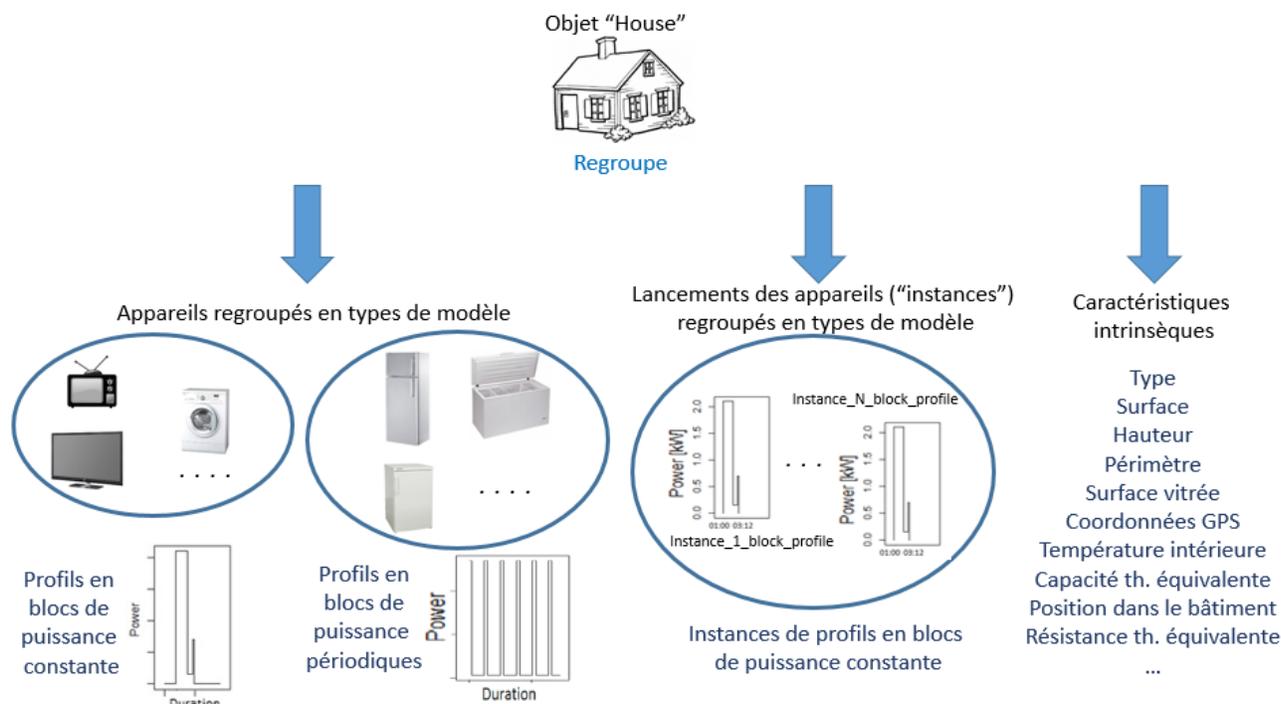


FIGURE 4.21 – Illustration des attributs de l’objet “House”

### 4.3.1 Partie “cœur de calcul”

Nous avons choisi la programmation C++ orientée objet pour implémenter les modèles. La brique de base étant l’objet “House” qui modélise un client, avec ses différentes caractéristiques (*e.g.* type de clients (*e.g.* résidentiel, bureaux, commerce), surface, hauteur, résistance thermique équivalente), ses appareils et leurs caractéristiques (*e.g.* type d’appareils, puissance, distribution des durées d’utilisation), et enfin les lancements de ces appareils (*i.e.* date de départ, profils de puissance).

La figure 4.21 schématise les attributs de l’objet “House”.

Le regroupement des différents objets est fait sous forme vectorisée. Les détails de l’implémentation sont fournis en annexe 8.2.2, en figure 8.3.

### 4.3.2 Étapes de la simulation

Pour mieux comprendre comment l’outil fonctionne globalement, nous présentons ici les différentes étapes de la simulation. Celles-ci sont illustrées en figure 4.22 et sont détaillées dans les paragraphes suivants.

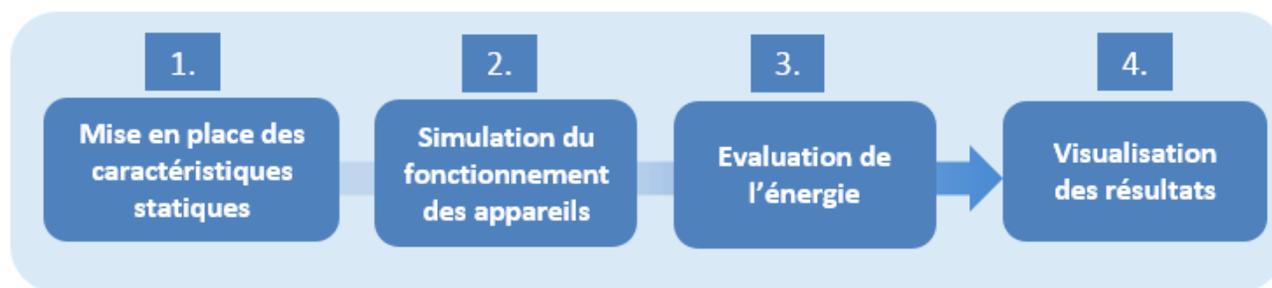


FIGURE 4.22 – Différentes étapes de la simulation

### Mise en place des caractéristiques statiques

Dans cette partie, les clients sont initialisés en allant chercher les différentes informations dans les bases de données (*e.g.* type, surface). Les usages sont installés chez les clients avec les données sur le parc d'appareils ( $Pa$ ), et les différentes puissances et répartitions des durées pour chaque appareil sont attribuées avec les données sur les technologies d'appareils ( $Ta$ ) et les comportements d'utilisation ( $Co$ ).

A la fin de cette étape, les objets “House” représentant les clients sont donc construits et les appareils sont installés chez les clients.

### Simulation du fonctionnement des appareils

Dans cette étape, on introduit les dates de début et de fin de la simulation, ainsi que les données météorologiques  $Me$  pour le lancement des appareils de chauffage et de climatisation. Les profils et les probabilités de lancement sont construits par client à partir des profils journaliers, hebdomadaires et saisonniers, provenant des bases de données sur les comportements d'utilisation. Le nombre de lancement est déterminé sur la période de simulation en lien avec le nombre de personnes pour les appareils modélisés par blocs de puissance constante, selon la météorologie pour le chauffage et la climatisation. Et puis les appareils sont lancés ; leur date de départ et la valeur de la puissance de chaque profil sont sauvegardées sous la forme d'objet “instance”, correspondant à un lancement d'appareils. Cette méthodologie permet de ne pas encore avoir de pas de temps à définir à cette étape, et permet donc de ne pas refaire la simulation quand on souhaite changer de pas de temps.

### Évaluation de l'énergie

C'est dans cette étape que le pas de temps est choisi. Une grille temporelle est construite et chaque lancement d'appareils est parcouru selon cette grille pour avoir finalement la consommation électrique à chaque pas de temps de la grille. La figure 4.23 illustre la manière avec laquelle est calculée l'énergie en parcourant chaque lancement.

### Visualisation des résultats

La simulation ainsi effectuée, il est alors possible de visualiser les résultats par appareil, par client, d'agréger les résultats par grande catégories d'usages ou de faire la somme totale. Un exemple de résultat de simulation est fourni en figure 4.24

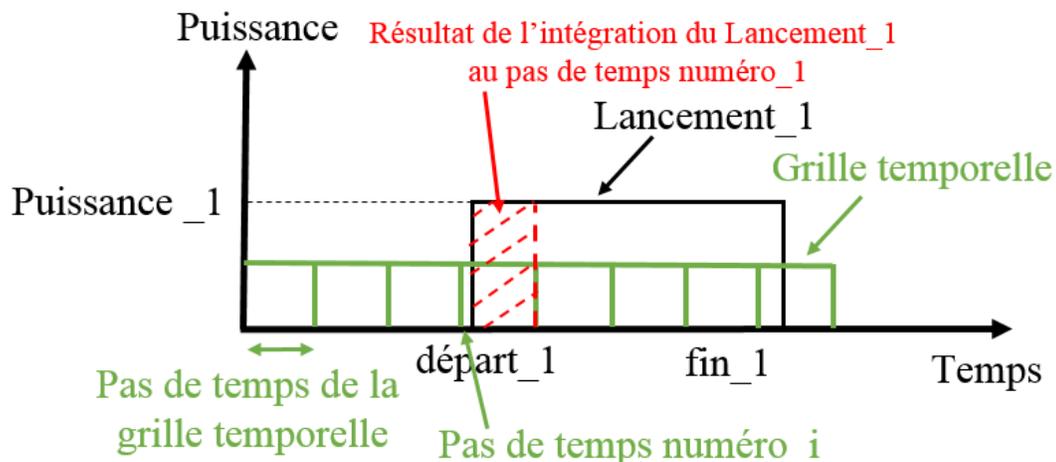


FIGURE 4.23 – Illustration de la manière avec laquelle est calculée l'énergie en sauvegardant chaque lancement d'appareils et en les parcourant avec la grille temporelle

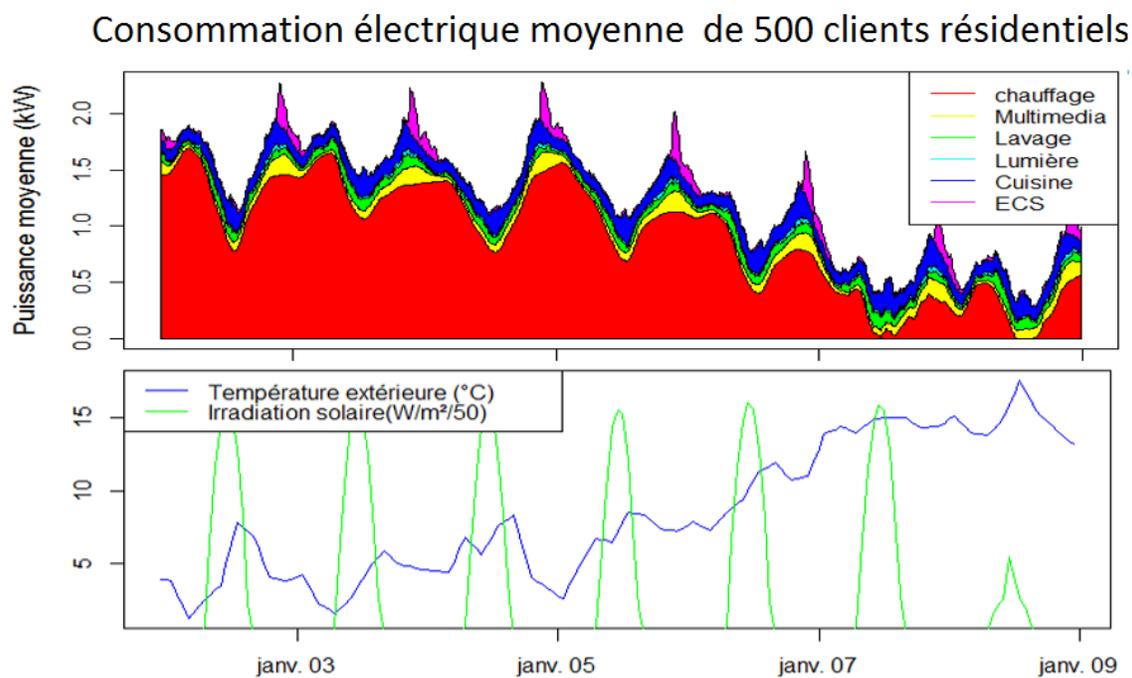


FIGURE 4.24 – Exemple de résultats de simulation de 500 clients résidentiels. Dans le cadre supérieur est représentée la consommation électrique ramenée par client, décomposée selon les grandes catégories d'usages. Dans le cadre inférieur figurent les conditions météorologiques

## 4.4 Conclusion et perspectives du chapitre

Dans ce chapitre, nous avons présenté de façon plus spécifique les besoins de modélisation bottom-up de la consommation électrique. Nous avons aussi montré comment le modèle développé s'inscrit dans la littérature, puis nous sommes rentrés dans le détail des modèles, résidentiels et tertiaires. Le lien entre les modèles et les données d'entrée disponibles a été décrit. Par la suite, nous avons détaillé comment ces modèles sont implémentés et quelle est la structure de l'outil de simulation. Le simulateur de courbes de charge, maintenant prêt à être utilisé, va être évalué et calibré sur les données de mesures dont nous disposons.

C'est un modèle en constante évolution, enrichi régulièrement par de nouvelles études, données, ou besoins de simulations. Les travaux actuels portent notamment sur l'amélioration de la modélisation des véhicules électriques dont le déploiement en cours est un enjeu majeur pour le réseau de distribution. De plus, une hypothèse importante du modèle actuellement implémenté est de considérer les lancements des différents appareils de façon indépendante, avec des profils de lancements probabilistes. Cette hypothèse n'influence pas la qualité des résultats de simulation des consommations agrégées à l'échelle de centaines de clients, mais en descendant plus précisément à la dizaine de clients, les résultats ne sont plus réalistes. En effet, ce type de modélisation ne prend pas en compte la corrélation qui peut être importante entre les différents appareils (*e.g.* corrélation de l'utilisation du lave-linge et du sèche-linge), de même que des absences individuelles liées aux périodes de vacances. Des modèles plus fins des bâtiments, en intégrant par exemple les parties communes, la position des logements, sont aussi des perspectives notables pour préciser les modèles et mieux exploiter les données disponibles. Une autre piste d'amélioration importante concerne le modèle de chauffage électrique. Son modèle thermique pourrait notamment intégrer une deuxième constante de temps, ce qui rendrait le modèle plus réaliste par exemple pour la simulation des effacements. Le modèle de chauffage et de climatisation gagnerait aussi en réalisme en intégrant les apports internes de chaleur, liés en particulier aux pertes électriques des appareils.

Enfin, l'amélioration des modèles du secteur tertiaire, en faisant dépendre la consommation du nombre de salariés (information notamment issue de la base SIRENE), et de l'occupation des locaux (*e.g.* modéliser le nombre de clients, de visiteurs) sont des perspectives notables d'amélioration.



## Chapitre 5

# Validation et calibration du modèle

### Résumé du chapitre

*Le simulateur de courbes de charge ainsi développé nécessite maintenant d'être validé et calibré avant d'être utilisé à des fins prospectives. L'état des lieux sur la littérature des modèles bottom-up a montré que les auteurs comparent principalement l'énergie et les profils simulés à des mesures de courbes de charge individuelles. Dans ce chapitre, nous avons dans un premier temps calibré l'outil de simulation afin que les résultats en énergie par usage pour le secteur résidentiel correspondent à des études nationales de référence.*

*Ensuite, les profils résultant de la simulation des usages non thermosensibles (i.e. qui ne dépendent pas de la température) ont été évalués avec les mesures des départs HTA d'Enedis. Afin de fixer une référence dans les résultats d'évaluation, un modèle statistique utilisant les mêmes données a été développé. Les résultats sur la partie non thermosensible montrent que le simulateur de courbes de charge suit les mêmes tendances sur les données que le modèle de référence, avec des niveaux d'erreurs acceptables. Des pistes de perfectionnement sont proposées pour diminuer ces erreurs en améliorant les modèles du simulateur.*

*Enfin, une méthode de calibration du modèle de chauffage électrique sur la partie thermosensible des mesures de courbes de charge de départs HTA est présentée. Elle permet de corriger par départ HTA des paramètres influant de façon importante sur la consommation thermosensible. La prévision et l'analyse de ces facteurs correctifs sont des perspectives notables de cette méthode, pour estimer ces paramètres sur des zones sans mesures, dans un contexte prospectif.*

## 5.1 Introduction du chapitre

La constitution de la base de données pour expliquer la consommation électrique a été décrite au chapitre 3. Le modèle de consommation électrique, quant à lui, a fait l'objet du chapitre 4. L'objet de ce chapitre est de valider et calibrer le modèle sur une partie des données disponibles. Après une description de l'état de l'art de la validation et de la calibration des modèles bottom-up, nous présentons dans ce chapitre différentes méthodes de validation et de calibration du modèle à partir des données disponibles, dans l'objectif de pouvoir l'utiliser à des fins prospectives.

## 5.2 État de l'art de l'évaluation et la calibration des modèles bottom-up

En utilisant des lois physiques combinées à des données détaillées sur les bâtiments, les occupants et les conditions extérieures pour simuler la consommation électrique, les modèles bottom-up utilisent majoritairement des connaissances et hypothèses *a priori* [148]. Ces connaissances *a priori* sont ensuite comparées à des mesures de consommation électrique réelles pour valider les modèles. Éventuellement, une calibration peut avoir lieu afin d'avoir des modèles dont les résultats soient plus proches de certaines caractéristiques des données mesurées.

Dans la plupart des cas étudiés dans la littérature, les auteurs utilisent des données individuelles d'au maximum quelques centaines de logements mesurés avec des compteurs intelligents, associées éventuellement à des questionnaires permettant de connaître les informations complémentaires sur ceux-ci (données socio-économiques, sur les locaux et sur les comportements).

On peut ainsi reprendre l'étude bibliographique vue au 4.1.2 et analyser les données utilisées pour l'évaluation de chacun des modèles et éventuellement leur calibration.

Nous avons présenté sous forme synthétique dans le tableau 5.1 les différents modèles bottom-up de la littérature à notre connaissance, les données utilisées pour la validation des modèles, ainsi que les méthodes de validation. Les modèles bottom-up que nous avons qualifiés de planification urbaine ([123], [120], [78]), ou de planification des infrastructures ([82], [97], [115]), ainsi que les modèles de Shimoda *et al.* [137] et Good *et al.* [84] ne sont pas représentés dans le tableau 5.1. Ils sont en effet validés de façon différente à des échelles plus macroscopiques, par comparaison à des données de consommations annuelles agrégées ou éventuellement à des consommations maximales mesurées sur de larges zones avec des profils type par catégorie de clients [119]. Quant au modèle GridLAB-D [100], bien que celui-ci soit utilisé pour modéliser la demande de départs HTA (*feeders* en anglais), les validations ne sont pas détaillées car ce sont des modules propres aux compagnies qui les utilisent [130]. Il semble cependant que les validations effectuées concernent plutôt les aspects réseau et calibration des modèles ZIP [54], [129] ou des aspects plus théoriques de la modélisation orientée agent [55] que les caractéristiques de la consommation électrique.

Auteur(s)	Référence	Données utilisées pour la validation (pas de temps)	Méthode utilisée
Dickert	[63]	Énergies et courbes de charge individuelles (cdci) typiques du secteur résidentiel venant du GRD allemand	Histogramme des consommations en énergie et comparaison qualitative de la courbe de charge journalière type
Yao et Steemers	[156]	Cdc résidentielle typique journalière du UK Electricity Associate Load Research Team (UK EALRT)	Corrélation avec la courbe de charge journalière type
Stokes	[141]	Cdc d'un quartier résidentiel (5min), Cdc de 13 logements (1min)	Comparaison énergie consommée sur la période de la simulation, des profils journaliers et de la puissance maximale
Paatero and Lund	[112]	Cdc de 702 logements pendant un an, Cdc de 1082 logements pendant 143 jours	Comparaison du profil-type journalier, de l'évolution annuelle de l'énergie journalière, de l'évolution de la variance avec le niveau d'agrégation, de l'histogramme de l'énergie journalière consommée
Walker	[152]	Cdc de 44 logements (10min)	Calcul du facteur normalisé de variation (NVF) et comparaison des consommations en énergie sur les courbes de charge journalières et moyennées sur 5 jours à différents niveaux d'agrégation
Capasso	[50]	Cdc de 95 logements dans la zone de Milan sur un mois (15min), avec questionnaire (informations socio-démographiques)	Calcul du NVF sur 22 jours et comparaison des courbes de charge journalières moyennées
Armstrong	[37]	Cdc des usages non thermosensibles venant de mesures d'Hydro Québec sur 2,5 années (15min)	Comparaison des consommations annuelles, maximales et moyennes sur 4 maisons sélectionnées. Comparaison visuelle des cdc, et comparaison des courbes de probabilité d'occurrence des valeurs de puissance
Widen <i>et al</i>	[154]	Cdc d'appareils individuels de 200 appartements et 200 maison en Suède (10min) avec questionnaire et activité des occupants (pour 5 d'entre eux)	Comparaison de la consommation annuelle, selon le type de logement, de la courbe de charge journalière moyenne selon type de jour. Calcul du NVF de quelques usages ; comparaison de la cdc journalière, de la puissance maximale

Auteur(s)	Référence	Données utilisées pour la validation (pas de temps)	Méthode utilisée
Widen and Wackekgard	[155]	Cdci d'appareils individuels de 14 logements en Suède (10min), avec questionnaire et activité des occupants (pour 5 d'entres eux)	Comparaison de la variance journalière des monotones de puissance. Comparaison de l'énergie annuelle, consommation moyenne, maximale, minimale, facteur de charge par usage. Comparaison des courbes de charge horaires moyennes, de l'énergie selon la taille du logement
Richardson <i>et al</i>	[122]	Cdci de 22 logements (1min)	Comparaison de l'énergie annuelle et évolution mensuelle de la moyenne journalière, courbe de charge journalière, volatilité de la demande à 1 min, facteur de diversité, demande maximale après désagrégation, monotones de puissance, facteur de puissance
Train	[148]	Cdci de 800 logements sur 2 ans avec questionnaire sur les différents usages et les données socio-économiques	Calcul de l'erreur quadratique moyenne pour la consommation agrégée de chaque usage
Berthou	[44]	Statistiques nationales sur les consommations annuelles par type d'appareils venant du CEREN	Comparaison des consommations annuelles moyennes par usage après simulation de 1000 logements

TABLE 5.1 – Différents modèles bottom-up de la littérature, données utilisées pour la calibration et méthode d'évaluation

Dans le cas de la thèse, nous ne disposons pas encore de courbes de charges individuelles sur de larges zones pour la France, le déploiement des compteurs linky étant en cours.

Dans ce chapitre, nous avons d'abord effectué la validation du simulateur de courbes de charge sur les répartitions de consommation en énergie à l'échelle nationale sur les différents secteurs et usages. Ensuite, nous avons validé les profils journaliers et mensuels avec la grande quantité de données des départs HTA. Finalement des méthodes d'évaluation de la partie thermosensible sont décrites et mises en pratique dans le chapitre suivant.

### 5.3 Validation et calibration des consommations en énergie du secteur résidentiel

#### Introduction

On a vu au chapitre 4 section 4.2.1 que les valeurs choisies pour régler les consommations annuelles, les dépendances au nombre de personnes et à la surface du logement sont issues de différentes campagnes de mesures qui ne sont pas forcément de la même année (voir 3.2.7).

Une étude de juin 2016 de l'ADEME [32], effectuée par Enertech précise que : “aujourd’hui, à notre connaissance, personne ne peut dire comment se répartit la consommation électrodomestique d’un foyer. Nous ne savons par exemple pas chiffrer l’impact de l’introduction incessante de nouveaux équipements liés au confort moderne. En effet, peu de campagnes de mesures ont été réalisées et dans tous les cas, leurs résultats sont très rapidement dépassés. Cette méconnaissance des enjeux limite la pertinence de l’action engagée pour réduire nos consommations d’électricité. Il y a un réel besoin de données mises à jour en permanence.” Pour pallier ces difficultés, et afin d’utiliser un référentiel accessible, une calibration du simulateur de courbes de charge sur des valeurs nationales prises à année donnée nous a paru indispensable.

### Données utilisées pour la calibration

De même que Berthou *et al.* [44], nous avons calibré les consommations annuelles de chaque usage résidentiel sur les données provenant du CEREN. Nous nous sommes basés sur leur publication la plus récente en date de la thèse, soit celle de 2015 [22], pour calibrer les consommations de chaque usage. Les valeurs retenues ont été mises dans le tableau 5.2. La colonne la plus à droite du tableau reprend la répartition des consommations résidentielles venant d’une autre source (EDF) pour vérifier la cohérence des calculs effectués. Au vu des proches valeurs par rapport à celle que nous avons calculées, il semble qu’EDF ait aussi utilisé les données CEREN pour la publication de ces valeurs [7]. Le détail du calcul est disponible en annexe 8.3.1.

Type de logement	Type d’usage	Consommation par logement équipé (kwh/an)	Répartition consommation logement équipé (%)	Répartition consommation selon EDF [7] (%)
Maison individuelle	Chauffage électrique	9304	60,2	62
Maison individuelle	ECS électrique	1755	11,4	11
Maison individuelle	Cuisson électrique	1139	7,4	7
Maison individuelle	Électricité spécifique	3246	21,0	20
Maison individuelle	Total	15443	100	100
Immeuble collectif	Chauffage électrique	7121	61,6	62
Immeuble collectif	ECS électrique	1603	13,9	14
Immeuble collectif	Cuisson électrique	781	6,8	7
Immeuble collectif	Électricité spécifique	2047	17,7	17
Immeuble collectif	Total	11552	100	100

TABLE 5.2 – Valeurs retenues pour la calibration de chaque usage du résidentiel

Afin de décomposer les usages de “l’électricité spécifique” nous avons utilisé la décomposition provenant d’EDF [7] de la figure 5.1.

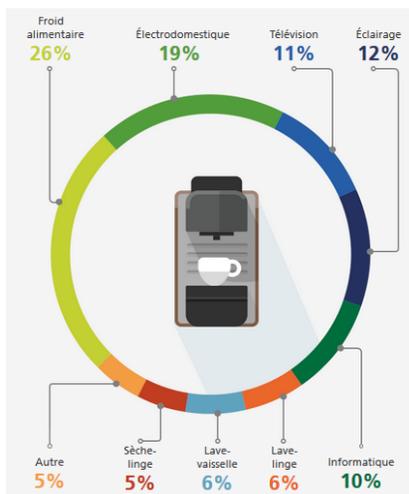


FIGURE 5.1 – Répartition de la consommation d’électricité d’un ménage français (hors chauffage, eau chaude et cuisson) [7]

### Méthode de calibration par usage

Nous avons vu au paragraphe précédent qu’à chaque usage nous avons fait correspondre une consommation cible moyenne par type de logement équipé. Nous avons ensuite modifié les valeurs par défaut de chaque usage du simulateur correspondant de façon à obtenir les valeurs cibles sur un grand nombre de simulations. De façon empirique, nous avons simulé 1000 logements de manière à atteindre une valeur de consommation stable, ce chiffre étant aussi confirmé par la littérature des simulateurs de courbes de charge [87]. Concernant les hypothèses prises pour la simulation, nous avons pris comme données pour la surface et le nombre de personnes une distribution gaussienne centrée selon les moyennes nationales par type de logement (voir tableau 5.3). Pour chaque gaussienne, l’écart-type ainsi que le minimum et le maximum de la distribution ont été pris d’après les données de l’enquête logement de l’INSEE de 2013.

Type de logement	Surface moyenne (m <sup>2</sup> )	Nombre de personnes par ménage moyen
Maison individuelle	113	2,45
Immeuble collectif	63	1,94

TABLE 5.3 – Surface et nombre moyen de personnes par type de logement [22]

Le tableau 5.4 recense les valeurs utilisées comme référence ainsi que la tolérance obtenue en re-simulant 10 fois chaque simulation de 1000 logements équipés. A noter que pour ce qui concerne les usages spécifiques, les valeurs correspondent non pas aux logements équipés mais aux moyennes nationales, donc les simulations pour ces usages tiennent compte du taux de possession. Les données du CEREN concernant le chauffage électrique, usage dépendant de la météorologie, sont publiées à climat normal. Ainsi pour la calibration de cet usage nous avons procédé de même, en utilisant une année météorologique normale, obtenue par Enedis.

Usage	Usage du simulateur	Consommation cible par MI(kWh/an)	Consommation simulée par MI (kWh/an)	Consommation cible par IC (kWh/an)	Consommation simulée par IC (kWh/an)	Méthode de calibration utilisée
Chauffage électrique	Chauffage	9 303,6	9305 ± 12	7 120,5	7121 ± 15	Coefficient correctif sur la matrice U
ECS électrique	ECS	1 755,0	1774 ± 5	1 603,4	1603 ± 5	Nombre de lancements par personne
Cuisson électrique	Four et plaque de cuisson	1 138,6	1139 ± 3	781,1	781 ± 2	Nombre de lancements par personne
Froid alimentaire	Réfrigérateur et congélateur	843,9	844 ± 3	532,3	532 ± 3	Puissance installée par personne
Télévision	Télévision	357,0	358 ± 1	225,2	225 ± 2	Nombre de lancements par personne
Éclairage	Éclairage	389,5	389 ± 3	245,7	246 ± 2	Puissance installée par unité de surface
Informatique	Ordinateur	324,6	324 ± 1	204,7	205 ± 1	Nombre de lancements par personne
Lave-linge	Lave-linge	194,8	195 ± 2	122,8	123 ± 1	Nombre de lancements par personne
Lave-vaisselle	Lave-vaisselle	194,8	195 ± 2	122,8	123 ± 1	Nombre de lancements par personne
Sèche-linge	Sèche-linge	162,3	162 ± 1	102,4	102 ± 1	Nombre de lancements par personne
Électro-domestique	Électro-domestique	616,7	617 ± 5	389,0	389 ± 3	Nombre de lancements par personne
Autre	Autre	162,3	162 ± 1	102,4	102 ± 1	Puissance installée par unité de surface
Total	-	15 443,2	15 445 ± 15	11 552,4	11552 ± 16	-

TABLE 5.4 – Valeurs, méthode retenue pour la calibration des usages et résultats obtenus à l'issue de la calibration

## 5.4 Validation et calibration sur les départs HTA d'Enedis

### 5.4.1 Introduction

Enedis dispose des mesures de ses départs HTA sur toute la France (voir section 3.2.1). Afin d'évaluer et de calibrer le simulateur de courbes de charge sur ces mesures, nous proposons ici une démarche qui sera mise en pratique au chapitre 6. La démarche s'appuie sur la décomposition des courbes de charge des départs HTA en une partie thermosensible, et une partie non thermosensible. Cette procédure de décomposition est couramment utilisée dans la littérature pour les courbes de charge agrégées (*e.g.* [102], [81]). Elle est décrite en détail en section 3.3.5, notamment avec l'équation 3.16.

Nous supposons dans la modélisation que les seuls usages thermosensibles sont le chauffage électrique et la climatisation. Tous les autres usages simulés contribuent à la partie non thermosensible de la courbe de charge. Nous utilisons la partie non thermosensible pour évaluer dans un premier temps la qualité des profils de la simulation. Pour avoir une référence dans cette évaluation des profils, une étude a été menée sur une méthode permettant de retrouver de façon optimale les profils à partir des données Enedis. Celle-ci a fait l'objet d'une publication dans un journal scientifique [81] et a été insérée en annexe 8.3.2. Cette référence a permis ensuite de comparer les erreurs obtenues avec utilisation du simulateur de courbes de charge. La partie non thermosensible permet aussi de s'assurer que la valeur de l'énergie totale simulée des usages non thermosensibles du simulateur est correcte par rapport aux mesures.

Concernant la partie thermosensible, celle-ci est un point clé notamment pour Enedis qui dimensionne les départs HTA selon la température la plus froide (voir section 6.1.2). Une méthode de calibration des paramètres influant le plus la thermosensibilité simulée est proposée puis mise en pratique dans le chapitre 6.

### 5.4.2 Validation et calibration des parties non thermosensibles

#### Mise en place d'une référence de comparaison

L'évaluation des profils sur des données réelles est une étape nécessaire à l'évaluation du modèle. L'état de l'art de l'évaluation et la calibration des modèles bottom-up a montré qu'aucun modèle n'est évalué sur des mesures agrégées de départs HTA. Cette échelle de mesure, bien que beaucoup plus agrégée que les mesures de consommations individuelles, présente d'autres challenges de modélisation dont nous avons abondamment parlé : disponibilité des données, associations, problèmes de maintenance et réorganisation des départs HTA etc.

Afin d'avoir des premiers résultats de simulation qui servent de référence dans l'évaluation des profils du simulateur, nous avons développé une méthode permettant de simuler la consommation de chaque départ HTA à partir de profils optimaux en terme d'erreur de simulation. Ceux-ci ont été obtenus à l'aide des données de mesure de consommation des clients d'Enedis. La méthode et les résultats ont été publiés dans un journal scientifique (voir [81]).

Cette méthode (décrite en détail en annexe 8.3.2) est basée uniquement sur les données des clients et des mesures de départs HTA d'Enedis présentées en 3.2.1. Elle se base sur la partie non thermosensible de la courbe de charge, obtenue par la même méthode que celle présentée en section

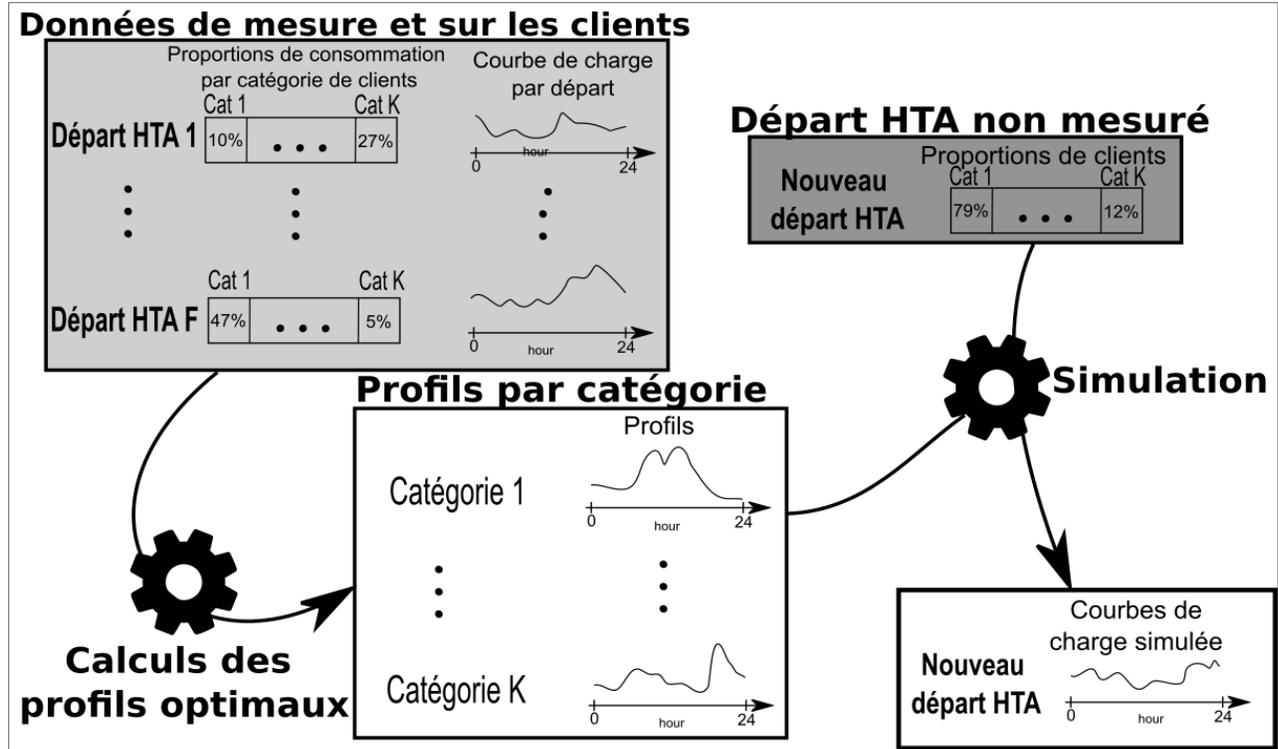


FIGURE 5.2 – Diagramme illustrant la méthode de référence pour la simulation de la partie non-thermosensible des courbes de charge de départ HTA d'Enedis [81]

3.3.5, et normalisée par sa consommation moyenne annuelle.

La figure 5.2 illustre la méthode de référence pour la simulation de la partie non-thermosensible des courbes de charge des départs HTA d'Enedis. Les données d'entrée sont les mesures et les informations sur les clients, provenant des bases de données d'Enedis. Pour chaque départ, on suppose les proportions de consommations de chaque type de clients connues (*e.g.* clients bleu domestiques, hôtels-restaurants). C'est bien le cas pour Enedis qui dispose des relevés de compteurs et des informations contractuelles par client. Cette méthode suppose que la partie non thermosensible normalisée de chaque départ HTA  $Pnts_n^z$  est la somme des profils par catégorie de clients  $P_k$  pondérés par la proportion de la consommation de chaque type de clients  $p_k^z$  :

$$Pnts_n^z = \frac{Pnts^z}{\text{mean}(Pnts^z)} = \sum_{k=1}^K p_k^z \times P_k \quad (5.1)$$

Avec  $\text{mean}(Pnts^z)$  la valeur moyenne de  $Pnts$ . Les  $Pnts_n^z$  et les  $p_k^z$  étant connus pour chaque départ HTA, un algorithme a été développé pour retrouver les profils par catégorie  $P_k$  en minimisant les erreurs de simulation. Une fois les profils de consommation obtenus par catégorie de clients, il est possible de simuler un départ non mesuré en connaissant uniquement la proportion de consommation annuelle de chaque type de clients.

Différentes catégorisations de clients ont été testées à partir de la classification de la base de données des clients d'Enedis. La quantification de la qualité des profils obtenus a été faite en simulant des nouveaux départs HTA non inclus dans le jeu de données initial d'apprentissage. Les résultats en terme d'erreur quadratique moyenne normalisée sont présentés dans le tableau 5.5.

Région	Année	2 catégories	8 catégories	9 catégories	12 catégories
BLOIS	2010	24.36 (2.32)	<b>23.90 (2.75)</b>	24.04 (2.99)	26.78 (2.91)
	2011	23.87 (1.62)	<b>22.79 (1.42)</b>	22.91 (1.16)	24.78 (2.09)
	2012	22.84 (1.26)	22.54 (1.17)	<b>22.09 (1.24)</b>	24.09 (2.17)
	2013	22.34 (2.06)	22.31 (1.98)	<b>21.32 (1.96)</b>	23.34 (2.04)
	Moyenne	23.35 (1.86)	22.89 (1.93)	<b>22.59 (1.98)</b>	24.75 (2.33)
LYON	2010	19.05 (2.39)	19.42 (2.71)	<b>18.29 (2.24)</b>	19.23 (1.94)
	2011	19.28 (1.24)	<b>18.06 (1.55)</b>	18.56 (1.20)	18.46 (1.42)
	2012	19.07 (1.35)	<b>18.21 (1.72)</b>	18.30 (1.34)	19.00 (1.86)
	2013	18.06 (1.03)	17.92 (2.03)	<b>17.49 (1.12)</b>	18.68 (1.91)
	Moyenne	18.87 (1.59)	18.40 (2.05)	<b>18.16 (1.58)</b>	18.84 (1.79)
RENNES	2010	22.57 (0.96)	21.67 (1.23)	<b>21.59 (1.04)</b>	22.70 (1.57)
	2011	22.62 (1.22)	<b>21.54 (1.48)</b>	21.57 (1.06)	22.10 (1.08)
	2012	22.75 (1.11)	22.96 (0.99)	<b>22.39 (0.98)</b>	22.61 (0.84)
	2013	24.94 (1.03)	<b>23.99 (1.37)</b>	24.14 (1.26)	24.08 (1.37)
	Moyenne	23.22 (1.08)	22.54 (1.28)	<b>22.42 (1.09)</b>	22.87 (1.25)

TABLE 5.5 – Erreurs quadratiques moyennes normalisées (NRMSE) (en %) du modèle de référence [81] pour les 3 différentes zones de simulation sur les 4 années de mesure en fonction du nombre de catégories. Les catégories sont issues de la base d’Enedis. La simulation a été effectuée 100 fois, et les NRMSE ont été reportées ainsi que la variance entre parenthèses

Ces taux d’erreurs ont été obtenus avec des profils calculés sur les départs HTA des trois zones (Blois, Lyon et Rennes). Il est important de noter que ces profils sont annuels et ont été obtenus à partir des mesures de façon à minimiser les erreurs quadratiques moyennes lors de la simulation d’un nouveau départ HTA. Dans cette section nous cherchons à évaluer les profils rentrés dans le simulateur de courbes de charge. Ces profils n’ont pas été trouvés en fonction des mesures, mais ont été renseignés *a priori* suite à des résultats de campagnes de mesures. Ainsi ces valeurs de référence constituent plutôt une borne basse pour la simulation de type de bottom-up avec ce type de données.

### Simulation de la partie non thermosensible du simulateur et comparaison à la référence

De façon à évaluer les profils du simulateur dans les mêmes conditions que la méthode de référence, nous avons simulé avec le simulateur de courbes de charge la partie non thermosensible de chaque départ HTA. Nous avons ensuite normalisé cette consommation par rapport à sa moyenne annuelle, ce pour chaque départ. Concernant les données d’entrée, celles-ci proviennent pour la partie résidentielle des résultats de l’association par méthode d’optimisation de la base des clients résidentiels d’Enedis et des logements de l’INSEE, décrite en section 3.3.4. Pour les données d’entrée du secteur tertiaire, nous avons supposé la consommation annuelle de chaque client connue, comme c’est le cas pour la méthode de référence. Nous avons ensuite calculé les surfaces correspondantes pour obtenir la consommation estimée de chaque client, nous permettant ensuite de générer les courbes de charge avec le simulateur.

Avec ces entrées du modèle, la partie non thermosensible de chaque départ HTA a été simulée avec le simulateur de courbes de charge pour l’année 2012.

Les résultats sont présentés dans le tableau 5.6. Nous avons donné les erreurs quadratiques moyennes normalisées (NRMSE) entre la partie thermosensible simulée et celle mesurée pour les

TABLE 5.6 – Erreur quadratique moyenne normalisée entre la partie thermosensible simulée et celle mesurée. Les différents quantiles sont en colonne, les zones de simulation en ligne. Pour la moyenne et l'écart type, les valeurs entre parenthèses correspondent aux valeurs de référence du tableau 5.5 pour l'année 2012 avec 12 catégories

	0%	25%	50%	Moyenne	Ecart type	75%	100%
Blois	8,5%	21,2%	25,5%	26,2% (24.1%)	7,4% (2.17%)	31,0%	48,2%
Lyon	8,3%	16,8%	20,5%	21,8% (19,0%)	7,2% (1.86%)	25,0%	47,6%
Rennes	13,1%	21,5%	25,4%	25,8% (22.61%)	5,7% (0.84%)	29,2%	41,3%

mêmes départs HTA des 3 zones correspondantes à celle de l'étude de référence [81]. La NRMSE a été définie comme étant :  $NRMSE = \sqrt{\text{mean}((\frac{P_{meas}}{\text{mean}(P_{meas})} - \frac{P_{simu}}{\text{mean}(P_{simu})})^2)}$ . Avec  $P_{meas}$  puissance mesurée du départ HTA et  $P_{simu}$  la puissance simulée par le simulateur de courbes de charge. On constate des NRMSE d'environ 3 points supérieurs à celle de la référence, ce qui s'explique par le fait que les profils du simulateur de courbes de charge ne sont pas trouvés de façon optimale, mais *a priori*. Ils résultent de profils de consommation par usages de campagnes de mesures. On constate cependant, comme dans l'étude de référence, que la consommation non thermosensible de la zone de Lyon présente des erreurs de simulation plus faibles de quelques pour cent par rapport aux deux autres zones.

Concernant l'écart type des erreurs, celui-ci est bien plus important avec le simulateur qu'avec l'étude de référence. Les profils résultant des campagnes de mesures semblent ne pas toujours être bien ajustés aux départs HTA à simuler et certains départs ont des NRMSE importantes, supérieures à 30%. La figure 5.3 représente l'évolution des NRMSE en fonction de la consommation moyenne des départs. On constate qu'une partie des erreurs ( $R^2=0.165$ ) est expliquée par la consommation moyenne non thermosensible du départ. Les résultats suggèrent que plus le départ HTA est important, plus la consommation électrique est prévisible et précisément simulable.

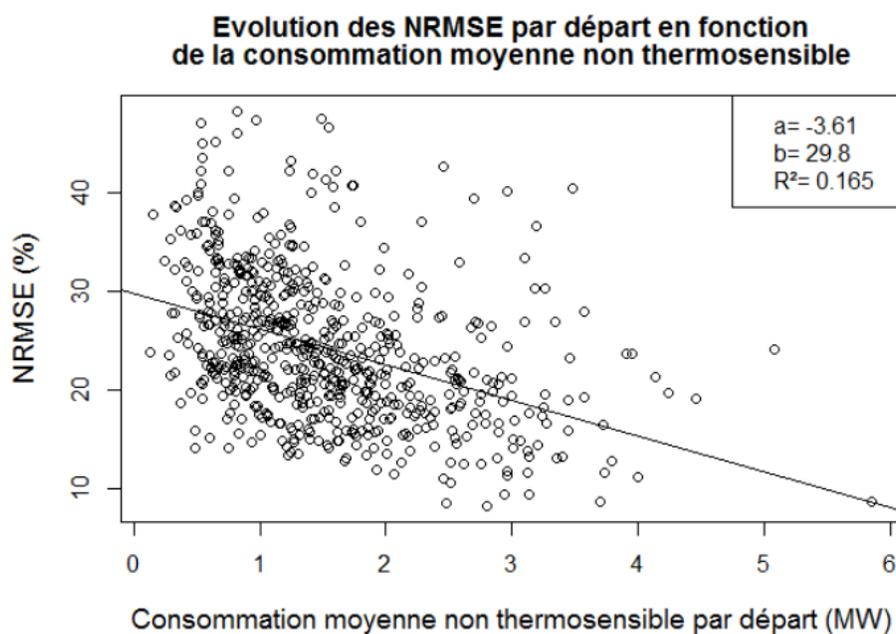


FIGURE 5.3 – Evolution des NRMSE par départ avec la consommation moyenne non thermosensible

### Comparaison sur la consommation en énergie de la partie non thermosensible

Les clients tertiaires ont été simulés de manière à respecter la consommation mesurée, mais pas les clients résidentiels. La comparaison entre la consommation annuelle non thermosensible mesurée et celle simulée est présentée dans le tableau 5.7. L'erreur relative est calculée comme étant la différence entre la consommation non thermosensible issue de mesures et celle simulée avec le simulateur de courbes de charge. On remarque une tendance assez marquée pour les 3 zones à surestimer la consommation annuelle, entre 13 et 20% en moyenne. Cela est assez surprenant étant donné la calibration en énergie effectuée des usages résidentiels à la section 5.3. De plus, les pertes dans le réseau n'ont pas été modélisées. On aurait donc tendance à sous estimer la simulation par rapport à la consommation mesurée au niveau des départs HTA de quelques pour cent (on a 6% en moyenne sur le réseau d'Enedis). Cette surestimation de l'énergie non thermosensible simulée vient probablement du fait qu'une partie de la consommation des usages hors chauffage est aussi thermosensible. Par exemple, l'arrivée d'eau des chauffe-eau est plus froide l'hiver que l'été. L'éclairage est aussi plus utilisé l'hiver que l'été, car les journées sont plus longues. Il y a probablement aussi un lien entre température extérieure et présence des personnes dans les bâtiments, menant à des modifications dans les consommations hors chauffage avec la température. Une dernière explication pourrait provenir du fait que la dépendance au nombre de personnes et à la surface de la consommation électrique a été modélisée comme linéaire pour de nombreux usages (voir figure 4.19, 4.20). Le domaine de validité de cette loi linéaire, bien qu'issue de campagnes de mesures, est probablement limité à 3-4 personnes par logements, maximum souvent observé dans les études. Dans les départs HTA avec une proportions importantes de foyers avec un nombre de personnes élevé, les consommations peuvent être aussi surestimées.

Des validations plus précises des consommations annuelles simulées par rapport à celles mesurées par IRIS sont prévues pour valider globalement la simulation des consommations en énergie résidentielle. Une simulation de tous les usages du simulateur et une calibration prenant en compte tous les paramètres, à la fois liés à la partie non thermosensible et thermosensible, permettrait de calibrer au mieux le simulateur de courbes de charge.

TABLE 5.7 – Distribution des erreurs relatives dans la simulation de l'énergie annuelle non thermosensible (%)

	Erreur relative dans la simulation de l'énergie annuelle non thermosensible (%)						
	0%	25%	50%	Moyenne	Ecart type	75%	100%
<b>Blois (abs)</b>	0,1	9,3	20,1	24,0	18,1	36,0	84,2
<b>Lyon (abs)</b>	0,0	9,8	19,1	22,1	16,2	30,2	98,9
<b>Rennes (abs)</b>	0,3	13,0	24,3	24,8	16,3	33,0	75,6
<b>Blois</b>	-84,2	1,3	15,6	15,2	26,0	32,2	71,8
<b>Lyon</b>	-98,9	-1,6	15,1	12,9	24,3	28,2	70,3
<b>Rennes</b>	-41,1	6,9	21,5	19,5	22,4	30,5	75,6

### 5.4.3 Calibration des parties thermosensibles

#### Introduction

Enedis dimensionne son réseau en prenant en compte à la fois la température la plus froide et la dépendance entre la température et la consommation électrique (voir section 6.1.2). En France, la plupart des départs HTA présentent uniquement une thermosensibilité négative, liée au chauffage électrique. Le cas de la thermosensibilité positive, lié à la présence de la climatisation, n'est pas traité dans cette thèse, bien qu'il soit d'intérêt croissant, notamment pour les régions du sud de la France, comme vu en section 6.2.1. Nous considérons ici uniquement le cas où la thermosensibilité est négative, et supposons que le chauffage électrique est le seul usage qui contribue à cette thermosensibilité, bien que nous ayons vu au paragraphe précédent les limites de cette hypothèse : des usages autres que le chauffage électrique participent à la thermosensibilité. Nous avons vu à la section 4.2.4 que de nombreux paramètres impactaient grandement la consommation de chauffage électrique : la surface, la position dans le cas d'un appartement, le niveau de performance, la consigne de chauffage, le COP du chauffage ainsi que la température extérieure. Le taux de possession du chauffage est aussi un élément clé.

#### Décomposition de la thermosensibilité par catégorie de clients

Nous avons vu au 3.3.5 que la thermosensibilité par départ était explicable en partie importante par la puissance souscrite totale par départ (voir par exemple figure 3.7,  $R^2$  de 0.814).

En prenant comme variable pour prédire la thermosensibilité non pas la somme des puissances souscrites par départ mais la puissance souscrite par catégorie de clients (*e.g.* bleu domestique, bleu agricole, bleu hôtel restaurant), on obtient :

$$Th^D \sim_l (P_{sous}^D) | K \quad (5.2)$$

Avec  $Th^D$  le vecteur contenant la thermosensibilité de chaque départ,  $P_{sous}^D$  le vecteur contenant la puissance souscrite de chaque départ et  $K$  l'ensemble de toutes les catégories de clients.

Cette régression, permet d'obtenir une thermosensibilité par unité de puissance souscrite pour chaque catégorie de clients. Nous avons présenté les résultats de cette régression pour l'année 2012 dans le tableau 5.8. On constate que les thermosensibilités de chaque branche sont négatives, sauf pour les hôtels et restaurants. Cela est certainement dû à la saisonnalité, faisant qu'il y a plus de personnes utilisant les hôtels et restaurants l'été que l'hiver, et cela n'est pas modélisé dans le simulateur actuel. Ainsi, en l'état, nous simulons les hôtels et restaurants avec une thermosensibilité positive pour prendre en compte cet effet. Cette hypothèse n'impacte pas grandement la simulation : la branche "café, hôtel, restaurant" représente seulement 7% des surfaces chauffées dans le secteur tertiaire [25], et 1.5% de la consommation totale en moyenne par départ HTA d'Enedis (chiffre issu de la base des clients d'Enedis). On remarque aussi que la catégorie "industrie" a une thermosensibilité par puissance souscrite particulièrement élevée par rapport aux autres, qui peut s'expliquer par le fait que cette catégorie peut dimensionner sa puissance souscrite au plus juste de la consommation, ce qui n'est pas forcément le cas des autres catégories. La thermosensibilité par unité de puissance souscrite des clients résidentiels est quant à elle la plus basse, probablement car à la puissance souscrite résidentielle n'est à l'inverse pas choisie au plus juste de la consommation des clients.

TABLE 5.8 – Résultats de la régression de la puissance souscrite par catégorie de clients en fonction de la thermosensibilité par départ, sur les trois zones de disponibilité des données, pour l'année 2012 (le  $R^2$  de la régression est de 0.899)

catégorie de clients $k$	Thermosensibilité par unité de puissance souscrite $Th^{Psous}$ (W/°C/kVA)
HTA	-7,3
Agricole	-5,2
Commerce	-4,9
Bureau	-11,2
Hôtel-Restaurant	7,6
Industrie	-19,5
Domestique	-2,9

Grâce au tableau 5.8 on a une thermosensibilité par unité de puissance souscrite pour chaque catégorie de clients  $k$ , notée  $Th_k^{Psous}$

Pour connaître la proportion de la contribution à la thermosensibilité de chaque catégorie de clients par départ HTA  $Prop_k^{Th,z}$ , on a :

$$Prop_k^{Th,z} = \frac{Th_k^{Psous} \times \sum_{j \in k} Psous^j}{\sum_{k \in K} Th_k^{Psous} \times \sum_{j \in k} Psous^j} \quad (5.3)$$

Avec  $K$  l'ensemble des catégories de clients,  $Psous^j$  la puissance souscrite du client  $j$ .

Cette proportion obtenue, il est alors possible de déterminer la thermosensibilité de chaque catégorie de clients  $Th_k^z$ , connaissant sa thermosensibilité du départ HTA mesuré  $Th_{meas}^z$  (thermosensibilité calculée comme vu à la section 3.3.5).

$$Th_k^z = Prop_k^{Th,z} \times Th_{meas}^z \quad (5.4)$$

### Secteur tertiaire et industriel

**Données à renseigner.** Pour le secteur tertiaire et industriel, les bases de données disponibles ne permettent de renseigner que la surface (*e.g.* repertoire SIRENE, base des impôts fonciers). Les autres informations qui influencent le chauffage électrique, à notre connaissance, ne sont présentes dans les études qu'à l'échelle nationale.

**Méthode de calibration.** Pour la simulation du chauffage du secteur tertiaire et industriel, nous retiendrons un coefficient de performance d'isolation moyen égal à 2, une température de consigne fixe à 19°C. Nous supposons que tous les clients tertiaire possèdent un chauffage électrique et ajustons le coefficient de performance du chauffage par catégorie de clients ( $COP_k$ ) de façon à retrouver les thermosensibilités de chaque catégorie de clients  $Th_k^z$ .

### Secteur résidentiel

**Données à renseigner.** Pour le secteur résidentiel, la plupart des informations pour le chauffage sont renseignées par les bases de données : la surface, le taux de possession, et le type de logement (maison ou appartement) sont renseignés avec la base logement de l'INSEE ; la position pour les appartements est déterminée selon un tirage dans une répartition d'un bâtiment moyen ;

la température extérieure est renseignée avec les données météorologiques. Concernant le COP, le taux de pénétration de pompe à chaleur étant très faible pour les années mesurée (2010-2013) [127], celui-ci est pris égal à 1. Concernant la température de consigne moyenne, celle-ci est habituellement prise dans les logements aux alentours de  $20^{\circ}\text{C}$ , bien qu'elle puisse évoluer grandement d'un logement et d'une pièce à l'autre [105]. Le niveau de performance d'isolation est déterminé en relation avec l'année de la dernière rénovation majeure (voir section 4.2.3, en particulier la figure 4.16). Cependant, la matrice permettant de calculer la résistance équivalente par logement à partir du niveau de performance correspond à des valeurs moyennes issues de statistiques nationales. Ainsi, pour chaque départ HTA, deux variables sont renseignées par défaut avec des incertitudes importantes : la température de consigne moyenne et les valeurs de conductivité thermique.

**Méthode de calibration.** La méthode de calibration du secteur résidentiel retenue s'appuie sur le modèle implémenté de chauffage électrique dans le simulateur de courbes de charge, décrit au 4.2.3. En reprenant l'équation de thermique des bâtiments 4.3, et en l'intégrant sur un an, avec un  $COP$  pris égal à 1, l'énergie annuelle consommée par départ HTA pour le chauffage résidentiel  $Pts_{an,res}^z$  dans le simulateur est :

$$Pts_{an,res}^z = \sum_{j \in z} \sum_{t \in an, t | T_{out}(t) < T_{seuil}} \frac{T_{cons}^j(t) - T_{out}^z(t)}{R^j} - S_{ves}^j \times E^z(t) \quad (5.5)$$

Avec  $T_{cons}^j$  la température de consigne du client  $j$ ,  $T_{out}^z$  et  $E(t)^z$  respectivement la température moyenne extérieure et l'irradiation solaire sur plan vertical orienté sud de la zone  $z$  desservie par le départ HTA.  $R^j$  est la résistance thermique équivalente du client,  $S_{ves}^j$  sa surface vitrée équivalente sud.  $T_{seuil}$  est la température seuil de non chauffage, prise par défaut à  $15^{\circ}\text{C}$ , comme Enedis dans ses études [71].

A partir de l'équation 5.5, on voit que la thermosensibilité résidentielle simulée du départ HTA  $Th_{simu,res}^z$  correspond au coefficient multiplicatif de  $T_{out}$ . Il vient donc :

$$Th_{simu,res}^z = \sum_{j \in z} \frac{1}{R^j} \quad (5.6)$$

La thermosensibilité totale des clients résidentiels du départ HTA à partir des mesures  $Th_{meas,res}^z$  est estimée avec l'équation 5.4.

Afin d'ajuster la thermosensibilité simulée pour obtenir la même que celle provenant des mesures, nous introduisons un facteur correctif par départ HTA pour toutes les résistances thermiques résidentielles  $Ucorr_{res}^z$  :

$$Ucorr_{res}^z = Th_{meas,res}^z \times \sum_{j \in z} R^j \quad (5.7)$$

Ce coefficient correctif trouvé, il est appliqué aux résistances thermiques équivalentes de chacun des clients résidentiels.

Concernant le réglage de la température de consigne par client  $T_{cons,res}^j$ , celle-ci est réglée de façon à minimiser les erreurs entre la courbe thermosensible simulée et celle mesurée. Dans un premier temps, la température de consigne est prise constante et égale pour tous les clients.

$$T_{cons,res}^z = \operatorname{argmin} \left( \sqrt{\sum_{t \in an} (Pts_{meas}^z(t) - Pts_{simu}^z(t))^2} \right) \quad (5.8)$$

Avec  $an$  correspondant à la série de dates sur l’année étudiée,  $Pts_{meas}^z$  la partie thermosensible de la courbe de charge mesurée du départ HTA desservant la zone  $z$ ,  $Pts_{simu}^z$  la partie thermosensible de la courbe de charge simulée.

## 5.5 Conclusion et perspectives du chapitre

Dans ce chapitre, nous avons proposé des méthodes d’évaluation et de calibration du simulateur de courbes de charge.

L’étude bibliographique sur l’évaluation et la calibration des modèles bottom-up a montré que c’est souvent les consommations en énergie ainsi que les profils qui sont évalués et calibrés. Ceci est effectué par les auteurs à partir de données de mesures individuelles (au plus quelques centaines de courbes de charge). Dans le cadre de la thèse, nous disposons de milliers de mesures agrégées à l’échelle des départs HTA sur plusieurs années. Notons que ceci sera à reconsidérer lorsque les compteurs intelligents seront pleinement déployés en France et que leurs mesures seront disponibles.

Pour la partie calibration en énergie du secteur résidentiel, nous nous sommes basés sur des données nationales de consommations par usage issues du CEREN [22] et d’EDF [7]. En comparant les énergies simulées à celles mesurées par les départs HTA, nous nous sommes rendu compte d’une surestimation importante, de l’ordre de 20%, de l’énergie non thermosensible résidentielle simulée. Celle-ci n’a pas encore trouvé d’explication, mais des évaluations de l’énergie simulée sur des mailles plus fines (*e.g.* maille IRIS) sont en cours pour trouver les causes de cette surestimation.

Concernant le secteur tertiaire, à notre connaissance, la seule étude à grande échelle permettant de faire les mêmes calibrations en énergie que pour le secteur résidentiel est l’“Enquête sur les consommations d’énergie dans le tertiaire / ECET” [11]. Cependant, celle-ci n’est pas accessible au public. Une demande au service du secret permettrait éventuellement d’accéder à l’étude pour calibrer au mieux le secteur tertiaire du simulateur de courbes de charge. Il en est de même pour le secteur industriel, avec l’enquête sur les consommations d’énergie dans l’industrie / EACEI [12]

Cette calibration en énergie effectuée, nous nous sommes basés sur une décomposition de la courbe de charge de chaque départ HTA en une partie thermosensible et une partie non thermosensible pour évaluer le simulateur de courbes de charge sur les données d’Enedis.

Pour la partie non thermosensible, nous avons développé une méthode de référence donnant une borne basse des niveaux d’erreur des profils non thermosensibles des départs HTA. Nous avons ensuite évalué les profils simulés par le simulateur de courbes de charge et constaté que les résultats étaient en adéquation avec la référence, bien que les erreurs soient plus élevées et plus étendues. Ces erreurs pourront être améliorées en modifiant les profils d’entrée du simulateur de courbes de charge, qui sont pour le moment des données d’entrée par défaut issues de campagnes de mesures. Ces modifications pourront être faites dans un premier temps en réglant les plages d’heures creuses et en ajustant le lancement des chauffe-eau en conséquence. Ensuite, des modifications automatiques des profils par défaut journaliers, hebdomadaires et annuels pourront être effectuées afin de minimiser les erreurs de simulation. Dans la perspective d’un simulateur de courbe de charge intégrant des modèles d’occupation des locaux, ceux-ci pourront aussi être calibrés en fonction des profils de courbes de charge mesurées.

Concernant la partie thermosensible, nous nous sommes intéressés uniquement au chauffage

électrique. A partir de régressions sur les données d'Enedis, nous avons déterminé une méthode permettant d'estimer, pour chaque départ HTA, la proportion de la contribution à la thermosensibilité de chaque catégorie de clients.

Pour la simulation des clients tertiaires et industriels, nous avons supposé que chaque client tertiaire ou industriel possède un chauffage électrique et une isolation moyenne, et avons ajusté les coefficients de performance des chauffages électriques de façon à retrouver les thermosensibilités de chaque catégorie. Notons que cette méthode ne prend pas en compte les spécificités locales de la thermosensibilité des clients tertiaires. Elle suppose que les consommations de chauffage électrique sont uniformément distribuées chez les clients d'une même catégorie, en étant pondérées par la surface des clients. Parmi les pistes d'amélioration, on peut citer la prise en compte de la saisonnalité, comme on a pu le voir de façon marquée avec une thermosensibilité positive pour la catégorie de clients hôtel-restaurant. Une optimisation globale sur toute la courbe de charge mesurée des différents paramètres par défaut aussi bien pour le résidentiel que pour le tertiaire fait aussi partie des champs d'étude à mener.

Une méthode de calibration de la partie résidentielle a enfin été proposée. Elle consiste à calculer par départ HTA un correctif global sur les résistances thermiques équivalentes des clients, de façon à obtenir la thermosensibilité résidentielle issue des mesures. Enfin, les températures de consigne des chauffages électriques sont réglées de façon à minimiser les erreurs de simulation par rapport aux mesures. Cette méthode nécessite de connaître la thermosensibilité et la consommation thermosensible du départ HTA.

Dans une logique prospective, pour simuler des départs HTA non mesurés par exemple, une étude de prévision *a posteriori* des valeurs obtenues pour la calibration de ces paramètres sur la grande quantité de données fait l'objet de perspectives notables. En effet, les valeurs correctives pourraient être mises en lien avec les informations de la zone desservie qui sont issues des bases de données vues au chapitre 3, et des modèles prédictifs pourraient être développés.

Enedis dispose de relevés de compteurs de ses clients qui sont effectués souvent plusieurs fois par an. L'analyse de l'évolution de l'énergie consommée par client avec les relevés de compteur permettrait d'autres calibrations du simulateur de courbes de charge, par exemple sur le nombre de lancements sur la période considérée ou sur la thermosensibilité du client. L'arrivée de données provenant de compteurs individuels intelligents "linky" constituera aussi une opportunité importante de validation et calibration des différents modèles. Les courbes de charge individuelles peuvent donner de nombreuses informations sur chaque client : par exemple la possession d'un chauffage électrique, l'utilisation des appareils en heure creuse, l'estimation des périodes de vacances, l'évolution de l'énergie journalière consommée selon l'année, la puissance des appareils en veille.



## Chapitre 6

# Simulations prospectives pour l'aide à la planification

### Résumé du chapitre

*Les gestionnaires du réseau de distribution utilisent des méthodes de prévision de la consommation électrique future pour dimensionner leur réseau. Elles se basent majoritairement sur des mesures passées et des tendances économétriques ou démographiques. Dans ce chapitre, nous décrivons la méthode traditionnelle utilisée par Enedis pour planifier son réseau HTA et BT, pour comprendre l'intérêt d'utiliser le simulateur de courbes de charge en complément de ces méthodes. Nous détaillons une manière d'utiliser le simulateur de courbes de charge à ces fins. Nous décrivons les principaux facteurs influençant la consommation future, et des pistes pour modéliser leur évolution. La planification du réseau de distribution doit aussi prendre en compte l'intégration de plus en plus importante des énergies renouvelables décentralisées. Le simulateur de production renouvelable utilisé par le centre PERSEE est donc aussi présenté ainsi que son utilisation avec différents scénarios. Un bilan sur la démarche de simulations prospectives est finalement proposé et la contribution des différentes parties de la thèse dans cette démarche est soulignée.*

*Deux cas d'étude sont décrits et mettent en pratique la démarche et les outils développés pendant la thèse.*

*Le premier est un cas d'étude sur 10 départs HTA de la ville de Brest. Il illustre les différentes fonctionnalités issues de l'association des bases de données et de l'utilisation du simulateur de courbes de charge par rapport aux méthodes traditionnelles d'Enedis. Il montre aussi les améliorations à apporter au modèle et à la démarche, notamment sur la calibration en énergie et la qualité des mesures utilisées pour la calibration.*

*Le deuxième cas d'étude, effectué sur deux départs HTA dans la région de Lyon, propose d'étudier l'impact d'un fort taux de pénétration d'énergies renouvelables sur un critère de dimensionnement du réseau HTA. Les résultats suggèrent qu'en ne considérant que le critère du nombre d'heures de dépassement de la puissance maximale, les énergies renouvelables peuvent être introduites de façon importante sans renforcement du réseau. Dans le cas de l'éolien, une forte pénétration peut même amener à une diminution de la contrainte liée à ce critère. Ces résultats sont cependant à nuancer car l'étude ne tiens pas compte d'autres problématiques comme celles de tension ou de déséquilibre des phases.*

## 6.1 Introduction du chapitre

Après avoir validé le modèle de consommation électrique, celui-ci peut être utilisé pour différents objectifs, que nous avons recensés au 4.1.1. Dans ce chapitre, nous proposons quelques pistes pour utiliser le simulateur de courbes de charge pour simuler la consommation électrique future à des fins d'aide au dimensionnement du réseau de distribution.

### 6.1.1 Contexte du dimensionnement du réseau de distribution

Les gestionnaires du réseau de distribution (GRD) doivent s'assurer de desservir en électricité leurs différents clients. Les ouvrages (*e.g.* lignes électriques, transformateurs, disjoncteurs) qui remplissent la fonction de desserte ont une durée de vie importante, souvent de plus de quarante ans, et ont des coûts d'investissement importants [51].

Avant d'installer ces ouvrages, le GRD doit s'assurer qu'ils seront capables de répondre à un certain nombre de critères pour leur dimensionnement. Parmi ces critères, on a la tenue en tension, en courant maximal, aux conditions extérieures (température, humidité, vent etc.), le vieillissement, le rendement (notamment car les pertes ont un coût sur toute la durée de vie de l'ouvrage) et le coût. Le GRD doit donc notamment estimer le courant maximal auquel sera soumis chacun des ouvrages sur le long terme. C'est une phase importante de la planification du réseau de distribution. Pour faire cela, les méthodes actuelles se concentrent sur l'estimation de la puissance maximale de la demande. Cependant, une estimation dynamique de la demande, sous forme de courbes de charge annuelles donnerait beaucoup d'informations utiles pour la planification. Cela permettrait notamment, en analysant l'allure temporelle et les différents contributeurs de la consommation, d'étudier de nouvelles stratégies comme la gestion de la demande [73], le stockage couplé aux énergies renouvelables [104], ou encore la diminution de la puissance maximale consommée par la production renouvelable locale [124]. Tout cela permettrait de quantifier et éventuellement limiter des investissements sur le réseau de distribution.

### 6.1.2 Méthode de planification du réseau d'Enedis

#### Planification du réseau HTA

La méthode de planification du réseau HTA d'Enedis est décrite en [146], et est résumée ci-dessous. Deux calculs de puissance sont utilisés pour le dimensionnement du réseau :  $P^*max$  et  $Ptmb$ . Ces valeurs correspondent à des valeurs de puissance relevées par périodes de 10 minutes. Notons que ces valeurs dimensionnantes servent ensuite d'entrée dans le modèle de calcul du réseau d'Enedis "ERABLE" [98].

- $P^*max$  est la puissance mesurée le jour le plus chargé de l'année. Cette valeur est corrigée en fonction de la température normale  $T_n^z$  de la zone considérée; définie comme étant la température moyenne statistiquement observée par les services de la météo le 15 janvier, sur plusieurs décennies.
- $Ptmb$  est la puissance estimée à la température minimale de base  $T_{mb}^z$  de la zone considérée (température froide dont la probabilité d'occurrence est de 1 jour par an).  $Ptmb$  est calculée à partir de  $P^*max$  qui est ramenée à la température minimale de base.

$P^*max$  sert pour le régime de secours, aussi appelé "situation N-1", quand le réseau est en situation d'incident. La puissance  $P^*max$  est révisée tous les ans grâce à la mesure des intensités transitant dans les départs HTA. Quand à  $Ptmb$ , celle-ci sert à dimensionner le réseau pour le régime normal, dans les conditions les plus sévères [98].

En pratique, la correction en température s'effectue en calculant la thermosensibilité du départ HTA considéré, comme décrit au 3.3.5. Ce calcul est rendu possible avec les mesures effectuées sur les départs HTA depuis des années (voir 3.2.1). On note  $P^z$  la consommation électrique d'un départ HTA sur un an,  $T^z$  la série temporelle de la température moyenne de la zone desservie par le départ HTA. Une fois la thermosensibilité  $Th^z$  du départ calculée, les puissances sont calculées comme il suit :

$$P^*max = P^z(t^*) + Th^z \times (T_n^z - T^z(t^*)) \quad (6.1)$$

$$Ptmb = P^*max - Th^z \times (T_{mb} - T_n^z) \quad (6.2)$$

Avec  $t^* = argmax(P^z)$ ,  $T_n^z$  et  $T_{mb}$  respectivement la température normale et la température minimale de base de la zone du départ HTA.

Grâce aux historiques de mesures et de températures, ces deux variables dimensionnantes peuvent être calculées sur plusieurs années. Cependant, en l'état, il n'est pas possible de simuler d'autres scénarios influençant la consommation (*e.g.* d'évolution socio-économique ou d'efficacité énergétique, voir 6.2.1). Grâce au simulateur de courbes de charge, au préalable évalué et calibré sur les départs HTA (voir chapitre 5), il est possible de faire d'autres scénarios, et de calculer les  $Ptmb$  et  $P^*max$  correspondantes [124].

## Planification du réseau BT

Bien que dans certains cas (*e.g.* sur la Côte d'Azur) la pointe de consommation apparaisse en été, la condition retenue comme la plus représentative pour la pointe de consommation est la température minimale de base ( $T_{mb}$ ) [146], que nous avons précédemment décrite.

Cependant, les départs BT ne sont pas ou peu instrumentés. Les seules mesures systématiques qui sont effectuées sont celles au niveau des compteurs individuels des clients, où des relevés sont effectués de façon annuelle ou éventuellement pluriannuelle.

Pour reconstruire la courbe de charge, Enedis se base sur un modèle probabiliste nommé Bagheera. C'est un modèle qui estime la courbe de charge théorique de chaque client en associant à chaque client un profil-type, à l'aide de ses caractéristiques contractuelles (*i.e.* puissance souscrite, activité du client, option tarifaire) [42]. Des exemples de courbes de charge reconstruites avec Bagheera à partir des profils-types des clients et de leurs caractéristiques sont présentés en figure 6.1.

La thermosensibilité de chaque client est estimée à l'aide des relevés de compteurs, comme illustré à la figure 8.1 de l'annexe 8.1.1. Enedis applique ensuite un profil thermosensible type correspondant à chaque type de clients.

Les différents profils-types correspondants ont été déterminés par Enedis lors de campagnes de mesures sur des clients individuels représentatifs de chaque type de clients avec des compteurs intelligents au pas de temps de 1h, ce pendant des années.

Cependant, ces profils sont statiques et leur représentativité a été déterminée avant d'effectuer les mesures. De plus, chaque client est associé à son profil-type uniquement par ses caractéristiques contractuelles. Les bases de données autres que celles d'Enedis ne sont pas utilisées dans l'attribution des profils. Les profils ne sont pas dépendants des facteurs socio-économiques, sur les locaux, les appareils et les comportements d'utilisation (voir les définitions en section 2.2.2), contrairement au simulateur de courbes de charge. Les résultats de simulation des flux de puissance BT de Bagheera ne sont pas calibrés sur les mesures de départs HTA, contrairement à l'outil de simulation de courbes de charge. Bagheera permet de calculer les puissances dimensionnantes uniquement avec

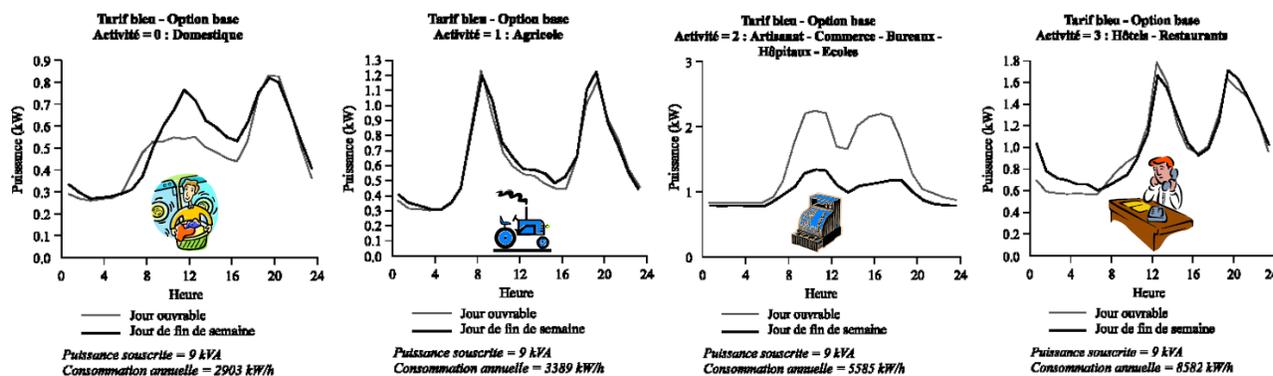


FIGURE 6.1 – Exemple de courbes de charge reconstruites à partir des profils-types des clients et de leurs caractéristiques (Source : Enedis)

des scénarios météorologiques différents, ainsi que l'ajout d'éventuels nouveaux clients, mais il ne permet pas la simulation d'autres scénarios car les profils ne sont liés qu'aux caractéristiques contractuelles des clients. Le simulateur de courbes de charge, quant à lui, permet, comme nous le présentons au 6.2.1, de simuler de nombreux autres scénarios (*e.g.* évolution socio-démographique ou de l'efficacité énergétique des appareils). Le simulateur de courbes de charge apparaît comme ayant des fonctionnalités intéressantes, mais la qualité de ses résultats doivent encore être étudiés à l'échelle des départs BT, notamment ceux ayant quelques dizaines de clients. Les pistes à explorer à ces échelles où le foisonnement est faible, s'orientent sur l'étude des puissances dimensionnantes probabilisées. Elles seront possibles notamment avec des données de courbes de charge individuelles.

### Dimensionnement avec la prise en compte des énergies renouvelables

Les énergies renouvelables installées sur le réseau de distribution contribuent au calcul du dimensionnement du réseau, particulièrement en cas de forte pénétration de celles-ci. En effet, il peut arriver dans le cas où il y a beaucoup de production et peu de consommation, à une inversion du flux de puissance, et la puissance de production soustraite à celle de consommation peut devenir la puissance dimensionnante. Cela est illustré à la figure 6.2.

Pour prendre en considération ce cas, Enedis calcule une troisième puissance dimensionnante  $P^*_{prod}$  calculée comme il suit [145] :

$$P^*_{prod} = P_{prod} - 0.2 \times P^*_{max} \quad (6.3)$$

Avec  $P_{prod}$  la puissance installée d'énergies renouvelables.

On constate que le calcul de cette valeur dimensionnante est assez simpliste et ne prend pas en compte les spécificités locales, notamment l'effet du foisonnement production/consommation ou le facteur de charge des installations. Il semble important pour Enedis d'avoir une simulation sous forme de courbes de charge de la consommation électrique mais aussi de production renouvelable pour pouvoir prendre en compte de façon plus fine les spécificités de chaque zone considérée.

### Conclusion sur la planification du réseau d'Enedis

Comme nous l'avons vu aux sous-sections précédentes, l'outil de simulation de courbes de charge présente de nombreuses autres possibilités de scénarisation des puissances dimensionnantes par rapport à leurs outils traditionnels. C'est pour cela que l'outil de simulation de courbes de charge

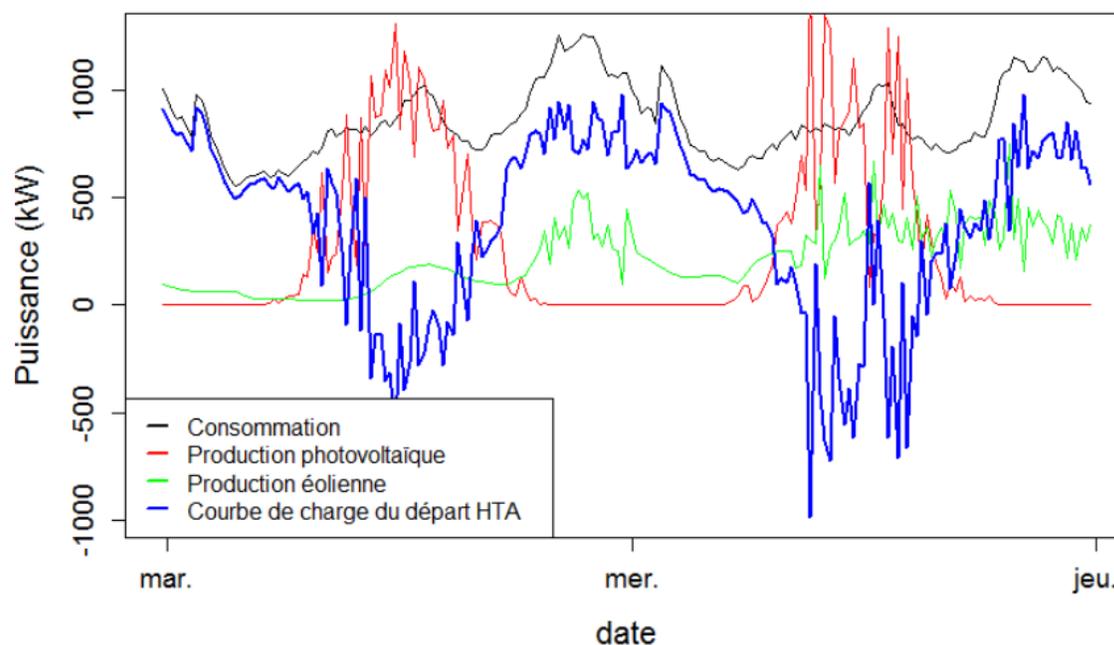


FIGURE 6.2 – Exemple illustratif d’une simulation d’un fort taux de pénétration d’énergie renouvelable sur un départ HTA où la production peut être dimensionnante. Simulation d’un départ HTA à Lyon avec 871 clients BT, 11 clients importants, 1,5 MW de puissance installée en photovoltaïque et 1.5 MW en éolien

développé pendant la thèse, appelé chez Enedis “MOSAIC” est utilisé dans de nombreux projets de planification. Il est particulièrement utile lors de grands projets de développements urbains, en collaboration avec les collectivités qui disposent de données précises notamment pour les nouveaux bâtiments construits (*e.g.* projet Transform avec Lyon métropole [73], projet avec la Brest métropole [98]).

## 6.2 Utilisation des outils de simulation et scénarisation

### 6.2.1 Utilisation du simulateur de courbes de charge

Afin de pouvoir simuler la consommation électrique d’une zone, il est nécessaire de disposer des données définissant chacun des consommateurs d’électricité de la zone. Nous avons vu au chapitre 3 les différentes bases de données disponibles utiles pour expliquer la consommation électrique. Nous avons vu au chapitre 4 quelles étaient les données d’entrée du simulateur, le réglage par défaut des différentes entrées, et les données qui devaient nécessairement être renseignées pour une simulation, avec leur influence sur la consommation par usage (section 4.2.4). En plus de ces données nécessaires, si d’autres informations supplémentaires sur les clients de la zone sont disponibles, les résultats n’en seront que meilleurs. Si la zone à simuler dispose de mesures de consommation électrique, une phase préalable de calibration du simulateur, en particulier de la partie thermosensible, améliorera grandement les résultats (voir 5.4.3).

Dans tous les cas, Enedis mesurant tous les départs HTA, il est intéressant, pour avoir les simulations les plus précises possible, d’effectuer en premier lieu une calibration de tous les clients du départ HTA, ce qui permet d’avoir des correctifs, notamment pour les parties thermosensibles, puis ensuite d’appliquer ces correctifs aux clients de la zone à simuler.

### Différentes scénarisations possibles

Pour simuler la consommation électrique d'une zone dans le futur, nous avons listé 5 grands types d'incertitudes. Dans ce paragraphe, nous les détaillons succinctement, en proposant quelques pistes pour prendre en compte et estimer ces incertitudes. Nous effectuerons par la suite deux cas d'étude prenant en compte ces scénarios.

- Incertitudes sur les évolutions des clients
  - Nouveaux clients et évolution des secteurs d'activité
  - Évolution du nombre de personnes par logement
  - Présence et localisation des bornes de chargement de véhicules électriques
- Incertitudes d'efficacité énergétique
  - Changement de vecteur énergétique
  - Rénovation de bâtiments
  - Évolution des appareils
  - Évolution des réglementations thermiques
- Incertitudes météorologiques (impact sur la partie thermosensible)
  - Évolution de la température
  - Évolution de l'irradiation solaire
- Incertitudes de comportement
  - Utilisation des véhicules électriques
  - Utilisation des appareils

**Nouveaux clients et évolution des secteurs d'activité.** Ces informations peuvent être estimées de façon locale en lien avec les collectivités avec les prévisions de nouvelles constructions, ou les tendances de dépôt de permis de construire. En l'absence d'information, des taux de croissance pris par défaut égaux aux valeurs nationales peuvent être utilisés, avec des caractéristiques des nouveaux clients qui peuvent par exemple être en premier lieu similaires aux clients déjà existants sur la zone.

**Évolution du nombre de personnes par logement.** Parmi les évolutions socio-économiques, une tendance de ces dernières dizaines d'années en France pour le résidentiel est à l'augmentation du nombre de ménages, mais à la diminution du nombre de personnes. Cela découle du vieillissement de la population et des changements dans les modes de vie (*e.g.* hausse des séparations, diminution des familles nombreuses) [53]. Le nombre de personnes par logement est intimement lié à la consommation des usages (voir notamment la figure 4.19). Cette information joue donc un rôle important dans la prévision de la consommation future.

**Présence et localisation des bornes de chargement des véhicules électriques.** Le parc de véhicules électriques est en pleine expansion, et bénéficie d'aides de la part de l'État français [15]. La consommation électrique des bornes de chargement est importante pour Enedis qui doit prévoir l'emplacement des différents points de recharges, aussi bien publics que privés. Pour ce faire, un premier estimatif du parc de véhicule électrique dans la zone considérée est effectué, puis les différentes bornes de chargement sont réparties sur les différents départements BT selon les scénarios possibles.

**Changement de vecteur énergétique.** La consommation de chauffage électrique contribue en général fortement à la consommation totale (voir notamment le cas d'étude 6.3). Des changements importants dans les vecteurs énergétiques (*i.e.* moyen de chauffage) peuvent avoir lieu, notamment quand il s'agit de projets de collectivités locales (*e.g.* mise en place de réseau de chaleur), et ainsi changer grandement la consommation électrique.

L'évolution de pompes à chaleur est aussi un élément-clé, car les coefficients de performance de ces dernières modifient les consommations de chauffage de façon importante. Une analyse du marché des ventes de pompes à chaleur, comme proposée par [58], permet d'avoir des estimatifs de la diffusion de cette technologie, bien que les répartitions par secteur d'activité et par territoire restent difficilement prévisibles. De plus, le nombre de pompes à chaleur vendues dans la grande distribution est difficilement estimable [58]. RTE propose des évolutions nationales des pompes à chaleur selon trois scénarios qui peuvent servir de première référence [127]. L'efficacité des pompes à chaleur en situation réelle représente aussi un challenge de modélisation important [77].

Enfin, la climatisation est un usage qui pourrait devenir dimensionnant pour le réseau électrique, en particulier dans les régions les plus chaudes (*e.g.* sud de la France). La climatisation est déjà en expansion importante suite aux vagues de chaleurs ces dernières années [127]. En outre, elle représente un enjeu sanitaire important. Les fortes températures impactent le fonctionnement biologique, pouvant aller jusqu'à augmenter le taux de mortalité moyens durant des vagues de chaleur [106]. Les projections du climat futur suggèrent que les épisodes chauds seront plus longs et intenses au cours du 21ème siècle [111]. L'usage de la climatisation, dépassant alors le confort pour devenir un nécessité en terme de santé, devrait donc être encore plus importante dans le futur [149]. Il est cependant difficile d'estimer le taux de croissance et la répartition géographique de cet usage.

**Rénovation de bâtiment.** Les rénovations de bâtiments réduisent les consommations de chauffage électrique. En France, différentes subventions sont proposées pour inciter à cette pratique [16]. De plus, la loi relative à la transition énergétique pour la croissance verte crée, à son article 14, une obligation de mettre en œuvre une isolation thermique à l'occasion de travaux importants de rénovation des bâtiments, comme un ravalement de façade, une réfection de toiture, ou encore la transformation de garages ou de combles en pièces habitables. Le décret est en vigueur depuis le 1er janvier 2017 [5], et ces rénovations sont à prendre en compte dans les scénarios de consommation électrique au long terme.

**Évolution des appareils.** La performance énergétique des équipements électriques s'améliore fortement : la plupart des équipements électriques des ménages sont soumis à des directives européennes sur l'écoconception et l'efficacité énergétique. RTE propose dans son bilan prévisionnel des estimatifs de consommation d'appareils selon leur étiquette énergie, comme illustré en figure 6.3. Bien qu'il soit difficile d'estimer le taux de renouvellement des appareils, des tendances nationales par rapport aux ventes des appareils peuvent être utilisées en valeur par défaut.

Consommation unitaire des équipements					
Blanc		Réfrigérateur	C 473 kWh	A+ 221 kWh	A+++ 119 kWh
		Congélateur indépendant	C 492 kWh	A+ 247 kWh	A+++ 135 kWh
		Lave-linge	C 251 kWh	A+ 183 kWh	A+++ 154 kWh
		Sèche-linge	C 434 kWh	A+ 242 kWh	A+++ 142 kWh
		Lave-vaisselle	C 303 kWh	A+ 216 kWh	A+++ 172 kWh
TIC		TV 42"	C 138 W	A+ 54 W	A+++ 24 W
		Informatique	Fixe 250 kWh	Portable 50 kWh	Tablette 10 kWh
Cuisson		Plaques électriques	Fonte 300 kWh	Vitrocéramique 230 kWh	Induction 180 kWh
		Four 60 L	C 315 kWh	A+ 155 kWh	A+++ 84 kWh
Éclairage		Lampe 800 lumens	Incandescent 60 W	Halogène Éco 43 W	LED 10 W

\* La consommation unitaire d'un ménage prend en compte les taux d'équipement et de multi-équipement moyens des ménages ainsi que les durées d'utilisation moyennes associées (estimations RTE). Les usages domestiques regroupent les usages froid & lavage (blanc), TIC, cuisson, éclairage.

FIGURE 6.3 – Estimation de la consommation annuelle par ménage équipé selon la technologie [127]

**Évolution des réglementations thermiques.** Une nouvelle réglementation thermique, la RT2020 sera appliquée aux futurs bâtiments neufs, avec pour objectif de rendre les bâtiments dits “à énergie positive” [14]. Ces nouvelles données sont à prendre en considération pour la simulation des bâtiments construits à partir de 2020. Dans le simulateur de courbes de charge, ces bâtiments sont simulés avec un coefficient de performance de 7 (voir figure 4.16).

**Évolution de la température et de l'irradiation solaire.** La température et l'irradiation solaire impactent la prévision de la consommation électrique. Le simulateur de courbes de charge utilise ces variables comme entrées, comme vu à la section 4.2.3. Une prévision précise de ces deux variables est donc requise pour minimiser l'incertitude dans les scénarios futurs de consommations électrique. Dans les cas d'étude proposés en sections 6.3 et 6.4, ces variables sont estimées à partir de l'historique mesuré (entre 10 et 30 ans). Or, le 21ème siècle est caractérisé par un climat changeant, avec notamment un réchauffement global des températures [111]. Pour une meilleure estimation de la consommation électrique, il pourrait être envisagé de prendre comme entrée des données issues de modèles climatiques afin de prendre en compte la dynamique du climat agissant sur la température et l'irradiation. Ces données peuvent notamment être proposées à des échelles temporelles et spatiales adaptées par des fournisseurs de services climatiques (eg Drias [6], C3S [4]).

**Utilisation des véhicules électriques.** Encore peu de données existent sur les comportements d'utilisation des véhicules électriques. Comme vu au 4.2.2, les hypothèses prises pour le modèle de comportement des véhicules électriques suivent celles prises par RTE [127].

**Utilisation des appareils.** Les comportements d'utilisation des appareils sont la résultante de nombreux facteurs, et il est fort possible que ceux-ci soient amenés à évoluer. Cependant, peu de données et d'études existent à ce sujet. Dans un premier temps, cette incertitude sera estimée en effectuant plusieurs simulations sur la zone à considérer. Nous supposons que les tirages aléatoires du simulateur représentent les variations de ces comportements.

## 6.2.2 Utilisation du simulateur de production renouvelable

### Description de l'outil

Le centre PERSEE a développé un outil permettant de simuler la production renouvelable normalisée d'une zone donnée en utilisant des données météorologiques telles que l'irradiation solaire sur plan horizontal, la température au sol et la vitesse du vent. Concernant la modélisation de la production photovoltaïque, les données d'irradiation utilisées proviennent de la base de données Helioclim 3 [48] qui permet d'avoir des données au pas de temps de 15min à la résolution spatiale de  $5\text{km} \times 7\text{km}$ . Le modèle de Muneer [144] est celui utilisé pour calculer la part d'irradiation diffuse sur la surface inclinée du module ainsi que pour le module physique du module. Les angles solaires, l'irradiation au sommet de l'atmosphère et les masses d'air optiques sont calculés à l'aide de la librairie C "solar\_geometry" [18]. Les détails sont disponibles en [49]. Les technologies des modules ainsi que les orientations et inclinaisons sont fixées par défaut en suivant des distributions statistiques de la littérature.

Pour la production éolienne, celle-ci est simulée à partir de données de vent venant de la base de données MERRA [17]. Ces données sont au pas de temps de 1h, à la résolution spatiale de  $50\text{km} \times 55\text{km}$ . La simulation peut être effectuée en rentrant les paramètres physiques de l'éolienne (*e.g* technologies de turbine, hauteur de la nacelle, orientation), qui permettra de générer la courbe de puissance et ainsi calculer l'énergie produite à partir des données de vent. Autrement, des courbes de puissance peuvent être renseignées, si elles sont connues. Les courbes de puissance par défaut du modèle sont calculées en prenant les courbes de productions régionales éoliennes couplées aux puissances installées et aux données de vent.

### Calibration sur la zone d'étude

Enedis dispose de mesures de production des producteurs HTA connectés au réseau de distribution. Connaissant la localisation de ces clients, la puissance installée et les courbes de charge, il est possible de calibrer pour adapter localement les paramètres d'orientation et d'inclinaison des modules ou des éoliennes [98].

### Différentes scénarisations possibles

L'incertitude sur la production renouvelable est le sujet de nombreuses études depuis des décennies (voir *e.g* [47], [95]).

Les données météorologiques utilisées étant à des échelles spatiales importantes (voir sous-section précédente 6.2.2), il est nécessaire pour avoir des courbes de productions plus réalistes

d'introduire de la variabilité spatiale. Une méthode d'introduction de variabilité pour la production photovoltaïque et une pour la production éolienne sont proposées par Rogeau *et al.* [124].

Pour la production photovoltaïque, la méthode consiste à définir 3 types de journées : une journée avec ciel clair, une nuageuse et une très nuageuse. Selon le type de journée, différentes variabilités sont introduites : une relativement faible pour les journées à ciel clair et nuageux, une variabilité importante pour le ciel nuageux. Les différents coefficients de variabilité sont ajustés sur des données de production de centrale mesurées.

Pour la production éolienne, la variabilité est introduite en ajoutant une variation dans la direction du vent à laquelle la ferme simulée est soumise. Tout comme la production photovoltaïque, les coefficients définissant la variabilité éolienne sont ajustés sur des données de production mesurées.

### 6.2.3 Bilan sur la démarche de simulations prospectives

A ce stade, nous avons tous les éléments pour effectuer la démarche de simulations prospectives qui utilise le simulateur de courbes de charge. Les différentes étapes de la démarche sont résumées sur la figure 6.4. Cette démarche est mise en pratique dans la section suivante, 6.3.

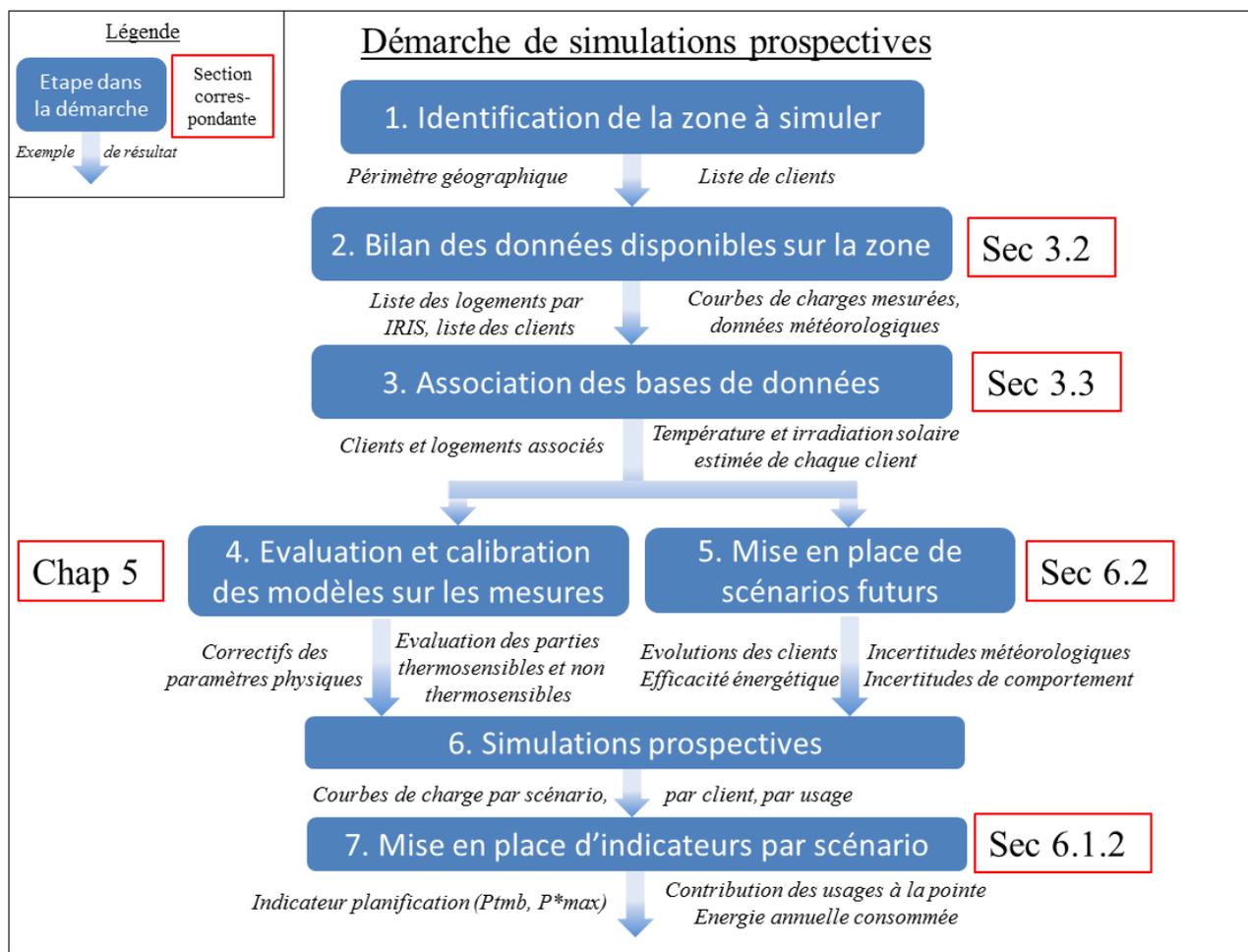


FIGURE 6.4 – Démarche globale de simulations prospectives développée pendant la thèse

### 6.3 Cas d'étude à Brest

Dans le cadre d'un projet effectué avec Enedis, la démarche développée pendant la thèse a été appliquée dans l'aide au dimensionnement de 10 départs HTA sur la ville de Brest en relation avec Brest métropole. Nous proposons un premier bilan sur les bases de données, leur association, la validation et la calibration du simulateur de courbes de charge puis son utilisation avec différents scénarios pour aider au dimensionnement des dix départs HTA. Ce cas d'étude a fait l'objet d'une publication dans une conférence internationale avec Enedis [98].

#### 6.3.1 Identification de la zone à simuler et données disponibles

##### Données d'Enedis

L'étude porte sur dix départs HTA, dont la liste des clients connectés a été fournie par Enedis pour l'année 2016. Les caractéristiques de ces départs sont recensées dans le tableau 6.1. Les mesures de ces départs ont été fournies par Enedis pour l'année 2012 au pas de temps de 10 minutes. La qualité des mesures sur cette année est tout à fait correcte, en moyenne les 10 départs HTA n'ont que 7,8% de réorganisations et valeurs manquantes (voir 3.2.1), avec un maximum à 16,2%.

TABLE 6.1 – Caractéristiques des 10 départs HTA étudiés

Caractéristique	Valeur minimale	Valeur moyenne	Valeur maximale
Nombre de clients desservis	357	751	1471
Fraction de consommation résidentielle	34%	62%	91%
Valeurs manquantes et réorganisations	0,5%	7,8%	16,2%

##### Données de l'INSEE

La zone desservie par les 10 départs compte 27 IRIS de la ville de Brest, ainsi que 12 IRIS extérieurs à la ville de Brest. Les données du recensement de la base logement de l'INSEE permettent d'apporter de nombreuses informations utiles à la simulation de type bottom-up (voir 3.2.2). Nous prenons les données du recensement de 2013, qui sont les données les plus récentes du recensement à la date de la thèse.

Le tableau 6.2 présente les caractéristiques principales des ménages des 39 IRIS. On constate que d'un IRIS à l'autre, les caractéristiques des ménages varient grandement. Les zones à simuler peuvent aussi bien être des zones fortement urbaines que des quartiers de banlieue avec des maisons individuelles.

TABLE 6.2 – Récapitulatif des différentes caractéristiques des logements des 39 IRIS du cas d'étude de la ville de Brest

Quantile	0%	25%	50%	Moyenne	75%	100%
Nombre de personnes	54	1726	2193	2123	2699	4066
Nombre de personnes par logement	1,45	1,69	2,20	2,13	2,47	2,79
Taux de possession du chauffage électrique	0,01	0,22	0,31	0,31	0,40	0,65
Surface moyenne par logement (m <sup>2</sup> )	57,6	71,6	97,6	91,1	105,8	128,2
Proportion de résidences principales	0,72	0,88	0,93	0,92	0,96	1,00
Proportion de maisons individuelles	0,00	0,16	0,33	0,49	0,90	1,00

## Données météorologiques

Concernant les données de température, celles-ci proviennent d'une station météorologique à Brest fournie par Enedis, et issue de Météo France. Les mesures fournies vont de l'année 1983 à 2015, et sont au pas de temps de 1h. Concernant l'irradiation, qui est une entrée du modèle de chauffage du simulateur (voir figure 4.18), celle-ci est disponible avec la base de données Helioclim 3 [48], de 2005 à 2015 au pas de temps de 15 minutes. Pour les années précédentes, l'année médiane d'irradiation a été utilisée. La mise au pas de temps de 1h des données d'irradiation s'est faite en moyennant les valeurs à 15min.

### 6.3.2 Association de la base Enedis et INSEE

La méthode d'association par l'algorithme d'optimisation permet d'associer à chaque client un logement et vice-versa. Nous appliquons ici la méthode décrite à la section 3.3.4, qui permet ainsi d'obtenir une liste de logements associés par départ HTA. Le tableau 6.3 récapitule les différentes caractéristiques des logements associés à chaque départ HTA.

TABLE 6.3 – Récapitulatif des différentes caractéristiques des logements des 10 départs HTA du cas d'étude de la ville de Brest

Quantile	0%	25%	50%	Moyenne	75%	100%
Nombre de foyers	306	470	600	637	796	957
Nombre de personnes	798	1064	1225	1295	1543	1972
Nombre de personnes par foyer	1,43	1,70	2,25	2,12	2,50	2,65
Taux de possession chauffage électrique	0,26	0,28	0,31	0,32	0,35	0,45
Surface moyenne par logement (m <sup>2</sup> )	66,8	90,0	102,7	100,5	113,6	126,2
Proportion de résidences principales	0,71	0,91	0,94	0,91	0,96	0,99
Proportion de maisons individuelles	0,02	0,12	0,36	0,36	0,53	0,82

### 6.3.3 Évaluation et calibration du simulateur de courbes de charge sur les mesures

#### Partie non thermosensible

La vérification de la qualité des profils des simulations effectuées par le simulateur de courbes de charge est une étape importante, décrite à la section 5.4.2. Pour ce faire, les courbes de charge de 2012 de chaque départ ont été au préalable décomposées en une partie thermosensible et une partie non thermosensible. Ensuite, les résultats de la simulation de la partie non thermosensible du simulateur ont été comparés à ceux de la partie non thermosensible mesurée.

Les résultats sont présentés dans le tableau 6.4. On constate que certains départs HTA ont des erreurs quadratiques moyennes normalisées (NRMSE) de simulation sur les profils non thermosensibles importantes, supérieures pour plus de la moitié à 30%. De même, la différence relative entre l'énergie simulée sur un an et celle mesurée est assez importante, de l'ordre de 18.6% en valeur absolue. En analysant plus en détail les courbes de charge, les erreurs importantes peuvent s'expliquer par des comportements erratiques dans les mesures. La figure 6.5 illustre cela en montrant le tracé de trois semaines typiques pour un départ à erreur médiane. On voit sur cette figure que la mesure est beaucoup moins régulière que la simulation. Il faudra donc prendre les résultats des départs avec ces taux d'erreur avec précaution. A l'inverse, certains autres départs HTA montrent de résultats bien meilleurs, comme illustré en figure 6.6. On constate une sous-estimation des pics

TABLE 6.4 – Résultats de l'évaluation de la partie non thermosensible des départs HTA du cas d'étude

Quantile	NRMSE profil thermosensible (%)	Différence énergie simulée/mesurée (%)
0%	17,2	-28,6
25%	23,4	-17,1
50%	30,1	-3,9
Moyenne de la valeur absolue	30,2	18,6
75%	36,3	16,6
100%	44,8	31,0

des heures creuses dans les deux cas, qui peuvent être corrigés notamment par un réglage adéquat du lancement des chauffe-eau, qui n'a pas été effectué dans cette étude.

### Partie non thermosensible

La méthode de calibration de la partie thermosensible décrite au 5.4.3 a été appliquée sur les deux départs HTA du cas d'étude. Ainsi, pour chaque départ, les coefficients de performance des différents types de tertiaire ont été déterminés, de même que le facteur correctif pour les résistances thermiques et la température moyenne de consigne pour le résidentiel.

#### 6.3.4 Mise en place de scénarios prospectifs

Dans ce cas d'étude, l'incertitude météorologique a fait l'objet de scénarisations. Pour ce faire, 33 années de température ont été testées pour constater leur influence sur les indicateurs utilisés pour la planification par Enedis.

Tous ces scénarios, couplés aux résultats de l'évaluation et de la calibration, permettent de simuler des courbes de charges et de calculer les différents indicateurs.

#### 6.3.5 Résultats des simulations et calculs des indicateurs

Le simulateur de courbes de charge, calibré pour l'année 2012 est ensuite utilisé avec la même liste de clients mais sur 33 années météorologiques. Le poste HTA/BT auquel chaque client est rattaché est connu par Enedis, et le simulateur de courbes de charge simulant de façon bottom-up client par client, il est ainsi possible d'estimer la courbe de charge de chaque poste HTA/BT, en sommant la consommation de chaque client du poste en question. Cela est illustré en figure 6.7, où l'on voit pour un départ HTA de la ville de Brest la décomposition de la consommation simulée et calibrée sur différents postes.

Ainsi, il est alors possible de calculer les puissances dimensionnantes pour Enedis,  $P_{tmb}$  et  $P^*_{max}$  non seulement selon les différents scénarios par départ HTA, mais aussi par poste HTA/BT.

Pour le dimensionnement d'un poste HTA/BT, on peut prendre comme puissance de référence le  $P^*_{max}$  calculé pour l'année 2012, année sur laquelle on a effectué la calibration. Ensuite, on recalcule le  $P^*_{max}$  pour chaque scénario météorologique. Ces valeurs servent ensuite d'entrée dans le modèle de calcul du réseau d'Enedis "ERABLE" [98]. La figure 6.8 donne la fréquence de dépassement de la puissance de référence par année météorologique pour un poste HTA/BT. On voit notamment que l'année 1985 a été particulièrement froide, menant à un dépassement de la puissance de référence de 0.3% (soit environ 26h par an). En moyenne, la fréquence de dépassement pour ce poste est de 0.082%, soit environ 7h par an.

### Tracé de 3 semaines représentatives pour un départ HTA de la zone de Brest

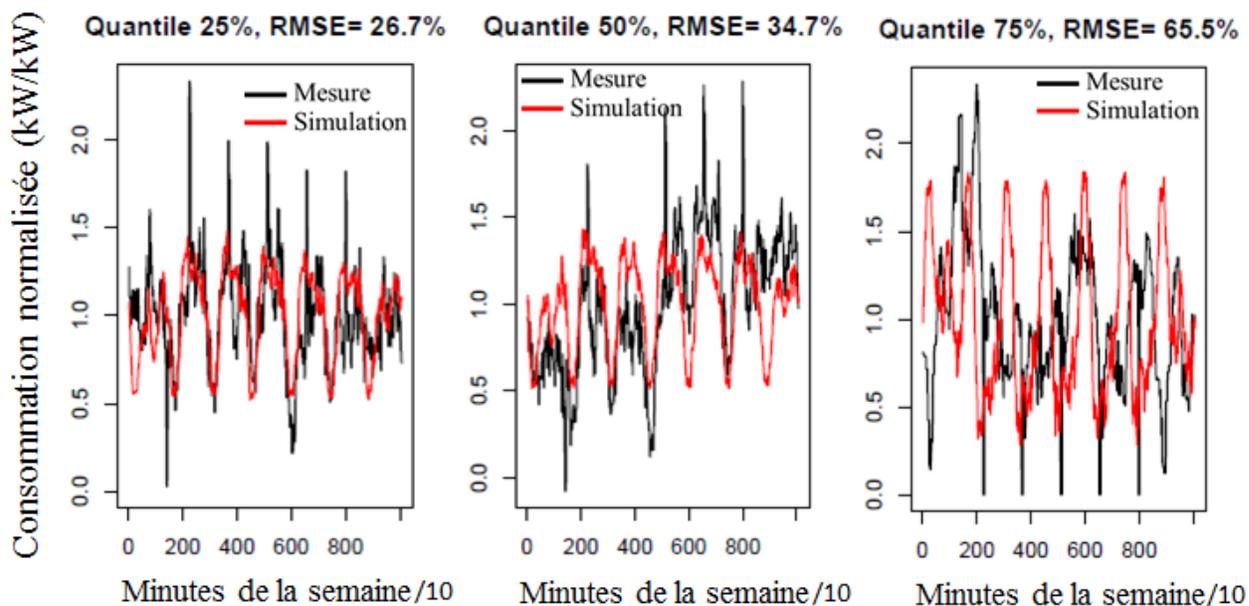


FIGURE 6.5 – Tracé de trois semaines représentatives d'un départ HTA à erreur médiane

### Tracé de 3 semaines représentatives pour un départ HTA de la zone de Brest

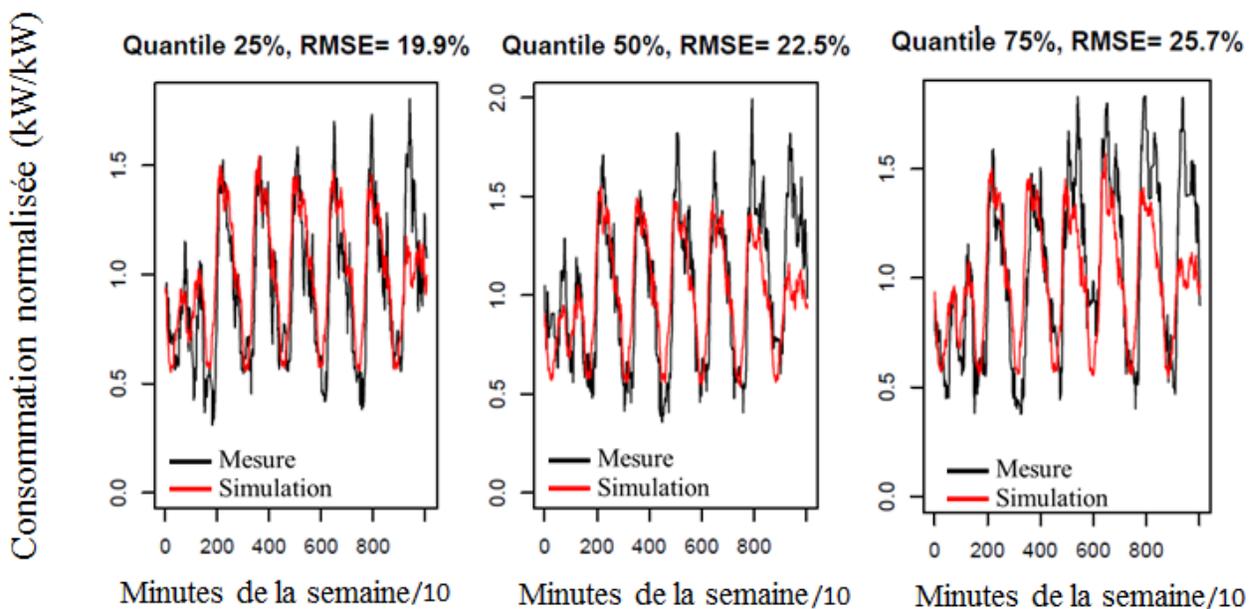


FIGURE 6.6 – Tracé de trois semaines représentatives d'un départ HTA à erreur dans le dernier quartile

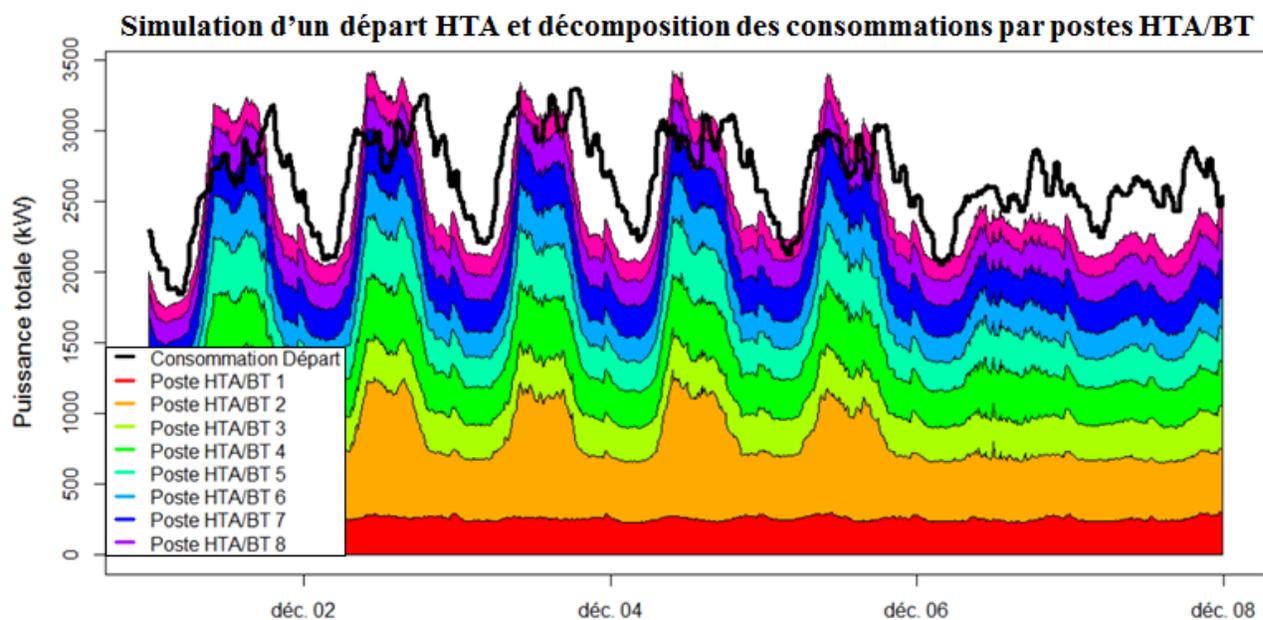


FIGURE 6.7 – Tracé sur une semaine de la consommation simulée d'un départ HTA avec la décomposition des consommations par postes HTA/BT

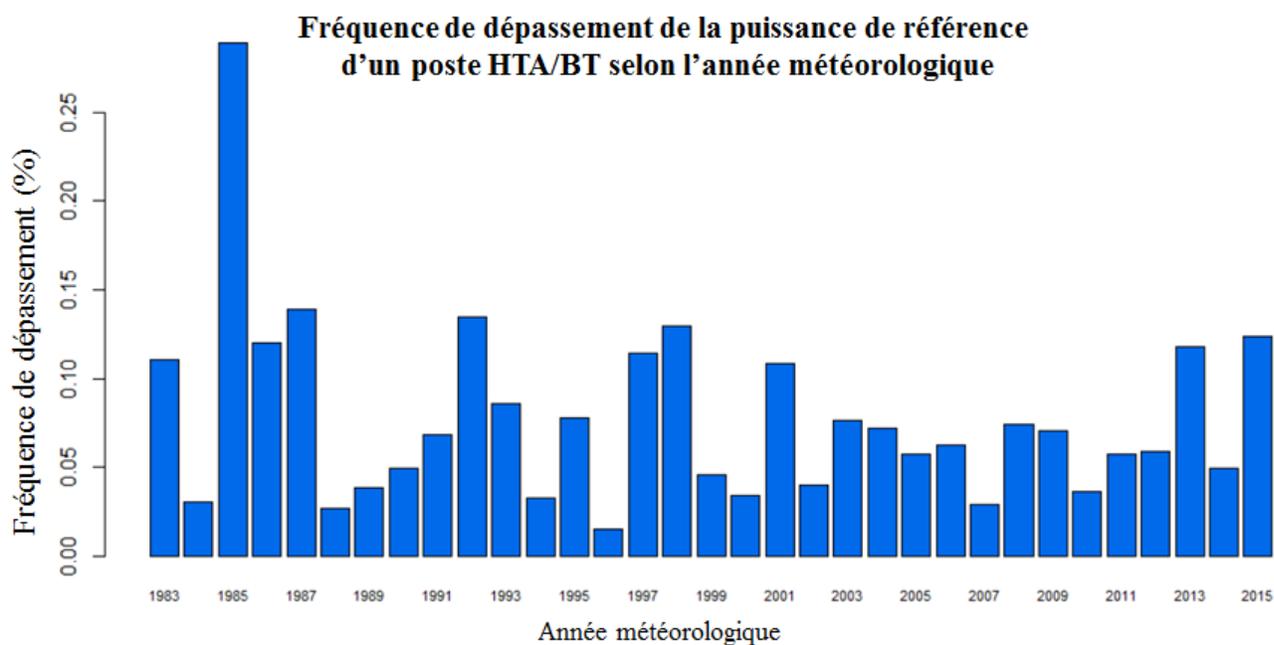


FIGURE 6.8 – Fréquence de dépassement de la puissance de référence d'un poste HTA/BT selon l'année météorologique [98]

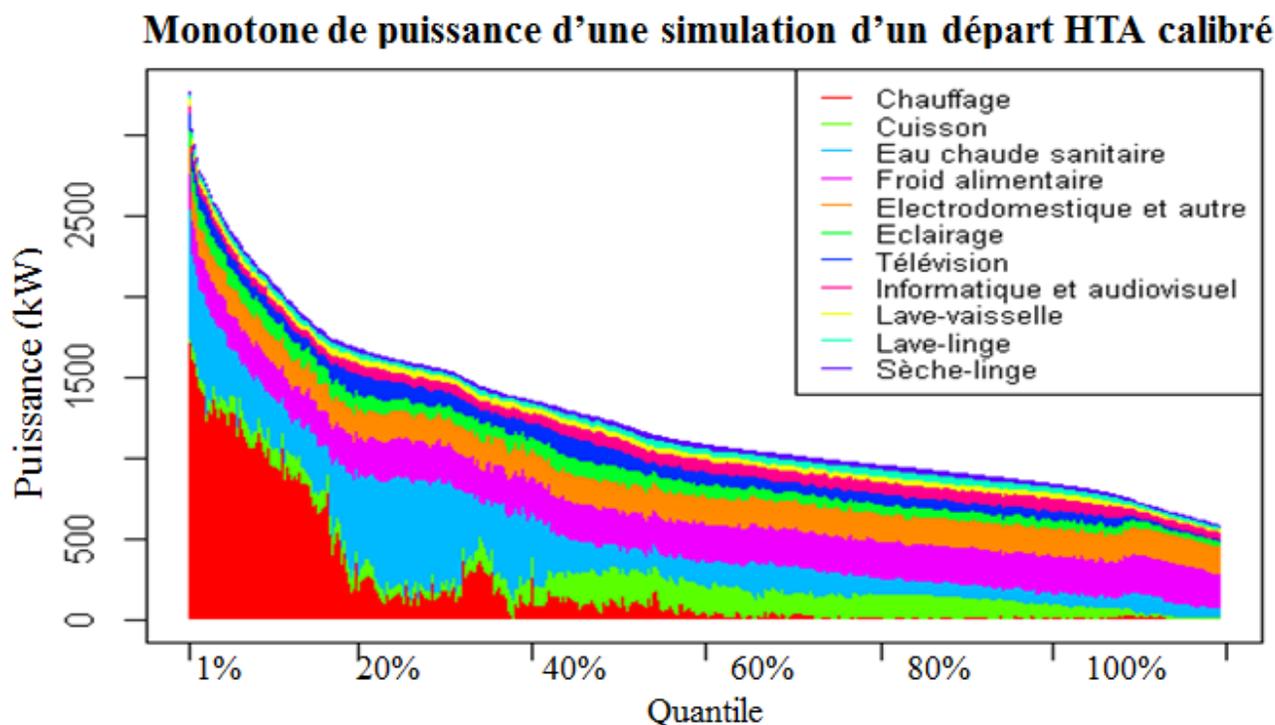


FIGURE 6.9 – Monotone de puissance dans un départ HTA avec des clients majoritairement résidentiels

Enfin, soulignons que l'analyse de la contribution à la pointe de chaque usage est un point clé pour le dimensionnement, notamment comme aide à la décision sur les choix de projets d'urbanisme. La figure 6.9 nous renseigne sur la contribution de chaque usage dans la monotone de puissance dans un départ HTA constitué de clients à majorité résidentiels. On peut ainsi mettre en évidence que dans ce cas, le chauffage électrique est clairement l'usage dimensionnant principal, et que tous les scénarios modifiant les paramètres en lien avec la consommation de chauffage (*e.g.* isolation des bâtis, taux de pénétration de chauffage électrique) influenceront grandement sur le dimensionnement du réseau électrique.

## 6.4 Cas d'étude sur l'évolution des critères dimensionnant selon le taux de pénétration des énergies renouvelables

### 6.4.1 Introduction

Le déploiement des énergies renouvelables comme moyen de production d'électricité est un point clé comme stratégie d'atténuation au changement climatique [110]. Dans le cas de la production photovoltaïque, la capacité installée en 2020 devrait être deux fois supérieure à celle prévue par la feuille de route prévue effectuée en 2010 [60]. En France en 2016, la filière éolienne et la filière solaire photovoltaïque concentrent plus de 40 % des capacités en énergie renouvelable [128]. En 2016, la capacité installée pour le photovoltaïque est actuellement de 6 772MW, dont 85% installé sur le réseau d'Enedis. Pour l'éolien, la puissance installée est de 11 670MW, dont 89% sur le réseau d'Enedis. Les projets à venir prévoient des installations à majorité sur le réseau d'Enedis : 2 279MW sont prévus pour le photovoltaïque dont 92% sur le réseau Enedis ; 11 397MW sont prévus pour

l'éolien dont 49% sur le réseau Enedis [128]. Ainsi, le réseau électrique de distribution d'Enedis accueille de plus en plus de production renouvelable décentralisée. Nous avons déjà constaté les limites des méthodes actuelles de dimensionnement du réseau en prenant en compte les énergies renouvelables en section 6.1.2. Dans ce cas d'étude, nous étudions comment se comporte un indicateur de dimensionnement avec un nombre important de scénarios météorologiques couplé aux incertitudes de comportement. Ce cas d'étude a fait l'objet d'une publication dans une conférence internationale avec Enedis [124].

## 6.4.2 Identification de la zone à simuler et données disponibles

### Données d'Enedis

Pour ce cas d'étude, 2 départs HTA desservant principalement des zones résidentielles de la région de Lyon ont été retenus. Leurs caractéristiques sont listées dans le tableau 6.5.

TABLE 6.5 – Caractéristiques des 2 départs HTA étudiés

Nom du départ HTA	a	b
Nombre de clients	2509	6982
Puissance souscrite totale (MVA)	20,25	51
Fraction de consommation résidentielle	80%	73%
Valeurs manquantes et réorganisations	18%	5%

### Données de l'INSEE

La zone desservie par les 2 départs compte 5 IRIS répartis autour de la ville de Lyon. Les données du recensement de la base logement de l'INSEE permettent d'apporter de nombreuses informations utiles à la simulation de type bottom-up (voir 3.2.2). Nous prenons les données du recensement de 2013, qui sont les données les plus récentes du recensement à la date de la thèse.

Le tableau 6.6 présente les caractéristiques principales des ménages des 5 IRIS. On constate que les zones à simuler sont à majorité urbaines ou péri urbaines, avec un taux relativement faible de maisons individuelles. Le nombre de personnes par logement quant à lui, ne varie que peu selon les IRIS.

TABLE 6.6 – Récapitulatif des différentes caractéristiques des logements des 5 IRIS du cas d'étude de la ville de Lyon

Quantile	0%	25%	50%	Moyenne	75%	100%
nombre de personnes	2338	2493	5177	11076,6	20194	25181
nombre de personnes par logement	2,12	2,41	2,49	2,45	2,58	2,67
Taux de possession du chauffage électrique	0,20	0,32	0,34	0,33	0,36	0,42
Surface moyenne par logement (m <sup>2</sup> )	74,7	92,6	97,9	98,9	114,5	114,6
Proportion de résidences principales	0,90	0,94	0,94	0,93	0,94	0,94
Proportion de maisons individuelles	0,11	0,19	0,33	0,40	0,54	0,85

### Données météorologiques

Les données de température (entrée du simulateur de courbes de charge) et de vent (pour la production éolienne) proviennent de la base MERRA [17]. Les mesures fournies vont de l'année

2006 à 2013, et sont au pas de temps de 30min. L'irradiation solaire, nécessaire à la fois pour la simulation de la production photovoltaïque, et comme entrée du simulateur de courbes de charge, est disponible avec la base de données Helioclim 3 [48], de 2006 à 2013, au pas de 15min.

### 6.4.3 Association de la base Enedis et INSEE

La méthode d'association par l'algorithme d'optimisation permet d'associer à chaque client un logement et vice-versa. Nous appliquons ici la méthode décrite à la section 3.3.4, qui permet ainsi d'obtenir une liste de logements associés par départs HTA. Le tableau 6.7 récapitule les différentes caractéristiques des logements associés à chaque départ HTA.

TABLE 6.7 – Récapitulatif des différentes caractéristiques des logements des 2 départs HTA du cas d'étude de la ville de Lyon

Depart HTA	a	b
Nombre de logements	1957	5232
Nombre de personnes	4736	9600
Nombre de personnes par logement	2,42	1,83
Taux de possession du chauffage électrique	0,41	0,22
Surface moyenne par logement (m <sup>2</sup> )	113,3	74,0
Proportion de résidences principales	0,94	0,90
Proportion de maisons individuelles	0,20	0,84

### 6.4.4 Évaluation et calibration du simulateur de courbes de charge sur les mesures

#### Partie non thermosensible

La vérification de la qualité des profils des simulations effectuées par le simulateur de courbes de charge est une étape importante, décrite à la section 5.4.2. Pour ce faire, les courbes de charge de 2012 de chaque départ ont été au préalable décomposées en une partie thermosensible et une partie non thermosensible. Ensuite, les résultats de la simulation de la partie non thermosensible du simulateur ont été comparés à ceux de la partie non thermosensible mesurée.

Les résultats pour l'évaluation des profils non thermosensibles pour les deux départs sont présentés dans le tableau 6.8. Ces résultats illustrent la bonne qualité de la simulation de la partie non thermosensible. La figure 6.10 montre la simulation et les mesures d'une semaine à erreur médiane pour chacun des départs. On voit pour le départ a que le pic lié aux heures creuses n'est pas correctement modélisé dans le simulateur de courbes de charge, qui est l'objet de travaux actuels au centre PERSEE. Le simulateur ainsi évalué dans sa partie non thermosensible pour ces deux départs va pouvoir être utilisé pour différentes scénarisations comme aide au dimensionnement.

TABLE 6.8 – Résultats de l'évaluation de la partie non thermosensible des deux départs HTA du cas d'étude de la région de Lyon

Départ HTA	NRMSE profil thermosensible (%)	Différence énergie simulée/mesurée (%)
a	20,6	2,8
b	17,7	-15,9

### Tracé de la semaine à erreur médiane pour chacun des deux départs

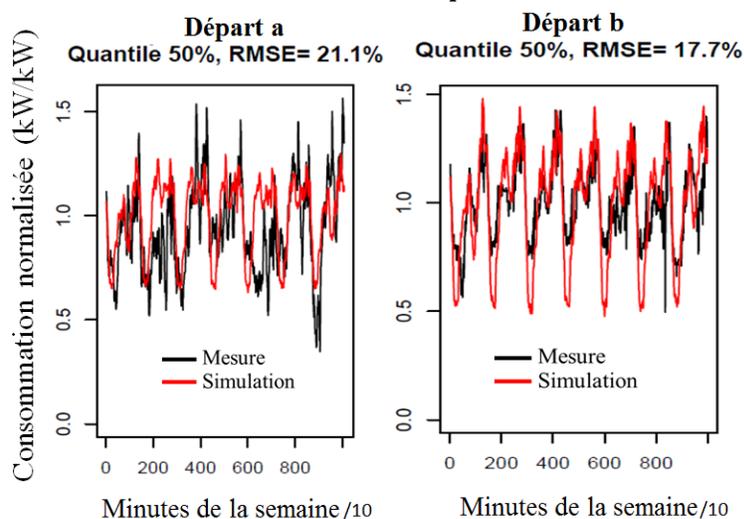


FIGURE 6.10 – Tracé des deux semaines médianes des deux départs HTA du cas d'étude de la région de Lyon

#### Partie thermosensible

La méthode de calibration de la partie thermosensible décrite au 5.4.3 a été appliquée sur les deux départs HTA du cas d'étude. Ainsi, pour chaque départ, les coefficients de performance des différents types de tertiaire ont été déterminés, de même que le facteur correctif pour les résistances thermiques et la température moyenne de consigne pour le résidentiel.

#### 6.4.5 Scénarisations

Dans ce cas d'étude, 8 années météorologiques sont testées (2006-2013) à la fois pour la simulation de production et la consommation.

De plus, 5 scénarios de comportements sont générés avec le simulateur de courbes de charge en relançant la simulation 5 fois. Au total, c'est donc 40 scénarios de consommation qui ont été générés. Notons les  $(Sconso_i)_{i \in 1,2,\dots,40}$ .

Pour la production renouvelable, une procédure d'ajout de variabilité décrite en 6.2.2 fait aussi l'objet de 8 scénarisations possibles. La procédure plus complète est décrite en [124]. Au total, c'est donc 64 scénarios de productions qui ont été générés. Notons les  $(Sprod_i)_{i \in 1,2,\dots,64}$ .

Finalement, en couplant production et consommation, on obtient 2560 scénarios pour chacun des 2 départs HTA.

#### 6.4.6 Critère de dimensionnement et influence du taux de pénétration d'énergies renouvelables

En intégrant de la production d'énergie renouvelable sur un départ HTA, celle-ci peut éventuellement réduire la puissance maximale. En ne considérant que le critère de dimensionnement correspondant au nombre d'heures de dépassement de la puissance à température minimale de base ( $P_{tmb}$  vue à la section 6.1.2), nous étudions ici quel serait le gain potentiel de dimensionnement sur le réseau en fonction de la pénétration de la production solaire et éolienne.

Afin de fixer une première référence, la température minimale de base  $T_{mb}$  (voir 6.1.2) a été calculée pour les 8 années météorologiques. Ensuite, chacun des deux départs a été simulé par le simulateur de courbes de charge pendant un an, soumis constamment à la température minimale de base. La puissance de référence dimensionnante  $P_{ref}^z$  pour chacun des deux départs desservant une zone  $z$  donnée a été choisie comme étant le maximum de cette courbe de charge simulée :

$$P_{ref}^z = \max[P^z(\text{rep}(T_{mb}, 52560))] \quad (6.4)$$

Avec  $\text{rep}(T_{mb}, 52560)$  signifiant la répétition de la température minimale de base pendant un an au pas de temps de 10min, soit 52560 points. Ensuite, un temps de dépassement moyen par an  $t_{dep}^z$  de la puissance de référence  $P_{ref}^z$  par départ a été calculé en prenant en compte les 40 scénarios de consommations par départ :

$$t_{dep}^z = \text{mean}_{i \in 1,2,\dots,40}(P^z(Sprod_i) > P_{ref}^z) \quad (6.5)$$

Enfin, les 64 scénarios de productions ont été combinés au 40 scénarios de consommation. Une nouvelle puissance dimensionnante  $P_{new}^z$  a été calculée comme étant la puissance pour laquelle on a le même temps de dépassement  $t_{dep}^z$  correspondant au scénario sans production. Afin de prendre en compte le cas où la production devient supérieure à la consommation, la puissance dimensionnante est calculée sur la valeur absolue de la courbe de charge.

Le gain en dimensionnement  $G$  avec apport de la production renouvelable est calculé comme il suit :

$$G^z = 1 - P_{new}^z / P_{ref}^z \quad (6.6)$$

Ce gain a été tracé pour différents taux de pénétration des énergies renouvelables, et les résultats sont présentés en figure 6.11, où seules les valeurs de gain positives ont été représentées.

On peut noter que pour les deux départs, le gain maximal est d'environ 4% de  $P_{ref}^z$ . Les deux départs HTA ont des comportements similaires dans l'évolution du gain en dimensionnement par rapport au taux de pénétration d'énergies renouvelables. On peut noter que l'influence de l'éolien est relativement importante alors que celle du photovoltaïque est quasi négligeable. Cela s'explique par le fait que le photovoltaïque produit dans la journée alors que le pic de consommation, élément dimensionnant sans production décentralisée, est principalement le soir. On peut aussi noter que quand la capacité de production est trop importante, il n'y a plus de gain sur le dimensionnement : la production devient dimensionnante. Dans ces deux départs HTA, on observe que l'on peut monter jusqu'à près de 40% de capacité de production renouvelable par rapport à celle de la consommation sur le réseau sans qu'elle devienne dimensionnante, selon l'indicateur calculé. Il est important de noter que d'autres problématiques entrent en jeu et n'ont pas été modélisées. Par exemple, les problématiques de contraintes en tension, de déséquilibre des phases, de déconnexion imprévue des moyens de production n'ont pas été considérées alors qu'elle peuvent limiter l'intégration des énergies renouvelables au réseau de distribution. Des solutions de stockages décentralisés peuvent pallier notamment aux problématiques de tension (voir *e.g.* [90], [157]).

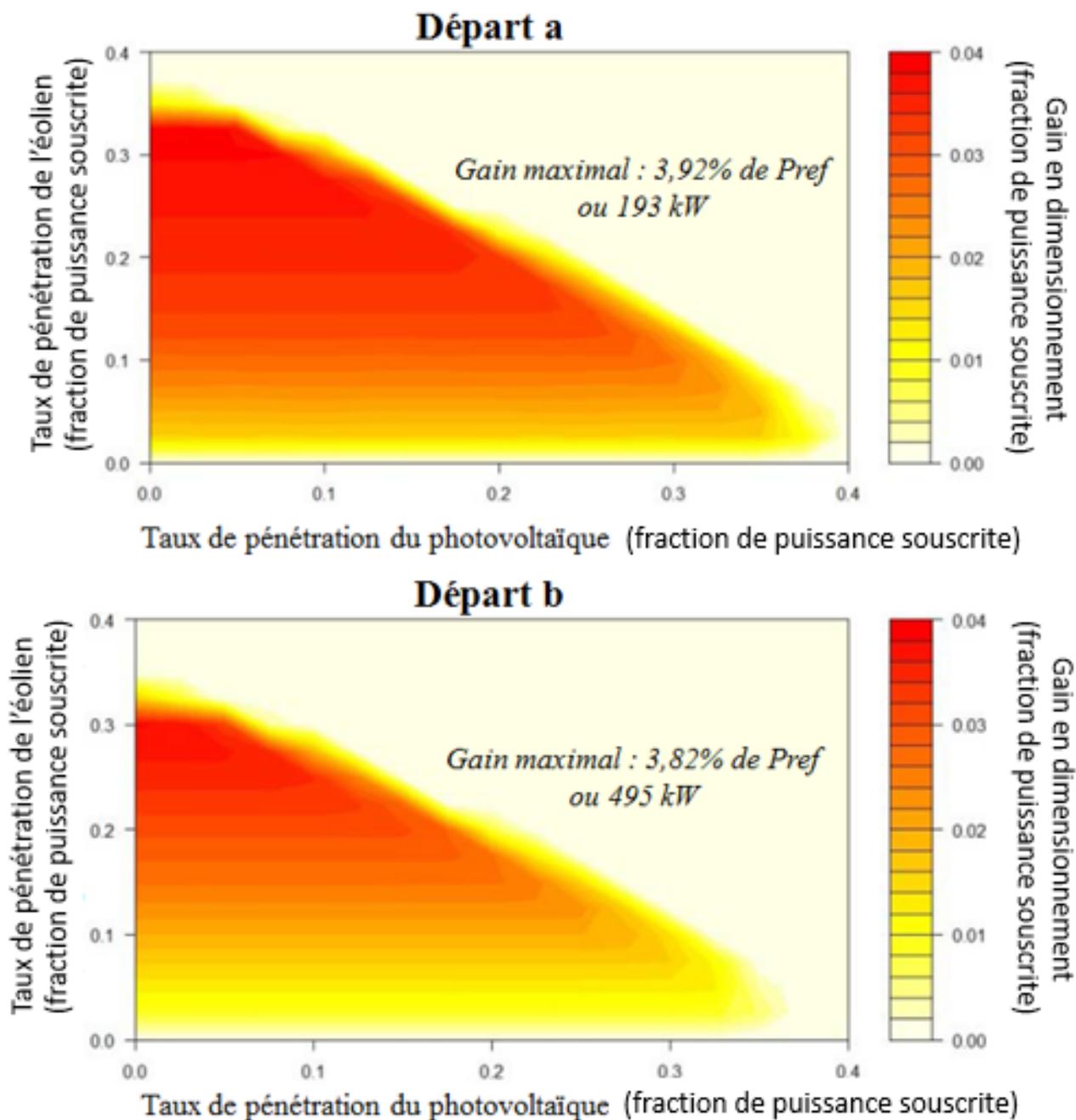


FIGURE 6.11 – Évolution du gain d'un critère de dimensionnement selon le taux de pénétration d'énergies renouvelables [124]

## 6.5 Conclusion et perspectives du chapitre

Ce chapitre a permis de mettre en application la démarche et les outils développés pendant la thèse. Nous avons dans un premier temps décrit comment utiliser le simulateur de courbes de charge pour simuler la consommation électrique future. Une telle simulation est utile notamment pour aider au dimensionnement du réseau de distribution. Pour cela, nous avons vu les méthodes traditionnelles de dimensionnement du réseau utilisées par Enedis, puis comment le simulateur de courbes de charges et toutes ses fonctionnalités peuvent compléter ces méthodes. Nous avons souligné que l'évolution future des facteurs impactant la consommation électrique est complexe et incertaine. Les méthodes traditionnelles se basant sur les tendances passées ne sont plus adaptées à un 21ème siècle caractérisé par des changements importants, notamment climatiques, sociaux et technologiques. Une démarche de simulations prospectives utilisant les différents outils et méthodes développés pendant la thèse a été présentée.

Enfin, deux cas d'étude ont été présentés et ont mis en pratique la démarche de simulations prospectives. Le premier a illustré les différentes fonctionnalités rendues possibles par l'association des bases de données et l'utilisation du simulateur de courbes de charge par rapport aux méthodes traditionnelles d'Enedis. Il a aussi montré les améliorations à apporter au modèle et à la démarche, notamment sur la calibration en énergie et la qualité des mesures utilisées pour la calibration. Le deuxième cas d'étude a proposé d'étudier l'impact d'un fort taux de pénétration d'énergies renouvelables sur un critère de dimensionnement du réseau HTA. Les résultats ont montré qu'en ne considérant que le critère du nombre d'heures de dépassement de la puissance maximale, les énergies renouvelables peuvent être introduites de façon importante sans renforcement du réseau. Dans le cas de l'éolien, une forte pénétration peut même amener à une diminution de la contrainte liée à ce critère. Ce travail ne tient cependant pas compte de nombreux autres critères comme le problème de tension ou de déséquilibre des phases.

La démarche mise en pratique dans ce chapitre pour la simulations prospectives ainsi que les cas d'étude n'ont pas vocation à être figés. Ils constituent plutôt une base pour les échanges. Les travaux à venir pourront étudier l'évolution d'indicateurs de dimensionnement sur les courbes simulées et mesurées sur plusieurs années et différentes données météorologiques afin de conforter la validité du simulateur de courbes de charge. Un travail d'analyse de sensibilité des indicateurs du dimensionnement vis à vis de différents scénarios est aussi une perspective importante de ce chapitre.

## Chapitre 7

# Conclusion et perspectives

Arrivé au terme de ce manuscrit, un bilan synthétique des travaux effectués est présenté. Les limites de ces travaux sont ensuite exposées, avant de proposer les principales perspectives.

### 7.1 Synthèse des travaux effectués

Cette thèse a proposé un modèle de consommation électrique de type bottom-up qui est alimenté, calibré et validé par de grandes quantités de données disponibles à l'échelle de la France. Ce modèle permet de mieux comprendre et prédire la consommation électrique à l'échelle du pays. La démarche proposée, correspondant au plan de la thèse, s'est voulue générale pour être reproductible pour d'autres régions ou pays.

Le panorama des modèles existants (chapitre 2) a montré qu'aucun modèle ne prenait en compte de façon globale la problématique de modélisation bottom-up de la consommation électrique à l'échelle d'un pays. Parmi les difficultés pour y arriver, figure le nombre important de données nécessaires. Une première réponse à ce challenge a été de proposer un bilan sous forme de classification des bases de données disponibles expliquant la consommation électrique (au chapitre 3). Parmi ces données classifiées, nous avons eu l'opportunité de travailler dans le cadre de cette thèse avec la grande quantité de données de mesures des départs HTA d'Enedis et leur liste des clients associés. Afin d'augmenter le potentiel de modélisation de la base des clients et mesures d'Enedis, nous avons enrichi la base de données des clients en y associant la base de données des logements de l'INSEE. Pour cela, une méthode d'association spécifique a été développée, visant à surmonter de façon statistiquement représentative la problématique de l'anonymisation des logements par l'INSEE. Nous avons montré l'intérêt quantitatif de cette association pour expliquer la thermosensibilité des départs HTA, indicateur important de la consommation électrique, notamment pour le dimensionnement du réseau de distribution.

Dans un second temps, les données nécessaires à la modélisation ayant été rassemblées et associées, un simulateur de consommation électrique bottom-up adapté à ces données a été conçu et implémenté (chapitre 4). Les modèles implémentés dans le simulateur sont le fruit d'un compromis entre complexité, amenant de la précision de simulation, et simplicité, assurant des entrées renseignables de façon réaliste avec les données disponibles tout en minimisant le temps de calcul. Cela a permis au simulateur de répondre au cahier des charges d'Enedis en terme d'outil d'aide à la planification, tout en étant adapté à des études plus larges propres au cadre de la recherche, telles que la simulation des alternatives énergétiques du futur.

Une fois élaboré, le simulateur a été validé et calibré (chapitre 5), et trois critères ont été étudiés. Le premier critère est la correspondance entre les énergies annuelles simulées et les valeurs de référence nationales, qui constituent la consommation cible. Le simulateur est calibré pour atteindre cette consommation cible. Le second critère est l'adéquation entre les profils simulés et mesurés. L'erreur entre les profils simulés et les mesures de départs HTA d'Enedis n'est pas facilement interprétable, de par le grand nombre et la variété des clients participant à la consommation mesurée. Pour pallier ces difficultés, des profils optimaux en termes d'erreurs de simulation des mesures HTA ont été extraits grâce à un modèle statistique, servant de référence pour les erreurs de simulation. La comparaison entre les profils simulés et les mesures montre des évolutions similaires avec la comparaison entre les profils optimaux et les mesures. Les tendances d'erreurs similaires entre la référence et le simulateur confortent les résultats obtenus avec le simulateur de charge concernant les profils. Le troisième critère est la comparaison entre la partie thermosensible de la consommation électrique mesurée et la partie simulée. C'est un nouveau critère d'évaluation des simulateurs bottom-up qui est rendu possible grâce aux mesures d'Enedis. Il garantit la qualité de la simulation de la consommation en termes de dépendance à la température, paramètre essentiel au dimensionnement du réseau de distribution. Nous avons ainsi proposé une méthode de calibration du modèle de chauffage électrique sur la partie thermosensible de la courbe de charge mesurée. Cette méthode permet de calculer des facteurs correctifs par départ HTA, qui sont ensuite utilisés dans la démarche de simulations prospectives.

Finalement, une démarche de simulations prospectives pour l'aide à la planification en utilisant les méthodes développées pendant la thèse a été proposée (chapitre 6). Après avoir étudié le contexte et les outils traditionnels utilisés par Enedis pour dimensionner son réseau, nous avons souligné l'intérêt d'utiliser le simulateur de courbes de charge comme complément à ces méthodes. Une démarche de simulations prospectives a été mise en place pour aider au dimensionnement du réseau de distribution, et testée sur deux cas d'études. Le premier cas d'étude a été mené sur 10 départs HTA de la ville de Brest et a exploré des nouvelles possibilités offertes par la démarche de simulations prospectives, pour aider au dimensionnement du réseau de distribution, par rapport aux outils traditionnels d'Enedis. Le second cas d'étude a étudié l'impact d'un fort taux de pénétration d'énergies renouvelables sur un critère de dimensionnement du réseau HTA d'Enedis, appliqué à deux départs HTA de la région de Lyon. Les résultats montrent qu'en ne considérant que ce critère, un fort taux d'énergies renouvelables peut être intégré sur le réseau de distribution sans renforcement de celui-ci. Ce travail ne tient cependant pas compte de nombreux autres critères comme le problème de déséquilibre des phases. Il a pour vocation de servir de base aux échanges, pour intégrer par la suite d'autres critères de dimensionnement et de nouvelles scénarisations.

## 7.2 Limites des travaux et perspectives notables

Le domaine de la modélisation de la consommation électrique est un sujet vaste, comme nous l'avons vu au chapitre 2. Quelques jalons supplémentaires pour la modélisation bottom-up de la consommation électrique française pour simuler des scénarios prospectifs ont été posés et présentés dans ce manuscrit. Les outils et les méthodes développés sont utilisables actuellement et en constante évolution, offrant des perspectives prometteuses pour le futur de ce projet.

Une première limite de cette thèse porte sur l'utilisation de la base des clients et des mesures d'Enedis associée à la base des logements de l'INSEE comme données d'entrée pour la simulation. L'utilisation de bases de données supplémentaires, telle le répertoire SIRENE, la base des impôts fonciers MAJIC, est une source d'amélioration importante. La deuxième limite est présente dans les modèles du simulateur de courbes de charge. Ce dernier implémente les modèles les plus simples de la littérature existante des modèles bottom-up. La version actuelle étant opérationnelle, elle peut servir de première référence. Toute proposition d'amélioration pourra alors être analysée par rapport à la référence actuelle, avec deux critères : temps de calcul pour les simulations et précision avec les résultats de l'évaluation sur les mesures de départs HTA.

Concernant l'évaluation et la calibration, celles-ci ont été menées de façon itérative et indépendante, d'abord sur les consommations en énergie, puis sur les parties non thermosensible et thermosensible des départs HTA. Une procédure plus unifiée de calibration prenant en compte tous les paramètres de façon globale améliorerait les résultats du processus de calibration sur les départs HTA.

Enfin, la démarche de simulations prospectives et les cas d'étude présentés ont montré que le simulateur de courbes de charge s'inscrit de façon satisfaisante par rapport aux outils traditionnels d'Enedis. Cependant, pour vérifier ses zones de validité, le simulateur nécessite d'être utilisé dans des cas plus nombreux. Dans ces cas, les résultats de simulation doivent être comparés, via les indicateurs utilisés pour le dimensionnement, avec les outils traditionnels d'Enedis.

D'autres applications plus orientées vers les alternatives énergétiques du futur sont à aussi développer. Il s'agit, par exemple, d'une étude sur l'impact que l'autoconsommation locale peut avoir au niveau des courbes de charge agrégées. Une telle étude utilise à la fois simulateur de courbes de charge, de production renouvelable, et inclut aussi une modélisation des moyens de stockage. Le simulateur de courbes de charge peut aussi être un module d'un modèle plus large. Notamment, il peut être utile pour la simulation des flux de puissance dans le réseau de distribution en prenant en compte les problématiques de tension, de congestion, d'arbitrage des prix de l'électricité et de gestion de la demande.



# Bibliographie

- [1] CEREN. <https://www.ceren.fr/activites/> consulté le 20/08/2017.
- [2] Collectivites-locales.gouv. <http://www.collectivites-locales.gouv.fr/fourniture-fichiers-informatises-fonciers> consulté le 01/07/2017.
- [3] Commission de régulation de l'énergie. <http://www.cre.fr/reseaux/reseaux-publics-d-electricite/qualite-de-l-electricite> consulté le 10/10/2017.
- [4] Copernicus. <http://clim4energy.climate.copernicus.eu/> consulté le 11/10/2017.
- [5] Décret n 2016-711 du 30 mai 2016 relatif aux travaux d'isolation en cas de travaux de ravalement de façade, de réfection de toiture ou d'aménagement de locaux en vue de les rendre habitables. <https://www.legifrance.gouv.fr/eli/decret/2016/5/30/LHAX1613394D/jo>. Consulté le 27/09/2017.
- [6] DRIAS. <http://www.drias-climat.fr/> consulté le 11/10/2017.
- [7] EDF. <https://travaux.edf.fr/electricite/raccordement/repartition-de-la-consommation-d-electricite-au-sein-d-un-foyer> consulté le 20/08/2017.
- [8] Enedis. <https://data.enedis.fr> consulté le 01/07/2017.
- [9] IGN. <http://professionnels.ign.fr/adressepremium> consulté le 01/07/2017.
- [10] INSEE. Enquête logement. <https://www.insee.fr/fr/statistiques/2409491?sommaire=2409559> consulté le 03/07/2017.
- [11] INSEE. Enquête sur les consommations d'énergie dans le tertiaire / ECET. <https://www.insee.fr/fr/metadonnees/source/s1215>. Consulté le 09/10/2017.
- [12] INSEE. Enquête sur les consommations d'énergie dans l'industrie / EACEI. <https://www.insee.fr/fr/metadonnees/source/s1214>. Consulté le 09/10/2017.
- [13] INSEE. <https://www.insee.fr/fr/statistiques?categorie=1> Consulté le 09/10/2017.
- [14] Ministère de la cohésion des territoires. <http://www.cohesion-territoires.gouv.fr/la-reglementation-thermique-2012> consulté le 27/09/2017.
- [15] Ministère de la transition écologique et solidaire. <https://www.ecologique-solidaire.gouv.fr/developpement-des-vehicules-propres> consulté le 27/09/2017.
- [16] Ministère de l'économie, des Finances, de l'Action et des Comptes publics. <https://www.economie.gouv.fr/particuliers/aides-renovation-energetique>. Consulté le 27/09/2017.
- [17] Modern Era-Retrospective analysis for Research and Applications, from NASA, <http://gmao.gsfc.nasa.gov>, consulté le 28/09/2017.
- [18] OIE. Solar Geometry. <http://www.oie.mines-paristech.fr/Valorisation/Outils/Solar-Geometry/> consulté le 03/10/2017.

- [19] RTE. <http://www.rte-france.com/fr/eco2mix/eco2mix-consommation> consulté le 25/09/2017.
- [20] SIRENE. <https://www.sirene.fr/sirene/public/static/contenu-base-sirene> consulté le 08/07/2017.
- [21] Référenciel foncier public national. guide d'utilisation. Technical report, Délégation à l'action foncière et immobilière; CETE Méditerranée, 2013.
- [22] Données statistiques du ceren. Technical report, CEREN, 08 2015.
- [23] Réglementation Thermique 2005. Règles th-i. caractérisation de l'inertie thermique des bâtiments. Technical report, 2006.
- [24] Réglementation Thermique 2005. Th-ce 2005. Technical report, 2006.
- [25] ADEME. Chiffres clés du bâtiment. Technical report, ADEME, 2013.
- [26] ADEME, EDF, and ENERTECH. Campagne de mesures de la consommation électrique de 20 piscines individuelles. rapport final. Technical report, 2007.
- [27] ADEME, EDF, and ENERTECH. Campagne de mesures des appareils de production de froid et des appareils de lavage dans 100 logements. Technical report, Projet AEE 2008, 2008.
- [28] ADEME, EDF, Union Européenne, and ENERTECH. Mesure de la consommation des usages domestiques de l'audiovisuel et de l'informatique. Technical report, Projet REMODECE, 2008.
- [29] ADEME and Enertech. Diagnostic électrique d'un supermarché de moyenne surfaces. rapport final. Technical report, 2001.
- [30] Ademe and Enertech. Technologies de l'information et éclairage. campagne de mesures dans 49 ensembles de bureaux de la région paca. Technical report, 2005.
- [31] ADEME and Enertech. Ineed. évaluation des performances énergétiques. rapport final. Technical report, 2010.
- [32] ADEME and ENERTECH. Campagne de mesures des appareils de production de froid, des appareils de lavage et de la climatisation. Technical report, Juin 2016.
- [33] ADEME and Cabinet Olivier Sidler. Campagne de mesures sur les usages électriques dans le secteur résidentiel en guyane. Technical report, 1998.
- [34] F.M. Andersen, H.V. Larsen, and T.K. Boomsma. Long-term forecasting of hourly electricity load : Identification of consumption profiles and segmentation of customers. *Energy Conversion and Management*, 68 :244 – 252, 2013.
- [35] F.M. Andersen, H.V. Larsen, and R.B. Gaardestrup. Long term forecasting of hourly electricity consumption in local areas in Denmark. *Applied Energy*, 110 :147 – 162, 2013.
- [36] F.M. Andersen, H.V. Larsen, N. Juul, and R.B. Gaardestrup. Differentiated long term projections of the hourly electricity consumption in local areas. the case of Denmark West. *Applied Energy*, 135 :523 – 538, 2014.
- [37] Marianne M. Armstrong, Mike C. Swinton, Hajo Ribberink, Ian Beausoleil-Morrison, and Jocelyn Millette. Synthetically derived profiles for representing occupant-driven electric loads in canadian housing. *Journal of Building Performance Simulation*, 2(1) :15–30, 2009.
- [38] T. Barbier, R. Girard, F. P. Neirac, N. Kong, and G. Kariniotakis. A novel approach for electric load curve holistic modelling and simulation. In *MedPower 2014*, pages 1–8, Nov 2014.

- [39] Thibaut Barbier, Robin Girard, Nicolas Kong, François-Pascal Neirac, Georges Kariniotakis, and Elena Magliaro. An optimisation algorithm for matching large scale databases on customers for improved characterisation of electricity consumption. In *MedPower 2016 - The 10th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion*, Belgrade, Serbia, November 2016.
- [40] Thibaut Barbier, Robin Girard, and François-Pascal Neirac. Development of an electric load curve simulator using a holistic modeling approach. Technical report, Mines ParisTech, 2014.
- [41] Robert Bartels, Denzil G. Fiebig, Michael Garben, and Robert Lumsdaine. An end-use electricity load simulation model. *Utilities Policy*, 2(1) :71 – 82, 1992.
- [42] Clément Baudot, Guillaume Roupioz, and Aymeric Billet. Modernizing distribution network management with linky smart meters - lessons learned in greenlys project. In *23rd International Conference on Electricity Distribution*, 2015.
- [43] Thomas Berthou. *Développement de modèles de bâtiments pour la prévision de charge de climatisation et l'élaboration de stratégies d'optimisation énergétique et d'effacement*. PhD thesis, 2013. Thèse de doctorat dirigée par Marchio, Dominique Energétique et Procédés Paris, ENMP 2013.
- [44] Thomas Berthou, Bruno Duplessis, Philippe Rivière, Pascal Stabat, Damien Casetta, and Dominique Marchio. SMART-E : a tool for energy demand simulation and optimization at the city scale. In *14th International Conference of the IBPSA*, Hyderabad, India, December 2015.
- [45] Thomas Berthou, Pascal Stabat, Raphael Salvazet, and Dominique Marchio. Development and validation of a gray box model to predict thermal behavior of occupied office buildings. *Energy and Buildings*, 74(0) :91 – 100, 2014.
- [46] Paolo Bertoldi and Bogdan Atanasiu. An in-depth analysis of the electricity end-use consumption and energy efficiency trends in the tertiary sector of the european union. *International Journal of Green Energy*, 8(3) :306–331, 2011.
- [47] Ricardo J. Bessa, Corinna Möhrlen, Vanessa Fundel, Malte Siefert, Jethro Browell, Sebastian Haglund El Gaidi, Bri-Mathias Hodge, Umit Cali, and Georges Kariniotakis. Towards Improved Understanding of the Applicability of Uncertainty Forecasts in the Electric Power Industry. *Energies*, 10(9) :1402, 2017.
- [48] Philippe Blanc, Benot Gschwind, Mireille Lefvre, and Lucien Wald. The helioclim project : Surface solar irradiance data for climate applications. *Remote Sensing*, 3(12) :343361, Feb 2011.
- [49] Arthur Bossavy, Robin Girard, and Georges Kariniotakis. Sensitivity analysis in the technical potential assessment of onshore wind and ground solar photovoltaic power resources at regional scale. *Applied Energy*, 182(Supplement C) :145 – 153, 2016.
- [50] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi. A bottom-up approach to residential load modeling. *IEEE Transactions on Power Systems*, 1994.
- [51] Philippe CARRIVE. Réseaux de distribution structure et planification. *Techniques de l'ingénieur Réseaux électriques de distribution publique*, base documentaire : TIB264DUO.(ref. article : d4210), 1991. fre.
- [52] CEE, EDF, ADEME, and cabinet Olivier Sidler. Maîtrise de la demande d'électricité. étude expérimentale des appareils de cuisson, de froid ménager et de séchage dans 100 logements. Technical report, Projet ECUEL., Janvier 1999.

- [53] J.-B. Champion, C. Collin, C. Lesdos Cauhapé, V. Quénechdu, and H. Valdelièvre. Des ménages toujours plus nombreux, toujours plus petits. *INSEE Première*, 2017.
- [54] D. P. Chassin, K. Schneider, and C. Gerkenmeyer. Gridlab-d : An open-source power systems modeling and simulation environment. In *2008 IEEE/PES Transmission and Distribution Conference and Exposition*, pages 1–5, April 2008.
- [55] David P Chassin, Jason C Fuller, and Ned Djilali. Gridlab-d : An agent-based simulation framework for smart grids. *Journal of Applied Mathematics*, 2014.
- [56] C.S Chen, M.S Kang, J.C Hwang, and C.W Huang. Synthesis of power system load profiles by class load study. *International Journal of Electrical Power & Energy Systems*, 22(5) :325 – 330, 2000.
- [57] G. Chicco, R. Napoli, and F. Piglione. Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, 21(2) :933–940, May 2006.
- [58] PAC & Clim’info. La climatisation. les pompes à chaleur. les chiffres du marché français de janvier à août 2016. Technical report, 2016.
- [59] European Commission. *Benchmarking smart metering deployment in the EU-27 with a focus on electricity*. European Commission, Brussels, June 2014.
- [60] Agence Internationale de l’Energie. Technology roadmap. solar photovoltaic energy. pages 9–37, 2014.
- [61] Vincent Debusschere, Kaustav Basu, and Seddik Bacha. Identification et prédiction non intrusive de l’état des charges dans les bâtiments résidentiels à partir de mesures compteur à échantillonnage réduit. In *Symposium de Génie Électrique 2014*, Cachan, France, July 2014.
- [62] J. Dickert and P. Schegner. Residential load models for network planning purposes. In *2010 Modern Electric Power Systems*, pages 1–6, Sept 2010.
- [63] J. Dickert and P. Schegner. A time series probabilistic synthetic load curve model for residential customers. In *2011 IEEE Trondheim PowerTech*, pages 1–6, June 2011.
- [64] Ni Ding. *Load models for operation and planning of electricity distribution networks with metering data*. Theses, Université de Grenoble, November 2012.
- [65] Alain Doulet and Pierre Bornard. Réseaux électriques avant-propos. *Techniques de l’ingénieur Généralités sur les réseaux électriques*, base documentaire : TIB261DUO.(ref. article : d4000), 2013. fr.
- [66] EDF, ADEME, and ENERTECH. Campagne de mesures de l’éclairage dans 100 logements en france. Technical report, Mars 2004.
- [67] Mabrouka El Guedri. *Caractérisation aveugle de la courbe de charge électrique : Détection, classification et estimation des usages dans les secteurs résidentiel et tertiaire*. Theses, Université Paris Sud - Paris XI, November 2009.
- [68] Enertech. Etude par mesure de l’ensemble des consommations d’énergie d’un hôtel. Technical report, 2001.
- [69] Enertech. Consommations d’électricité par usage dans un hôtel 3\* parisien. Technical report, 2003.
- [70] Enertech. Review of all existing european monitoring campaigns in households. Technical report, REMODECE, 2006.
- [71] ERDF and RTE. Règles relatives à la programmation, au mécanisme d’ajustement et au dispositif de responsable d’équilibre. chapitre f. relations entre le grd et le responsable d’équilibre. dispositions applicables pour l’estimation des courbes de charges. Technical report, 08 2015.

- [72] Explicit, Arene Ile de France, and ADEME. Enquête sur les consommations d'énergie des commerces et services de proximité en région ile-de-france. Technical report, Janvier 2008.
- [73] Aurélie Ferrage, Maxence Bocquel, Nicolas Kong, Thibaut Barbier, and Robin Girard. Using simulated predictive load curves to improve DSO's network development planning methods integrating smart grids functionalities. In *23rd International Conference on Electricity Distribution - CIRED 2015*, Lyon, France, June 2015.
- [74] Pierre Ferrand. Comptage d'électricité fondements de la tarification. *Techniques de l'ingénieur Problématiques communes des réseaux électriques : du fonctionnement au comptage*, base documentaire : TIB266DUO.(ref. article : d4951), 2001. fre.
- [75] Marisa B. Figueiredo, Ana de Almeida, and Bernardete Ribeiro. *An Experimental Study on Electrical Signature Identification of Non-Intrusive Load Monitoring (NILM) Systems*, pages 31–40. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [76] David Fischer, Andreas Härtl, and Bernhard Wille-Haussmann. Model for electric load profiles with high time resolution for german households. *Energy and Buildings*, 92 :170 – 179, 2015.
- [77] David Fischer and Hatef Madani. On heat pumps in smart grids : A review. *Renewable and Sustainable Energy Reviews*, 70(Supplement C) :342 – 357, 2017.
- [78] Jimeno A. Fonseca and Arno Schlueter. Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts. *Applied Energy*, 142 :247 – 265, 2015.
- [79] Carine Burricand François Gleizes. De plus en plus de foyers équipés de biens électroniques. Technical report, INSEE, 2015.
- [80] Krzysztof Gajowniczek and Tomasz Zabkowski. Short term electricity forecasting using individual smart meter data. *Procedia Computer Science*, 35 :589 – 597, 2014.
- [81] Alexis Gerossier, Thibaut Barbier, and Robin Girard. A novel method for decomposing electricity feeder load into elementary profiles from customer information. *Applied Energy*, 203 :752 – 760, 2017.
- [82] Luc Girardin, François Marechal, Matthias Dubuis, Nicole Calame-Darbellay, and Daniel Favrat. Energis : A geographical information based system for the evaluation of integrated energy conversion systems in urban areas. *Energy*, 35(2) :830 – 840, 2010. ECOS 2008.
- [83] Conseil général du Bas-Rhin, Ademe, and Enertech. Diagnostic énergétique de l'hôtel du département du bas-rhin. Technical report, 2005.
- [84] Nicholas Good, Lingxi Zhang, Alejandro Navarro-Espinosa, and Pierluigi Mancarella. High resolution modelling of multi-energy domestic demand profiles. *Applied Energy*, 137 :193 – 210, 2015.
- [85] Yannig Goude, Raphael Nedellec, and Nicolas Kong. Local short and middle term electricity load forecasting with semi-parametric additive models. *IEEE transactions on smart grid*, 5(1) :440–446, 2014.
- [86] A. Grandjean, J. Adnot, and G. Binet. A review and an analysis of the residential electric load curve models. *Renewable and Sustainable Energy Reviews*, 16(9) :6539 – 6565, 2012.
- [87] Arnaud Grandjean. *Introduction de non linéarités et de non stationnarités dans les modèles de représentation de la demande électrique résidentielle*. PhD thesis, 2013. Thèse de doctorat dirigée par Adnot, Jérôme Energétique Paris, ENMP 2013.
- [88] Bjorn Grinden and Nicolai Feilberg. Analysis of monitoring campaign in europe. Technical report, REMODECE project., Nov 2008.

- [89] E. Grover-Silva, R. Girard, and G. Kariniotakis. Multi-temporal optimal power flow for assessing the renewable generation hosting capacity of an active distribution system. In *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T D)*, pages 1–5, May 2016.
- [90] E. Grover-Silva, R. Girard, and G. Kariniotakis. Optimal sizing and placement of distribution grid connected battery systems through an socp optimal power flow algorithm. *Applied Energy*, 2017.
- [91] R. F. Hamilton. Synthetic or equivalent load curves. *Electrical Engineering*, 61(6) :369–381, June 1942.
- [92] Y. Han, X. Sha, E. Grover-Silva, and P. Michiardi. On the impact of socio-economic factors on power load forecasting. In *Big Data (Big Data), 2014 IEEE International Conference on*, Oct 2014.
- [93] Pecan Street Inc. *Dataport from Pecan Street. [Online]. Available : <https://dataport.cloud/>*, 2017.
- [94] Rory V. Jones, Alba Fuertes, and Kevin J. Lomas. The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. *Renewable and Sustainable Energy Reviews*, 43 :901 – 917, 2015.
- [95] Georges Kariniotakis, editor. *Renewable Energy Forecasting*. Woodhead Publishing, 2017.
- [96] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic. A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment*, 45(7) :1683 – 1697, 2010.
- [97] James Keirstead, Nouri Samsatli, and Nilay Shah. Syncity : an integrated tool kit for urban energy systems modelling. In *Fifth Urban Research Symposium*, 2009.
- [98] Nicolas Kong, Maxence Bocquel, Thibaut Barbier, Pelton Guillaume, Cauchois Pierre, Robin Girard, Elena Magliaro, and Georges Kariniotakis. Long term forecast of local electrical demand and evaluation of future impacts on the electricity distribution network. 2017.
- [99] M. I. Chan, E. N. Marsh, J. Y. Yoon, and G. B. Ackerman. Simulation-based load synthesis methodology for evaluating load-management programs. *IEEE Transactions on Power Apparatus and Systems*, PAS-100(4) :1771–1778, April 1981.
- [100] Pacific Northwest National Laboratory. Gridlab-d power system simulation software. Technical report, <http://www.gridlabd.org/>, 2017.
- [101] Kai le Zhou, Shan lin Yang, and Chao Shen. A review of electric load classification in smart grid environment. *Renewable and Sustainable Energy Reviews*, 24 :103 – 110, 2013.
- [102] Vincent Lefieux. *Modèles semi-paramétriques appliqués à la prévision des séries temporelles. Cas de la consommation d’électricité*. PhD thesis, Université Rennes 2, 2007.
- [103] Laurent Letinois. Methodologie de repartition spatiale de la population. Technical report, Laboratoire Central de Surveillance de la Qualité de l’Air, 2014.
- [104] Xing Luo, Jihong Wang, Mark Dooner, and Jonathan Clarke. Overview of current development in electrical energy storage technologies and the application potential in power system operation. *Applied Energy*, 137(Supplement C) :511 – 536, 2015.
- [105] Bruno Maresca, Anne Dujin, and Romain Picard. La consommation d’énergie dans l’habitat. entre recherche de confort et impératif écologique. Technical report, CREDOC, 2009.

- [106] Omid Mazdiyasn, Amir AghaKouchak, Steven J. Davis, Shahrbanou Madadgar, Ali Mehran, Elisa Ragno, Mojtaba Sadegh, Ashmita Sengupta, Subimal Ghosh, C. T. Dhanya, and Mohsen Niknejad. Increasing probability of mortality during indian heat waves. *Science Advances*, 3(6), 2017.
- [107] Eoghan McKenna, Ian Richardson, and Murray Thomson. Smart meter data : Balancing consumer privacy concerns with legitimate applications. *Energy Policy*, 41 :807–814, 2012.
- [108] Fintan McLoughlin, Aidan Duffy, and Michael Conlon. A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, 141 :190 – 199, 2015.
- [109] A. Mutanen, M. Ruska, S. Repo, and P. Jarventausta. Customer classification and load profiling method for distribution systems. *IEEE Transactions on Power Delivery*, 26(3) :1755–1763, July 2011.
- [110] Intergovernmental Panel on Climate Change. Mitigation of climate change. In *Climate Change 2014*, chapter Summary for Policymakers, pages 1–33. Cambridge University Press, Cambridge, 2015.
- [111] G Ouzeau, M Déqué, M Jouini, S Planton, and R Vautard. Le climat de la france au xxie siècle. *Rapport de la Direction générale de l'énergie et du climat, www. developpement-durable. gov. fr*, 2014.
- [112] Jukka V. Paatero and Peter D. Lund. A model for generating household electricity load profiles. *International Journal of Energy Research*, 30(5) :273–290, 2006.
- [113] Région PACA, ADEME, and Enertech. Etude des paramètres influant sur les consommations de climatisation dans les immeubles de bureaux. Technical report, 2002.
- [114] Région PACA and Enertech. Technologies de l'information et éclairage. enquêtes de terrain dans 50 bâtiments de bureaux. Technical report, 2004.
- [115] Jessen Page, Daniele Basciotti, Olivier Pol, José Nuno Fidalgo, Mario Couto, Rebecca Aron, Alice Chiche, and Laurent Fournié. A multi-energy modelling, simulation, and optimization environment for urban energy infrastructure planning. In *Conference of International Building Performance Simulation Association*, 2013.
- [116] T. Pons. La compensation des pertes à erdf. présentation aux membres du groupe de travail cre sur les pertes. Technical report, ERDF, 2009.
- [117] ERDF Direction Raccordement. Référentiel documentaire pour l'offre de raccordement et l'accès au réseau des installations des utilisateurs individuels du réseau public de distribution. Technical report, 2007.
- [118] Teemu Rasanen, Dimitrios Voukantsis, Harri Niska, Kostas Karatzas, and Mikko Kolehmainen. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Applied Energy*, 87(11) :3538 – 3545, 2010.
- [119] Christoph F. Reinhart and Carlos Cerezo Davila. Urban building energy modeling – a review of a nascent field. *Building and Environment*, 97 :196 – 202, 2016.
- [120] Christoph F Reinhart, Timur Dogan, J Alstan Jakubiec, Tarek Rakha, and Andrew Sang. Umi - an urban simulation environment for building energy use, daylighting and walkability. In *13th Conference of International Building Performance Simulation Association, Chambéry, France, August 26-28*, 2013.

- [121] Joshua D. Rhodes, Wesley J. Cole, Charles R. Upshaw, Thomas F. Edgar, and Michael E. Webber. Clustering analysis of residential electricity demand profiles. *Applied Energy*, 135 :461 – 471, 2014.
- [122] Ian Richardson, Murray Thomson, David Infield, and Conor Clifford. Domestic electricity use : A high-resolution energy demand model. *Energy and Buildings*, 42(10) :1878 – 1887, 2010.
- [123] Darren Robinson, Haldi F., Kämpf J., Leroux P., Perez D., Rasheed A., and Wilke U. Citysim : comprehensive micro-simulation of resource flows for sustainable urban planning. 2009.
- [124] Antoine Rogeau, Thibaut Barbier, Robin Girard, and Nicolas Kong. Evolution of electrical distribution grids sizing considering self-consumption of local renewable production. 2017.
- [125] Région Languedoc Roussillon and Enertech. Etude des consommations électriques de la cuisine du lycée jules guesde (montpellier). Technical report, 2009.
- [126] RTE. Consommation française d’électricité : Caractéristiques et méthode de prévision. Technical report, RTE, Novembre 2014.
- [127] RTE. Bilan prévisionnel de l’équilibre offre-demande d’électricité en france. Technical report, RTE, 2016.
- [128] RTE, Syndicat des Energies Renouvelables, Enedis, and ADEeF. Panorama de l’électricité renouvelable en 2016. Technical report, 2017.
- [129] K. P. Schneider, J. C. Fuller, and D. P. Chassin. Multi-state load models for distribution system analysis. *IEEE Transactions on Power Systems*, 26(4) :2425–2433, Nov 2011.
- [130] K. P. Schneider and T. F. Weaver. Volt-var optimization on american electric power feeders in northeast columbus. In *PES T D 2012*, pages 1–8, May 2012.
- [131] Anssi Seppala. *Load research and load estimation in electricity distribution*. Theses, Technical research center of Finland, VTT Publications, Jan 1996.
- [132] European Parliamentary Research Service. *Smart electricity grids and meters in the EU Member States*. European Parliament, Brussels, September 2015.
- [133] R. Sevljan and R. Rajagopal. Value of aggregation in smart grids. In *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 714–719, Oct 2013.
- [134] R. A. Sevljan and R. Rajagopal. A model for the effect of aggregation on short term load forecasting. In *2014 IEEE PES General Meeting | Conference Exposition*, pages 1–5, July 2014.
- [135] Zhen Shao, Fu Chao, Shan-Lin Yang, and Kai-Le Zhou. A review of the decomposition methodology for extracting and identifying the fluctuation characteristics in electricity demand forecasting. *Renewable and Sustainable Energy Reviews*, pages –, 2016.
- [136] Zhen Shao, Fei Gao, Qiang Zhang, and Shan-Lin Yang. Multivariate statistical and similarity measure based semiparametric modeling of the probability distribution : A novel approach to the case study of mid-long term electricity consumption forecasting in China. *Applied Energy*, 156 :502 – 518, 2015.
- [137] Yoshiyuki Shimoda, Takuro Fujii, Takao Morikawa, and Minoru Mizuno. Evaluation on residential energy efficiency programs using the city-scale end-use simulation model. In *Summer Study on Energy Efficiency in Buildings*, 2004.
- [138] LD Shorrock and JE Dunster. The physically-based model brehomes and its use in deriving scenarios for the energy use and carbon dioxide emissions of the uk housing stock. *Energy Policy*, 25(12) :1027 – 1037, 1997.

- [139] Cabinet Conseil Olivier Sidler. Maîtrise de la demande électrique. campagne de mesures par usage dans le secteur domestique. Technical report, Programme SAVE, Janvier 1995.
- [140] Olivier Sidler. Connaissance et maîtrise des usages spécifiques de l'électricité dans le secteur résidentiel. Technical report, Enertech, 2009.
- [141] Melody Stokes. *Removing barriers to embedded generation : a fine-grained load model to support low voltage network performance analysis*. PhD thesis, 2005.
- [142] Melody Stokes, Mark Rylatt, and Kevin Lomas. A simple model of domestic lighting demand. *Energy and Buildings*, 36(2) :103 – 116, 2004.
- [143] Lukas G. Swan and V. Ismet Ugursal. Modeling of end-use energy consumption in the residential sector : A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8) :1819 – 1835, 2009.
- [144] Muneer T. *Solar radiation and daylight models for the energy efficient design of buildings*. Architectural Press, 1997.
- [145] Enedis Direction Technique. Modèle de convention de raccordement au réseau public de distribution hta d'une installation de production d'énergie électrique relevant d'un schéma régional de raccordement au réseau des énergies renouvelables (srrrer). Technical report, 2016.
- [146] Enedis Direction Technique. Description physique du réseau public de distribution. Technical report, 2017.
- [147] Hélène Thélot. Les achats d'énergie dans le secteur tertiaire marchand en 2011. Technical report, INSEE, 2014.
- [148] Kenneth E. Train. An assessment of the accuracy of statistically adjusted engineering (sae) models of end-use load curves. *Energy*, 17(7) :713 – 723, 1992.
- [149] Brice Tremeac, Pierre Bousquet, Cecile de Munck, Gregoire Pigeon, Valery Masson, Colette Marchadier, Michele Merchat, Pierre Poeuf, and Francis Meunier. Influence of air conditioning management on heat island in paris air street temperatures. *Applied Energy*, 95(Supplement C) :102 – 110, 2012.
- [150] Joaquim L. Viegas, Susana M. Vieira, R. Melcio, V.M.F. Mendes, and Joo M.C. Sousa. Classification of new electricity customers based on surveys and smart metering data. *Energy*, 107 :804 – 817, 2016.
- [151] Éric Vorger. *Study of the influence of the inhabitants behavior on the energy performance of buildings*. Theses, Ecole Nationale Supérieure des Mines de Paris, December 2014.
- [152] C. F. Walker and J. L. Pokoski. Residential load shape modelling based on customer behavior. *IEEE Transactions on Power Apparatus and Systems*, PAS-104(7) :1703–1711, July 1985.
- [153] Yong Wang and Lin Li. Time-of-use electricity pricing for industrial customers : A survey of u.s. utilities. *Applied Energy*, 149 :89 – 103, 2015.
- [154] Joakim Widén, Magdalena Lundh, Iana Vassileva, Erik Dahlquist, Kajsa Ellegård, and Ewa Wäckelgård. Constructing load profiles for household electricity and hot water from time-use data—modelling approach and validation. *Energy and Buildings*, 41(7) :753 – 768, 2009.
- [155] Joakim Widén and Ewa Wäckelgård. A high-resolution stochastic model of domestic activity patterns and electricity demand. *Applied Energy*, 87(6) :1880 – 1892, 2010.
- [156] Runming Yao and Koen Steemers. A method of formulating energy load profile for domestic buildings in the uk. *Energy and Buildings*, 37(6) :663 – 671, 2005.

- [157] Abdelouadoud Seddik Yassine. *Intégration des énergies renouvelables au réseau de distribution d'électricité*. PhD thesis, 2014.
- [158] Jean Paul Zimmermann. End-use metering campaign in 400 households in sweden. assessment of the potential electricity savings. Technical report, Swedish Energy Agency, Enertech, 2009.

# Chapitre 8

## Annexes

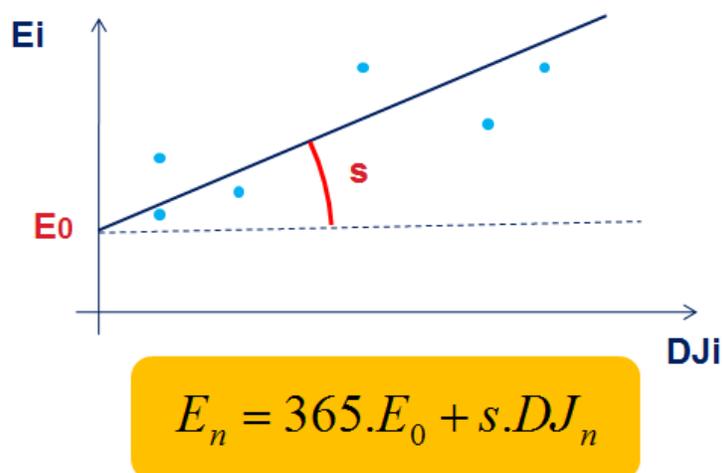
### 8.1 Annexes du chapitre 3 sur les bases de données

#### 8.1.1 Données Enedis

Information	Descriptif	Exemple
Code GDO du départ HTA	Identifiant Enedis du départ HTA sur lequel le client est connecté	ABERSC0001
Code GDO du transformateur HTA/BT	Identifiant Enedis du poste HTA/BT sur lequel le client est connecté	FTR60183313
Nom de la rue	Nom de la rue du lieu du poste de livraison du client ( <i>i.e.</i> du compteur)	Chemin du pommier
Complément d'adresse	complément d'adresse de la rue du lieu du poste de livraison du client ( <i>i.e.</i> du compteur)	bis
Code INSEE	Code INSEE de la commune	29019
Puissance souscrite	Puissance souscrite en kVA	3
Puissance de raccordement	Puissance de raccordement en kVA ( $\geq$ Puissance souscrite)	15
Statut client	Si inactif (plus de contrat) ou non	Actif
Consommation annuelle HP	Consommation annuelle heure pleine à température normale (Voir figure 8.1) en heure pleines en kWh	8250

Information	Descriptif	Exemple
Consommation annuelle HC	De même que précédemment mais pour les heures creuses en kWh	1851
Consommation annuelle en base ou EJP	De même que précédemment mais pour un tarif base ou effacement jour de pointe en kWh	11323
Libellé activité client	Classification du client	bleu domestique
Libellé Option Tarifaire	Option tarifaire : base, HP/HC, EJP, autre	base
Code APE	Code de l'activité principale exercée du client	01.13Z

TABLE 8.1 – Descriptif des informations disponibles dans la base des clients

FIGURE 8.1 – Illustration de la normalisation des relevés de compteur client (Enedis).  $E_n$  est la consommation normalisée, les points bleus sont des relevés de compteurs, DJ représentent les degrés jours

## 8.1.2 Algorithme d'association des adresses postales

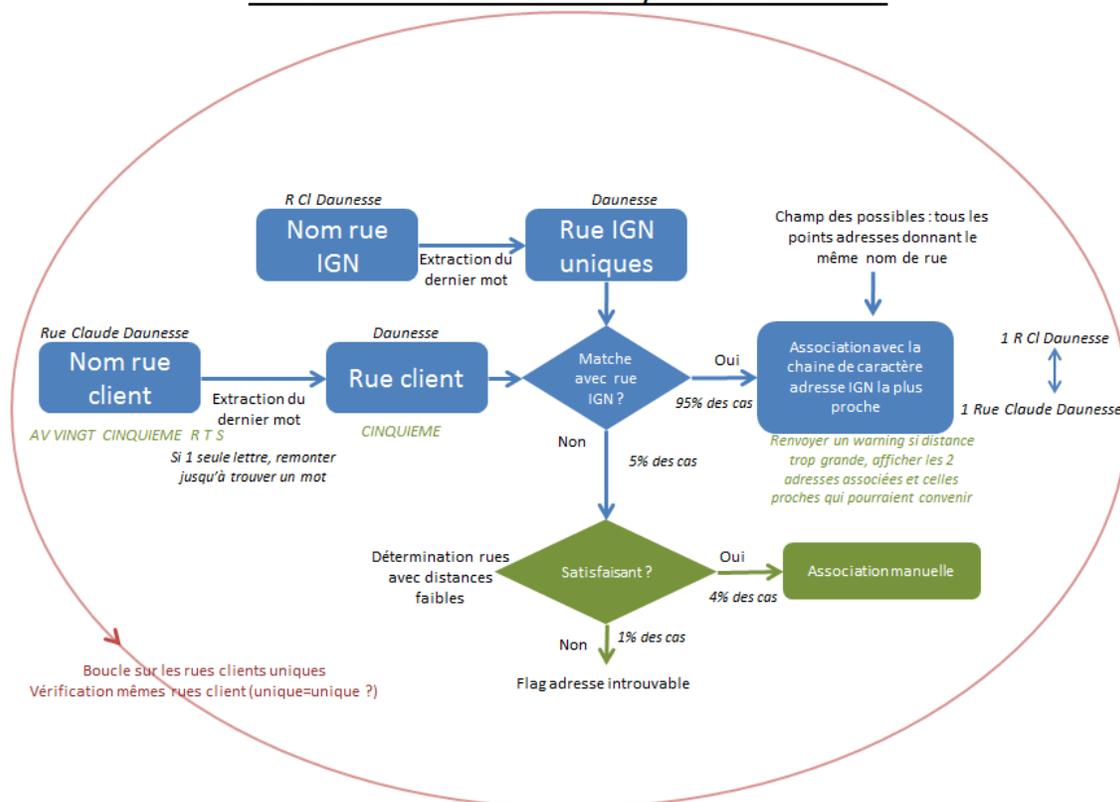
Association adresse client/adresse IGN v1

FIGURE 8.2 – Illustration de l'algorithme d'association des adresses clients et adresses IGN de la BD Adresse

Différentes associations possibles avec les drapeaux correspondant :

- 1 : Dernier mot de la rue en commun + un seul numéro correspondant (*e.g.* le 10 rue du Port avec le 10 avenue vieux Port, avec un 10 seulement dans chaque base)
- 2 : Deux mots en commun dans la rue + un seul numéro correspondant (*e.g.* le 10 rue du Port avec le 10 rue vieux Port, avec un 10 seulement dans chaque base)
- 3 : Rue exactement identique + un seul numéro correspondant (ex : le 10 rue du Port avec le 10 rue du Port, avec un 10 seulement dans chaque base)
- 4, 5 et 6 : idem que 1,2,3 mais on enlève les Bis, Ter etc
- 7 : Dernier mot de la rue en commun + Tirage au sort des numéros correspondant, car il y a plus de clients que de bâtiment, ou l'inverse (*e.g.* le 10 rue du Port avec le 10 avenue vieux Port, avec plusieurs 10 seulement dans chaque base)
- 8 : idem que 7, mais sans les Bis, Ter etc
- 9 : Dernier mot de la rue en commun + Tirage au sort des numéros différents (*e.g.* le 10,11 et 12 rue du Port avec le 8,15,20 avenue vieux Port)
- 10 : Pas de correspondance de rue trouvée

## 8.1.3 Tableau des erreurs de prévision de la consommation annuelle par IRIS

Variables du modèle	Amélioration de l'erreur (%)
NbPer (ACHLxSTOCDxCMBL)	50,37
NbPer (CATLxACHLxSTOCDxCMBL)	50,37
NbPer (ACHLxCHFLxSTOCD)	48,57
NbPer (CATLxACHLxCHFLxSTOCD)	48,57
NbPer (ACHLxTYPL)	45,95
NbPer (CATLxACHLxTYPL)	45,95
NbPer (ACHLxSTOCDxTYPL)	44,35
NbPer (CATLxACHLxSTOCDxTYPL)	44,35
NbPer (ACHLxCHFLxSTOCDxCMBL)	43,83
NbPer (CATLxACHLxCHFLxSTOCDxCMBL)	43,83
NbPer (ACHLxSTOCD)	43,74
NbPer (CATLxACHLxSTOCD)	43,74
NbPer (ACHLxTYPLxCMBL)	42,04
NbPer (CATLxACHLxTYPLxCMBL)	42,04
NbPer (ACHLxCHFLxTYPL)	41,63
NbPer (CATLxACHLxCHFLxTYPL)	41,63
NbPer (CHFLxSTOCD)	40,76
NbPer (CATLxCHFLxSTOCD)	40,76
NbPer (STOCDxCMBL)	39,72
NbPer (CATLxSTOCDxCMBL)	39,72
NbPer (CHFLxTYPL)	39,3
NbPer (CHFL)	32,61
NbPer (CMBL)	31,72
NbPer (ACHL)	23,72
NbPer (STOCD)	6,54
NbPer (TYPL)	0,73
NbPer (CATL)	0
NbPer	-23,8

TABLE 8.2 – Amélioration de l'erreur de prédiction de la consommation par IRIS en utilisant différentes variables intensives couplées au nombre de personnes par rapport au modèle de référence (surface par IRIS)

Variables du modèle	Amélioration de l'erreur (%)
NbPi (CATLxACHLxCHFLxSTOCDxCMBL)	71,34
NbPi (CATLxACHLxSTOCD)	70,57
NbPi (CATLxACHLxSTOCDxTYPLxCMBL)	70,27
NbPi (CATLxACHLxCHFLxCMBL)	69,36
NbPi (CATLxACHL)	67,64
NbPi (CATLxACHLxSTOCDxTYPL)	66,61
NbPi (CATLxACHLxSTOCDxCMBL)	66,46
NbPi (ACHLxSTOCD)	65,16
NbPi (ACHLxCHFLxSTOCDxCMBL)	64,91
NbPi (ACHLxSTOCDxTYPL)	64,78
NbPi (CATLxACHLxCHFLxSTOCD)	63,35
NbPi (CATLxACHLxCHFLxSTOCDxTYPL)	63,25
NbPi (ACHLxSTOCDxCMBL)	62,14
NbPi (ACHLxTYPLxCMBL)	61,56
NbPi (ACHLxCHFLxCMBL)	61,14
NbPi (ACHLxCHFLxSTOCD)	60,53
NbPi (ACHLxTYPL)	60,28
NbPi (ACHL)	60,27
NbPi (CATLxACHLxCMBL)	60,26
NbPi (ACHLxCHFLxTYPL)	60,06
NbPi (CMBL)	24,35
NbPi (CATL)	23,14
NbPi (CHFL)	22,01
NbPi (TYPL)	12,26
NbPi (STOCD)	3,97
NbPi	-21,5

TABLE 8.3 – Amélioration de l'erreur de prédiction de la consommation par IRIS en utilisant différentes variables intensives couplées au nombre de pièces par rapport au modèle de référence (surface par IRIS)

### 8.1.4 Tableau des erreurs de prévision de la thermosensibilité selon les différents algorithmes d'association des bases de données

Variables du modèle optim	Amélioration de l'erreur optim (%)	Variables du modèle optim sans chauffage	Amélioration de l'erreur optim sans chauffage (%)	Apport des informations non chauffage
NbPer (ACHLxSTOCD xTYPLxCMBL) Psous K	40,35	NbPer (STOCD) Psous K	37,68	2,67
NbPer (ACHLxCHFL) Psous K	40,08	NbPer (ACHLxSTOCD xTYPL) Psous K	37,08	3
NbPer (ACHLxCHFL xTYPL) Psous K	39,98	NbPer (ACHLxSTOCD) Psous K	36,98	3
NbPer (ACHLxTYP LxCMBL) Psous K	39,85	NbPer (STOCDxTYPL) Psous K	36,62	3,23
NbPer (ACHLxSTOCD) Psous K	39,59	NbPer (TYPL) Psous K	36	3,59
NbPer (ACHLxCMBL) Psous K	39,54	NbPer (ACHL) Psous K	35,64	3,9
NbPer (ACHLxCHFL xSTOCDxTYPL) Psous K	39,32	NbPer (ACHLxTYPL) Psous K	35,6	3,72
NbPer (ACHLxCHFL xCMBL) Psous K	39,23	NbPer (CATL) Psous K	-64,3	103,53
NbPer (ACHL) Psous K	39,14	NbPer (CATLxSTOCD) Psous K	-69,1	108,24
NbPer (ACHLxCHFL xTYPLxCMBL) Psous K	38,77	NbPer (CATLxACHL) Psous K	-131,3	170,07
NbPer (ACHLxCHFL xSTOCD) Psous K	38,66	NbPer (CATLxACHL xSTOCD) Psous K	-145,3	183,96
Psous K	0			

TABLE 8.4 – Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées au nombre de personnes, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode d'optimisation en utilisant ou non les données des logements avec chauffage électrique

Variables du modèle ref	Amélioration de l'erreur ref (%)	Variables du modèle ref sans chauffage	Amélioration de l'erreur ref sans chauffage (%)	Apport des informations non chauffage
NbPer (STOCDxCMBL) Psous K	29,51	NbPer (STOCD) Psous K	33,05	-3,54
NbPer (CATLxSTOCD xCMBL) Psous K	29,51	NbPer (CATLxSTOCD) Psous K	33,05	-3,54
NbPer (CHFLxSTOCD) Psous K	29,23	NbPer (TYPL) Psous K	32,26	-3,03
NbPer (CATLxCHFL xSTOCD) Psous K	29,23	NbPer (CATLxTYPL) Psous K	32,26	-3,03
NbPer (CHFLxSTOCD xCMBL) Psous K	28,86	NbPer (ACHLxSTOCD) Psous K	32,04	-3,18
NbPer (CATLxCHFL xSTOCDxCMBL) Psous K	28,86	NbPer (CATLxACHL xSTOCD) Psous K	32,04	-3,18
NbPer (ACHLxSTOCD xCMBL) Psous K	28,39	NbPer (STOCDxTYPL) Psous K	32,01	-3,62
NbPer (CATLxACHL xSTOCDxCMBL) Psous K	28,39	NbPer (CATLxSTOCD xTYPL) Psous K	32,01	-3,62
NbPer (CHFLxTYPL) Psous K	27,76	NbPer (ACHLxTYPL) Psous K	31,64	-3,88
NbPer (CATLxCHFL xTYPL) Psous K	27,76	NbPer (CATLxACHL xTYPL) Psous K	31,64	-3,88
NbPer (CHFLxTYPL xCMBL) Psous K	27,76	NbPer (CATL) Psous K	31,05	-3,29
NbPer (CATLxCHFL xTYPLxCMBL) Psous K	27,76	NbPer (ACHL) Psous K	30,31	-2,55
NbPer (ACHLxCHFL xSTOCD) Psous K	27,75	NbPer (CATLxACHL) Psous K	30,31	-2,56
NbPer (CATLxACHL xCHFLxSTOCD) Psous K	27,75	NbPer (ACHLxSTOCD xTYPL) Psous K	29,04	-1,29
NbPer (TYPLxCMBL) Psous K	27,74	NbPer (CATLxACHL xSTOCDxTYPL) Psous K	29,04	-1,3
Psous K	0			

TABLE 8.5 – Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées au nombre de personnes, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode de référence en utilisant ou non les données des logements avec chauffage électrique

Variables du modèle optim	Amélioration de l'erreur optim (%)	Variables du modèle optim sans chauffage	Amélioration de l'erreur optim sans chauffage (%)	Apport des informations non chauffage
Surf (ACHLxCHFL) Psous K	34,67	Surf (STOCDxTYPL) Psous K	33,04	1,63
Surf (ACHLxCMBL) Psous K	34,5	Surf (ACHLxSTOCD xTYPL) Psous K	33,03	1,47
Surf (ACHL) Psous K	34,17	Surf (CATLxSTOCD) Psous K	32,94	1,23
Surf (ACHLxCHFL xSTOCD) Psous K	33,94	Surf (CATLxACHL) Psous K	32,69	1,25
Surf (ACHLxCHFL xCMBL) Psous K	33,33	Surf (CATL) Psous K	31,77	1,56
Surf (ACHLxSTOCD) Psous K	33,13	Surf (CATLxACHL xSTOCD) Psous K	30,63	2,5
Surf (ACHLxSTOCD xCMBL) Psous K	31,72	Surf (STOCD) Psous K	29,48	2,24
Surf (ACHLxTYPL) Psous K	29,45	Surf (CATLxSTOCD xTYPL) Psous K	27,65	1,8
Surf (ACHLxCHFL xSTOCDxCMBL) Psous K	27,6	Surf (ACHLxSTOCD) Psous K	27,26	0,34
Surf (CATLxACHL) Psous K	27,23	Surf (CATLxTYPL) Psous K	25,43	1,8
Surf (CHFLxSTOCD) Psous K	27,05	Surf (ACHL) Psous K	21,13	5,92
Surf (STOCDxCMBL) Psous K	26,09	Surf (TYPL) Psous K	18,55	7,54
Surf (ACHLxCHFL xTYPL) Psous K	24,81	Surf (ACHLxTYPL) Psous K	18,47	6,34
Surf (CATLxACHL xSTOCD) Psous K	24,22	Surf (CATLxACHL xTYPL) Psous K	-71,4	95,62
Surf (CATLxACHL xCHFLxCMBL) Psous K	23,22	Surf (CATLxACHL xSTOCDxTYPL) Psous K	-82,1	105,32
Psous K	0			

TABLE 8.6 – Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées à la surface, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode d'optimisation en utilisant ou non les données des logements avec chauffage électrique

Variables du modèle ref	Amélioration de l'erreur ref (%)	Variables du modèle ref sans chauffage	Amélioration de l'erreur ref sans chauffage (%)	Apport des informations non chauffage
Surf (CATLxACHL xCMBL) Psous K	-16	Surf (CATLxACHL) Psous K	29,71	-45,71
Surf (CATLxACHL xSTOCDxCMBL) Psous K	-16,4	Surf (CATLxACHL xSTOCD) Psous K	29,54	-45,94
Surf (CATLxACHL xCHFL) Psous K	-17,4	Surf (CATL) Psous K	29	-46,4
Surf (CATLxACHL xTYPLxCMBL) Psous K	-18	Surf (CATLxTYPL) Psous K	28,88	-46,88
Surf (CATLxACHL xCHFLxCMBL) Psous K	-18,2	Surf (CATLxSTOCD) Psous K	28,83	-47,03
Surf (CATLxACHL xCHFLxSTOCD) Psous K	-18,4	Surf (CATLxSTOCD xTYPL) Psous K	28,15	-46,55
Surf (ACHLxCHFL xSTOCDxTYPL xCMBL) Psous K	-19,1	Surf (STOCDxTYPL) Psous K	26,75	-45,85
Surf (CATLxACHL xCHFLxTYPL) Psous K	-19,5	Surf (CATLxACHL xTYPL) Psous K	26,41	-45,91
Surf (CATLxACHL xSTOCDxTYPL xCMBL) Psous K	-19,7	Surf (ACHLxSTOCD xTYPL) Psous K	25,83	-45,53
Surf (CATLxACHL xCHFLxSTOCD xCMBL) Psous K	-21,2	Surf (ACHLxSTOCD) Psous K	25,75	-46,95
Surf (CATLxACHL) Psous K	-21,5	Surf (STOCD) Psous K	25,27	-46,77
Surf (ACHL) Psous K	-22,1	Surf (CATLxACHL xSTOCDxTYPL) Psous K	23,71	-45,81
Surf (ACHLxCHFLxSTOCD) Psous K	-22,5	Surf (TYPL) Psous K	17,32	-39,82
Surf (ACHLxSTOCD xCMBL) Psous K	-22,5	Surf (ACHL) Psous K	14,84	-37,34
Surf (ACHLxSTOCD xTYPLxCMBL) Psous K	-22,5	Surf (ACHLxTYPL) Psous K	14,27	-36,77
Psous K	0			

TABLE 8.7 – Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées à la surface, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode de référence en utilisant ou non les données des logements avec chauffage électrique

Variables du modèle optim	Amélioration de l'erreur optim (%)	Variables du modèle optim sans chauffage	Amélioration de l'erreur optim sans chauffage (%)	Apport des informations non chauffage
NbPi (ACHLxCHFL) Psous K	33,4	NbPi (ACHLxSTOCD xTYPL) Psous K	31,58	1,82
NbPi (ACHLxCMBL) Psous K	33,04	NbPi (CATLxACHL) Psous K	31,11	1,93
NbPi (ACHL) Psous K	32,45	NbPi (STOCDxTYPL) Psous K	30,77	1,68
NbPi (ACHLxCHFL xCMBL) Psous K	31,42	NbPi (CATLxSTOCD) Psous K	30,58	0,84
NbPi (ACHLxCHFL xSTOCD) Psous K	31,19	NbPi (CATLxACHL xSTOCD) Psous K	30,47	0,72
NbPi (ACHLxSTOCD) Psous K	29,78	NbPi (CATL) Psous K	28,45	1,33
NbPi (ACHLxSTOCD xCMBL) Psous K	29,22	NbPi (STOCD) Psous K	27,51	1,71
NbPi (ACHLxTYPL) Psous K	27,38	NbPi (CATLxSTOCD xTYPL) Psous K	26,12	1,26
NbPi (ACHLxCHFL xSTOCDxCMBL) Psous K	23,6	NbPi (ACHLxSTOCD) Psous K	25,74	-2,14
NbPi (CHFLxSTOCD) Psous K	22,01	NbPi (CATLxTYPL) Psous K	24,72	-2,71
NbPi (ACHLxCHFL xTYPL) Psous K	21,7	NbPi (ACHL) Psous K	18,59	3,11
NbPi (STOCDxCMBL) Psous K	21,34	NbPi (TYPL) Psous K	16,63	4,71
NbPi (CATLxACHL) Psous K	21,33	NbPi (ACHLxTYPL) Psous K	16,33	5
NbPi (CATLxACHL xCHFLxCMBL) Psous K	20,07	NbPi (CATLxACHL xTYPL) Psous K	-48	68,07
NbPi (CATLxACHL xSTOCD) Psous K	18,71	NbPi (CATLxACHL xSTOCDxTYPL) Psous K	-60,9	79,61
Psous K	0			

TABLE 8.8 – Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées au nombre de pièces, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode d'optimisation en utilisant ou non les données des logements avec chauffage électrique

Variables du modèle ref	Amélioration de l'erreur ref (%)	Variables du modèle ref sans chauffage	Amélioration de l'erreur ref sans chauffage (%)	Apport des informations non chauffage
NbPi (CATLxACHL xSTOCDxCMBL) Psous K	-18,1	NbPi (CATLxACHL) Psous K	28,54	-46,64
NbPi (CATLxACHL xCMBL) Psous K	-18,3	NbPi (CATLxACHL xSTOCD) Psous K	28,41	-46,71
NbPi (CATLxACHL xCHFL) Psous K	-20,3	NbPi (CATL) Psous K	27,7	-48
NbPi (CATLxACHL xCHFLxSTOCD) Psous K	-20,4	NbPi (CATLxSTOCD) Psous K	27,46	-47,86
NbPi (CATLxACHL xCHFLxCMBL) Psous K	-20,6	NbPi (CATLxTYPL) Psous K	26,57	-47,17
NbPi (CATLxACHL xTYPLxCMBL) Psous K	-21,4	NbPi (CATLxSTOCD xTYPL) Psous K	25,51	-46,91
NbPi (CATLxACHL xCHFLxSTOCD xCMBL) Psous K	-22,8	NbPi (STOCDxTYPL) Psous K	25,22	-48,02
NbPi (ACHLxCHFL xSTOCDxTYPL xCMBL) Psous K	-22,8	NbPi (CATLxACHL xTYPL) Psous K	23,94	-46,74
NbPi (CATLxACHL xCHFLxTYPL) Psous K	-24,1	NbPi (ACHLxSTOCD xTYPL) Psous K	23,84	-47,94
NbPi (CATLxACHL xSTOCDxTYPL xCMBL) Psous K	-24,9	NbPi (ACHLxSTOCD) Psous K	23,7	-48,6
NbPi (ACHLxSTOCD xCMBL) Psous K	-25,7	NbPi (STOCD) Psous K	23,32	-49,02
NbPi (ACHLxSTOCD xTYPLxCMBL) Psous K	-25,9	NbPi (CATLxACHL xSTOCDxTYPL) Psous K	20,96	-46,86
NbPi (ACHLxCHFL xSTOCD) Psous K	-26,1	NbPi (TYPL) Psous K	14,59	-40,69
NbPi (CATLxACHL) Psous K	-26,2	NbPi (ACHL) Psous K	12,1	-38,3
NbPi (CATLxACHL xCHFLxSTOCD xTYPL) Psous K	-27,6	NbPi (ACHLxTYPL) Psous K	11,57	-39,17
Psous K	0			

TABLE 8.9 – Amélioration de l'erreur de prédiction de la thermosensibilité par départ en utilisant différentes variables intensives couplées au nombre de pièces, par rapport au modèle de référence (puissance souscrite par catégorie) pour la base de données venant de l'association par la méthode de référence en utilisant ou non les données des logements avec chauffage électrique

## 8.2 Annexes du chapitre 4 sur le modèle de consommation électrique

### 8.2.1 Consommation et puissance appelée des principaux usages résidentiels

Type d'appareils	Taux d'équipement des foyers français (%)	Puissance pic appelée (W)	Energie moyenne consommée kWh/an des foyers équipés
Réfrigérateurs	99,8 [a]	190 [f]	356 [l]
Congélateur	87,3 [a]	190 [f]	556 [l]
télévision	97,5 [a]	100 [l]	307 [l]
télévision veille	97,5 [a]	4,2 [l]	36 [l]
périphériques (tv+audio)	97,5 [a]	100	203 [l]
Ordinateur Fixe en service	91 [b]	125 [l]	297 [l]
Ordinateur Fixe en mode Veille	91 [b]	10 [l]	23 [l]
Ordinateur Portable en service	38 [b]	70 [l]	54 [l]
Ordinateur Portable en mode veille	38 [b]	6 [l]	10 [l]
Jeux vidéos	47 [b]	160 [l]	30 [l]
Connexion internet (moyenne box, routeur, borne wifi)	60,5 [a]	5 [l]	44 [l]
Périphériques informatiques	60 [b]	10	28 [l]
Machine à laver	94,9 [a]	2500 [f]	169 [l]
Sèche linge	55 [c]	2500 [f]	400 [i]
Fer à repasser	59,8 [c]	1700 [h]	40 [i]
Cuisinière à four	25,2 [c]	3500 [h]	173 [h]
Plaques de cuisson	53 [c]	1300 [h]	277
Four classique	46,7 [c]	2300 [h]	224 [h]
Four micro-ondes	85,2 [a]	1400 [f]	75 [h]
Cafetière (filtre et expresso)	67,3 [c]	1000 [h]	30 [i]
Lave vaisselle	51,6 [a]	2200 [f]	273 [l]
Bouilloire	46,6 [c]	2000 [h]	70 [i]
Friteuse	36,4 [c]	1500 [i]	10 [i]
Hotte aspirante	57 [a]	150 [i]	15 [i]
Robot	51,1 [a]	700 [j]	40 [i]
ECS électrique	44 [b]	2000 [i]	700-5600 [i]

Type d'appareils	Taux d'équipement des foyers français (%)	Puissance pic appelée (W)	Energie moyenne consommée kWh/an des foyers équipés
Chauffage électrique	33 [b]	Modèle en fonction des maisons	
Climatisation		Modèle en fonction des maisons	
Eclairage	100,0	800 [g]	300 [l]
Aspirateur	75,2 [c]	1200 [i]	18 [i]
Nettoyeur vapeur	29,2 [c]	1500	110 [i]
Sèche-cheveux	69,3 [c]	1500	15 [i]
Piscine enterrée	2,5 [d]	1050 [d]	2123 [d]
Piscine hors-sol	1,6 [d]	475 [d]	382 [d]
VMC	35 (selon année et type du logement ) [e]	40	310 [i]

TABLE 8.10 – Principales caractéristiques des usages résidentiels.

Sources du tableau :

[a] Insee 2009.

[b] B. Grinden, N. Feilberg. Remodece 2008 [88].

[c] Insee 2010 (via GIFAM).

[d] Fédération des professionnels de la piscine (FPP), 2007.

[e] Observatoire de la qualité de l'air intérieur, CSTB. Etat de la ventilation dans le parc de logements français. Juin 2009.

[f] M. El Guedri, 2009 [67].

[g] M. Stokes *et al.*, 2004 [142].

[h] Projet ECUEL, 1999 [52].

[i] O. Sidler, 2009 [140].

[j] Programme SAVE, 1995 [139].

[k] ADEME, 2007 [26].

[1] Robert Angioletti, présentation du projet REMODECE, 2008.

Si aucune référence, il s’agit d’estimations prises directement sur les équipements.

*Merci à Elise Pupier, doctorante au CMA à MINES ParisTech pour son aide dans ce travail de collecte des valeurs.*

### 8.2.2 Implémentation du cœur de calcul C++

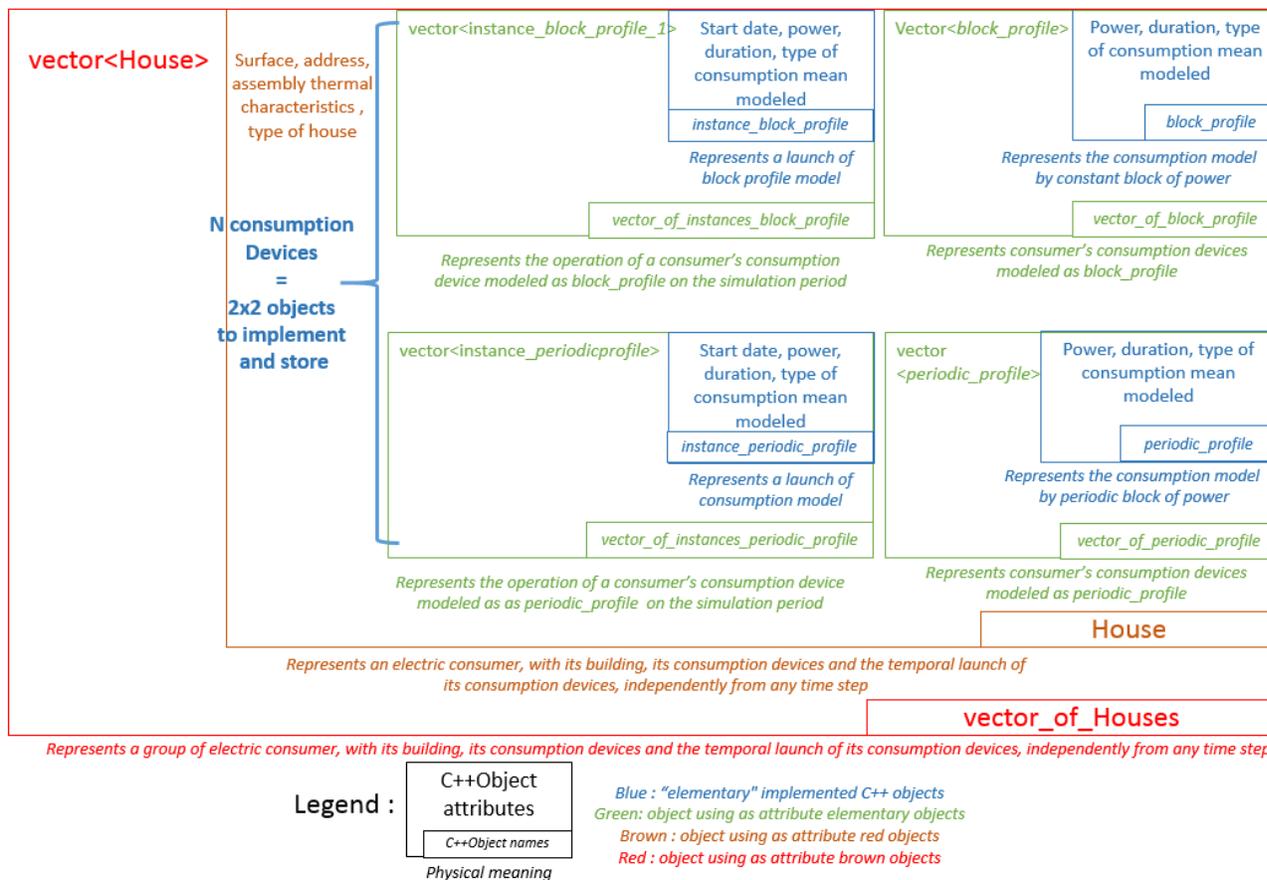


FIGURE 8.3 – Illustration de l’implémentation C++ du cœur de calcul du simulateur de courbes de charge

### 8.3 Annexes du chapitre 5 sur la validation et la calibration du modèle

#### 8.3.1 Calcul des consommation en énergie par usage de références pour le secteur résidentiel

Les données utilisés du rapport [22] ont été mises dans le tableau 8.11.

Type d'usage	Consommation électricité MI (TWh)	Pourcentage de consommation MI (%)	Taux de possession usage MI (%)	Consommation électricité IC (TWh)	Pourcentage de consommation IC (%)	Taux de possession usage IC (%)
Chauffage	32,1	30,4	21,7	13,5	27,6	15,6
ECS	13,6	12,9	48,8	6,6	13,5	33,8
Cuisson	8,1	7,7	44,7	3,9	8,0	41,4
Spécifique	51,6	49,0	100,0	24,9	50,9	100,0
Total	105,4	100	-	48,9	100	-

TABLE 8.11 – Tableau récapitulatif des données CEREN [22] utilisées

Les valeurs de consommation électrique par logement du tableau 5.2 sont alors obtenue en divisant par le nombre de logement de chaque type (voir les valeurs dans le tableau 8.12).

Type de logement	Nombre (milliers)
Maison individuelle	15897
Immeuble collectif	12162

TABLE 8.12 – Nombre de logements par type [22]

#### 8.3.2 Article de journal sur la méthode pour décomposer les départs HTA en profils élémentaires à partir des données d'Enedis

# A novel method for decomposing electricity feeder load into elementary profiles from customer information

Alexis Gerossier\*      Thibaut Barbier\*      Robin Girard

{alexis.gerossier},{thibaut.barbier}@mines-paristech.fr

\* Both authors equally contributed to this paper

*MINES ParisTech, PSL Research University, Center for processes, renewable energy and systems (PERSEE), 1 rue Claude Daunesse, 06904 Sophia Antipolis, France*

June 2017

## Abstract

To plan a distribution grid involves making a long-term forecast of sub-hourly demand, which requires modeling the demand and its dynamics with aggregated measurement data. Distribution system operators (DSOs) have been recording electricity sub-hourly demand delivered by their medium-voltage feeders (around 1,000—10,000 customers) for several years. Demand profiles differ widely among the various considered feeders. This is partly due to the varying mix of customer categories from one feeder to another. To overcome this issue, elementary demand profiles are often associated with customer categories and then combined according to a mix description. This paper presents a novel method to estimate elementary profiles that only requires several feeder demand curves and a description of customers. The method relies on a statistical blind source model and a new estimation procedure based on the augmented Lagrangian method. The use of feeders to estimate elementary profiles means that measurements are fully representative and continuously updated. We illustrate the proposed method through a case study comprising around 1,000 feeder demand curves operated by the main French DSO Enedis. We propose an application that uses the obtained profiles to evaluate the contribution of any set of new customers to a feeder peak load. We show that profiles enable a simulation of new unmeasured areas with errors of around 20%. We also show how our method can be used to evaluate the relevancy of different customer categorizations.

## 1 Introduction

### 1.1 Motivation

Electricity represented 18% of total final energy consumption in 2013 [3] and is expected to constitute a quarter of final energy consumption by 2040 [1]. 42% of global CO<sub>2</sub> emissions in 2012,

*i.e.* 13.8 gigatons of CO<sub>2</sub>, are due to electricity and heat production [2]. To reduce CO<sub>2</sub> emissions due to electricity, many states are developing energy transition strategies. This kind of transition involves significant changes to electricity flows in the distribution network (with *e.g.* decentralized production, improved efficiency of buildings and appliances, new uses and demand

### Nomenclature

$a^f$	Consumption trend of feeder $f$ relative to temperature
$b^f$	Temperature threshold of feeder $f$
$B$	Matrix of demand profiles of customer categories
$\beta$	Column vector associated with $B$
$c_k^f$	Annual consumption of category $k$ for feeder $f$
$d^f$	Demand of feeder $f$
$d_k$	Elementary profile of customer category $k$
$\varepsilon^f$	Residual term for modeling feeder demand $f$
$F$	Number of feeders
$f$	Feeder
$K$	Number of customer categories
$k$	Customer category
$m_k$	Average demand share of a given category $k$
$p_k^f$	Share of electricity used by category $k$ for feeder $f$
$\sigma_k^2$	Empirical variance of $p_k^1, \dots, p_k^F$
$T$	Number of instants
$t$	Instant
$T^f$	Outside temperature of feeder $f$
$u$	Vector $(1, \dots, 1)^\top$ of length $K$
$v$	Vector $(T^{-1}, \dots, T^{-1})^\top$ of length $T$
$V_{inter}$	Inter group variability
$V_{tot}$	Total variance
$X$	Matrix of feeder demands
$x$	Column vector associated with $X$
$y$	Year
$\otimes$	Kronecker product

response enabling energy consumption management [18]).

These changes impact the planning process of distribution system operators (DSOs). The current network planning process considers the two most extreme situations [16], *i.e.* maximum demand with minimum supply, and maximum supply with minimum demand. While planning with such a method does not require a deep modeling of the different dynamics and their correlations, it does not take into account the aggregation effect between supply and demand [15]. The above-mentioned changes make it necessary to model all of the aggregated demand dynamics.

## 1.2 Literature review

In this section we present two kinds of existing approach for modeling aggregated demand. The first is bottom-up, and uses individual customer profiles, which are summed to obtain aggregated demand. The second is a global approach in which the aggregated load curve is directly modeled using aggregated measurement data.

### 1.2.1 Bottom-up approaches

Measuring the electricity demand of individual electricity customers is a simple way to establish their load profiles and dynamics, and therefore a necessary step in bottom-up modeling. The current smart-meter roll-out in Europe will provide precise measurements of individual demand profiles. Around 80% of customers are scheduled to receive a smart-meter by 2020 [28]. However, this massive deployment is hindered by cost and privacy issues [21]. In 2014, only 23% of smart-meters in the European Union were installed in localized areas for private customers [13]. In some countries, this share is still insuffi-

cient to be representative, and the corresponding deployment is too recent to adequately cover long periods. To deal with the lack of individual measurements and characterize the behavior of electricity customers, researchers have attempted to classify them into different categories.

The classification of electricity demand profiles is a flourishing research topic (see reviews [19], [25]). Researchers use individual measurements from smart-meters as input and apply different clustering methods [31]. This reduces the dimension, which makes it easier to manipulate data [22]. With the resulting classification, each customer is associated with a cluster and its corresponding load profile [26]. The classification and the obtained load profiles can be used for a number of applications.

First, a fine classification can be made in order to help decision-makers design personalized policies for specific customers [7].

Secondly, the classification allows a DSO to plan its network and anticipate its investments [23, 27]. For example, the French DSO uses a model named "Bagheera" combining about 50 customer categories to plan its low-voltage network [16]. Classification is combined with the evolution of category distributions to forecast aggregated demand in prospective scenarios [5].

Last, classification and load profiles allow us to understand the contribution made by each category to aggregated demand [27].

Large measurement campaigns are necessary with these methods since a representative set of customers is required. This constraint makes continuous updating of the profiles difficult, which is an issue since it remains necessary to adapt the profiles to the changing consumption habits [4, 26].

### 1.2.2 Global approaches

In global approaches, models forecast aggregated electricity demand with past measurements and explanatory variables, such as expected temperature or sometimes economic progress [30].

In order to obtain past measurements, most DSOs have been recording the electric power delivered by their medium-voltage feeders (around 1,000—10,000 customers) for several years. These measurements are aggregated, but exhaustive, since all electricity customers' contributions are taken into account. This aggregated electricity demand data is considered as a "nonlinear, non-stationary series, and is often made up by a superposition of several distinct frequencies" [29] with daily to monthly periods in global models [8]. Additionally, the demand series can be divided into different parts (*e.g.* working time, holidays) [9, 17].

The global approach produces accurate forecasts. However, these are based on aggregated past measurements, which are not available when planning a new unmeasured zone. This type of planning is improved with specific information about customers, which DSOs possess thanks to the Customer Information System (CIS) [23]. The CIS stores information on all customers regarding their electric connection to the grid, annual energy consumption, type of contract, and contracted power.

In all of the reviewed global methods [29] for modeling demand dynamics, the explanatory variables used, such as expected temperature or sometimes economic changes [30], do not characterize the feeder-specific local features. In particular, none of them employs CIS general statistics.

Finally, the drawback of these methods when used for planning purposes is that they cannot adapt to a change in the mix of customer cate-

gories. For example, in the case of the development of a commercial area in a residential feeder, such methods fail to take into account the corresponding information. If the profile differences of the two sectors is not accounted for, this might result in an overestimation of the future peak and hence an over-sizing of the network.

### 1.3 Contributions

Our paper presents a novel method to estimate elementary profiles. The proposed method relies on a statistical model that takes into account the mix of customer categories. To do this, we assume that the demands aggregate different shares of elementary profiles associated with different customer categories. These profiles are optimally found by minimizing prediction errors in a new algorithm relying on the augmented Lagrangian method.

Unlike bottom-up methods, our method only requires several feeder demand curves and a description of customers. The advantages of aggregated measurements compared to a set of individual load curves are: the availability of long-term historical data, full representativeness, and continuous updates. We show that the method performs similarly or better than a bottom-up method in the literature when predicting new local areas.

We illustrate the proposed method through a case study comprising around 1,000 feeder demand curves operated by the main French DSO Enedis. The profiles obtained are essential to size the distribution network. This is illustrated by an application that evaluates the contribution of any set of new customers to a feeder peak load. We show that profiles enable a simulation of new unmeasured areas with errors of around 20%. We also show how our method can be used

to evaluate the relevancy of different customer categorizations.

### 1.4 Description of the paper

In section 2, the methodology is described. A case study is presented in section 3 with the resulting profiles by category. Section 4 describes two applications that use the obtained profiles. One is employed to estimate the contribution of set of new customers to a feeder peak load. The other evaluates forecasting errors for unmeasured areas, by testing different categories and comparing performances with a similar framework case study in the literature. Finally, some conclusions are presented and discussed in section 5.

## 2 Methodology

### 2.1 The problem of recovering load profiles and the forecasting method

Our paper assumes that the sub-hourly demands  $d^f(t)$  of a feeder  $f$  aggregate different profiles  $d_1(t), \dots, d_K(t)$  associated with  $K$  categories of customers with weights  $p_1^f, \dots, p_K^f$ ,

$$d^f(t) = \sum_{k=1}^K p_k^f d_k(t) + \varepsilon^f(t). \quad (1)$$

We take the elementary profiles  $d_k(t)$  to be common to all feeders, while the weights vary from one feeder to another. The corresponding residual term  $\varepsilon^f(t)$  is meant to be small. The time  $t$  can vary along any set. The aim is to recover unknown elementary electricity profiles  $d_k(t)$ . For each feeder  $f \in \{1, \dots, F\}$ ,  $d^f(t)$  is observed and, thanks to the CIS, for each category  $k \in \{1, \dots, K\}$ , we also have access to the weight  $p_k^f$ .

The process of obtaining proportions from the CIS and defining categories is the categorization step, and is described in subsection 2.3. Once the  $K$  profiles have been obtained on a set of feeders, it is possible to turn Equation (1) into a simulation algorithm. The process is described in Figure 1. In the signal processing community, the corresponding problem is called blind signal separation and is well-known (see *e.g.* [11]).

## 2.2 Optimization problem

The aim is to find the elementary profiles  $d_k(t)$  from aggregated demand  $d^f(t)$  according to Equation (1). We write and solve the following optimization problem.

To mathematically write this optimization problem, we define a matrix  $A$  of size  $(F, K)$  whose elements are proportions  $p_k^f$  for  $k \in \{1, \dots, K\}$  and  $f \in \{1, \dots, F\}$ . Aggregated demands  $d^f(t)$  for all feeders and instants  $\{1, \dots, T\}$  are gathered in a matrix  $X$  of size  $(F, T)$ . We are trying to compute demand profile  $d_k(t)$  for all categories and instants: these unknown values can be put in a matrix  $B$  of size  $(K, T)$ . It is useful to define  $\beta$  (resp.  $x$ ), the column vector obtained by stacking rows of  $B$  (resp.  $X$ ) on top of each other. Two constraints limit the values of matrix  $B$ :

1. Each component of  $\beta$  is an electricity demand. Since electricity producers are not considered in this paper, components should be positive.
2. For each class  $k$ , components should have an average unit, *i.e.*  $\sum_t d_k(t) = T$ , to have comparable profiles. To write this constraint in mathematical terms, we define the column of length  $K$ ,  $u = (1, \dots, 1)^\top$ , and the column of length  $T$ ,  $v = (T^{-1}, \dots, T^{-1})^\top$  in

order to write the average unit constraint, with a Kronecker product  $\otimes$ , as  $(I_K \otimes v^\top)\beta = u$ .

The optimization problem then writes

$$\begin{aligned} \min_{\beta} \quad & \|x - (A \otimes I_T)\beta\|^2 & (2) \\ \text{s.t.} \quad & \beta \geq 0 \\ & (I_K \otimes v^\top)\beta = u \end{aligned}$$

An alternating direction method of multipliers [10] is used to recursively solve problem (2):

1. minimize the function with the equality constraint by employing the augmented Lagrangian method,
2. retain only positive components to satisfy the positivity constraint,
3. adjust a penalty variable balancing positivity and the minimization.

The algorithm is implemented with the R language [24]. Special care is taken on the first step, since the minimization requires inverting a large matrix of size  $K(T + 1)$ . With common Kronecker product rules, matrix to be invert is reduced to size  $K$  divided the number of flops by approximately  $T^3$ .

## 2.3 Categorization of electricity customers

The aggregated demand profile  $d^f(t)$  of a feeder  $f$  aggregates a large group of customers (a few thousands). The CIS provides general features on these customers, *i.e.* annual consumption, type of contract, and contracted power, which can be used to cluster them into  $K$  different categories. Once the features are selected, the

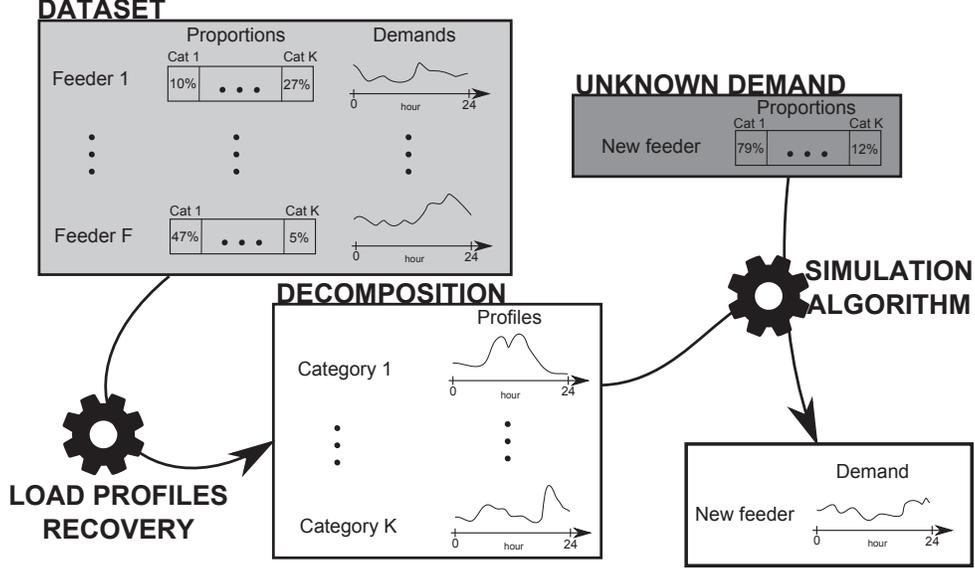


Figure 1: Diagram detailing the method. A dataset of  $F$  feeder measurements is used to find the  $K$  category profiles. Once the load profiles recovery is operated, a new feeder whose category distribution is known can be run through the simulation algorithm in order to obtain its expected demand.

total annual consumption  $c_k^f$  of a category  $k \in \{1, \dots, K\}$  in a feeder  $f \in \{1, \dots, F\}$  is computed from each annual individual consumption. The corresponding weight  $p_k^f$  is a normalized version of this consumption

$$p_k^f = \frac{c_k^f}{\sum_{k=1}^K c_k^f} \in [0, 1] \quad (3)$$

It is important that the size of the dataset  $F$  should be larger than the number of categories  $K$ . Empirically, it was observed that the condition  $F > 5K$  is preferable in order to obtain a wide range in the set of category distributions, and thus a more precise result. Features should be general enough to keep a reasonably low  $K$  for three reasons: (i) to obtain a robust profile, (ii)

to avoid an excessively long computing time, and (iii) to ensure that user privacy is not violated.

Figure 2 sets out four different categorizations, based on information from the CIS. The first categorization divides the total energy into two groups: residential and tertiary. The second splits the tertiary into 7 categories to make a total of 8 categories, *i.e.* residential, agriculture, commercial, public equipment, office and hospital, industry, restaurant and hotel, and medium-voltage (MV) customers (*e.g.* large buildings that have a specific contract with the operator). A 9-group division results from splitting the residential share into two groups: base tariff and special tariff<sup>1</sup>. Finally, an even more precise cat-

<sup>1</sup>Special tariff charges less during fixed off-peak peri-

egorization, *i.e.* 12 groups, is proposed. Commercial buildings are split into 2 categories reflecting low and high annual consumption. Similarly, MV customers are divided into 3 groups: low, medium and high.

On Figure 2, category heights for a category  $k$  represent the average demand shares for a given category  $m_k = \frac{1}{F} \sum_{f=1}^F p_k^f$ .

The share in category distribution is different for every feeder. For instance, there are more restaurants in a city center than in a rural area and so the two electricity shares are different. This share has to vary between feeders to efficiently compute the demand profiles. We computed the coefficients of variation

$$\frac{\sigma_k}{m_k} \quad (4)$$

where  $\sigma_k^2$  is the empirical variance of  $p_k^1, \dots, p_k^F$ . The coefficients are always higher than 40%, and thus the different categorizations are sufficiently spread from one feeder to another for our algorithm.

## 3 Case study

### 3.1 Data description

In this case study, we use electricity feeder demand measured every ten minutes in 3 geographical regions in France. Data come from the main French DSO, Enedis. The three regions encompass a large French city and the surrounding countryside. The three cities are Blois, Lyon and Rennes. Each region is divided into around 500 feeders, and each of these feeders provides electricity for about 1,000 customers. For each feeder, we know the demand measured for 4 years

from 2010 to 2013. We discard some feeders because the measures are too scarce and their overall quality is not sufficient. This can result from database errors or from network reconfiguration or physical injuries on the grid [17]. Ultimately, between 200 and 400 feeders are selected for each region.

### 3.2 Temperature effect and normalization

Aggregated demand measurements cannot be directly compared since some feeders are connected to more customers than others, causing a large discrepancy in average consumption. In order to be used as inputs in the method, measurements therefore need to be pre-processed. The two steps of this pre-processing are: removal of the temperature effect, and normalization by weekly consumption.

Electricity demand is mostly influenced by outdoor air temperature, as residents turn on electric devices to adjust their indoor temperature (heating and air conditioning). In France, the air conditioning effect is low and not considered in this paper, but the heating effect is high during cold weather. French electric demand represents 40% of the European thermal sensitivity [14]. Indeed, since most French heating devices are electric, demand strongly increases when temperature decreases. However, this effect is well understood and can be removed and treated separately with a method used by the French TSO [20, pp 11–12]: one linear regression for each hour of week. Therefore, for each feeder  $f$ , we can determine a temperature threshold  $b^f$  and a trend  $a^f > 0$  such as for each degree colder than threshold  $b^f$ , demand increases by  $a^f$ . A new demand series is defined from the

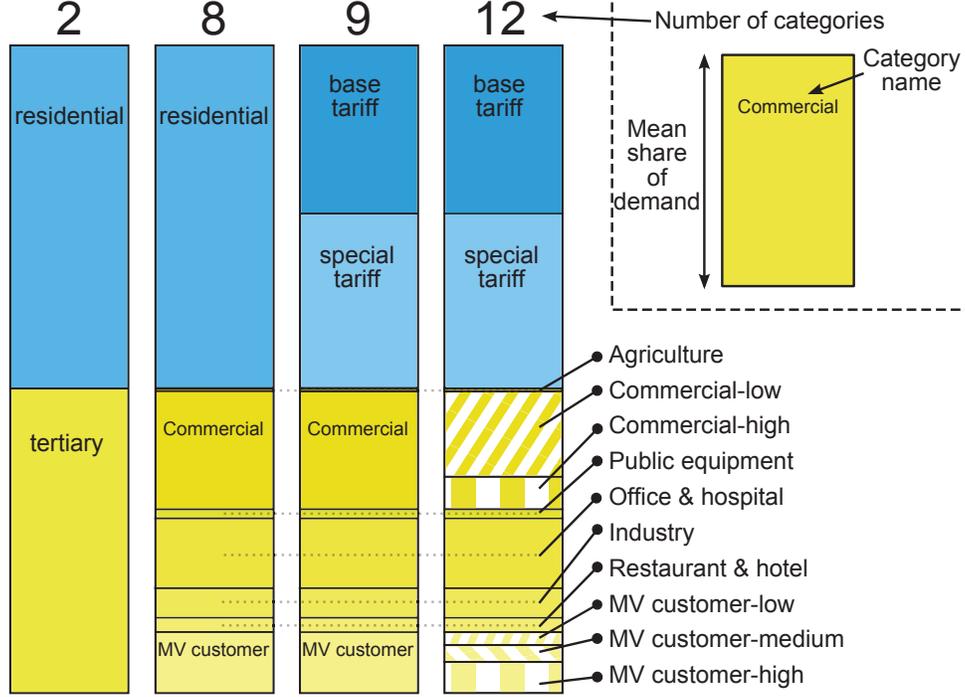


Figure 2: Example of different categorizations (in 2, 8, 9 or 12 groups) for the region near Lyon. There are  $F = 320$  feeders in this dataset. The height of a division shows the mean share of the category in all feeders in the region.

initial  $d_0^f(t)$

$$d_1^f(t) = \begin{cases} d_0^f(t) & \text{if } T^f(t) > b^f \\ d_0^f(t) - a^f (b^f - T^f(t)) & \text{otherwise.} \end{cases} \quad (5)$$

where  $T^f(t)$  is the outside temperature of feeder  $f$  at instant  $t$ . In fact, trends  $a^f$  and threshold  $b^f$  are calculated for each hour of the day but the hour index is omitted for clearer notation. The new series is thus supposed to be independent from the temperature, and demand dynamics are supposed to be similar during cold and warm periods.

To obtain comparable measurements between feeders, demand is normalized. Each measurement within a given week is divided by the energy it consumed during that week. This total energy can be predicted using different models, such as that employed in [6], and is thereafter supposed to be known. After the normalization, data values fluctuate around a dimensionless value equal to 1.

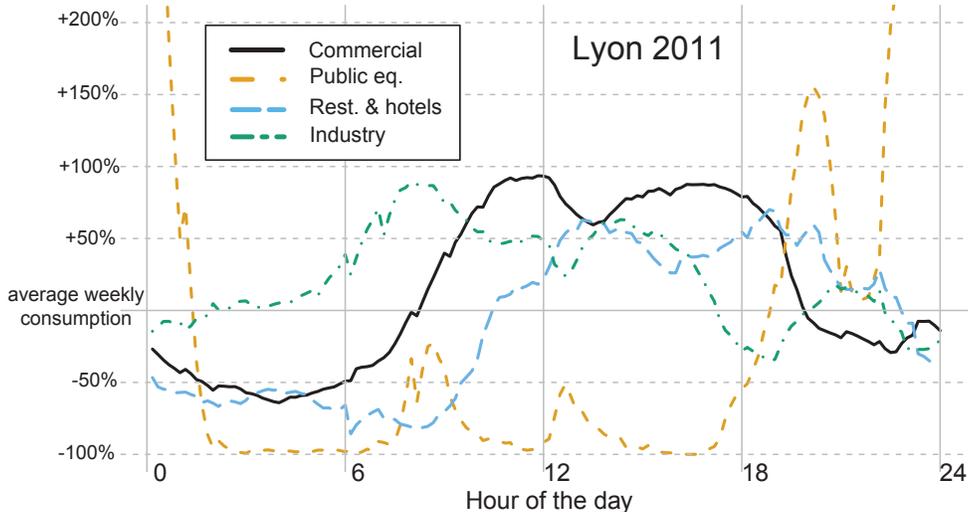


Figure 3: Weekday profiles of 4 different categories computed with the algorithm (9 overall categories) using aggregated consumption data relating to Lyon in 2011. Plots represent the variations around the average weekly consumption and not absolute consumptions.

### 3.3 Profiles

As previously described (see Figure 1), we disaggregated the electricity demand in order to recover a load profile  $d_k(t)$  for each category  $k \in \{1, \dots, K\}$ . The number of overall categories depends on the customer categorization: 2, 8, 9 and 12 categories were tried out (see Figure 2). A total of 12 datasets is formed (for each region: Blois, Lyon and Rennes; and for each year: from 2010 to 2013) and separately used as input into matrix  $X$  in problem (2).

Figure 3 presents the profiles obtained for  $K = 9$  with only 4 categories shown: commercial, public equipment, restaurant and hotel, industry. Profiles are computed with the demand dataset of Lyon in 2011. Profiles are presented for a typical weekday (144 values, once every 10 minute). Since we have normalized the data, the

variations around the average weekly consumption are displayed. Different effects are noteworthy, *e.g.* the electricity consumption of commercial buildings increases by around 75% during working hours, and decreases by 50% during the night. Conversely, the consumption of public equipment (mainly public lighting and lifts) greatly increases at night. These profiles are a pertinent way to understand electricity demand patterns. Profiles can be plotted for other datasets (another region or another year) in order to analyze specific characteristics.

## 4 Applications of the method

### 4.1 Estimation of the contribution of new customer sets to a feeder peak load

To plan the expansion of a new area, the DSO has to estimate the evolution of peak demand. The profiles obtained enable it to quantify and forecast the contribution of the new set of customers in the peak load demand. Indeed, for a feeder  $f$  at year  $y_0$  with proportions  $p_{1,y_0}^f, \dots, p_{K,y_0}^f$  we can determine the residuals  $\varepsilon_{y_0}^f(t)$  in Equation 1 and for new proportions  $p_{1,y_1}^f, \dots, p_{K,y_1}^f$  in a future year  $y_1$  the forecast demand is obtained by

$$d_{y_1}^f(t) = \sum_{k=1}^K p_{k,y_1}^f d_k(t) + \varepsilon_{y_0}^f(t). \quad (6)$$

Figure 4 depicts the peak change obtained with this formula in the case of different evolutions for both offices and special-tariff residential consumers. In this case study, the considered feeder is from the Lyon region and has the following distribution of customers: 30% commercial, 15% offices, 30% basic residential and 20% special special-tariff residential. The initial peak occurs at 12:10 and is 650 kW. The profiles used are taken from the 9-category breakdown. We quantify the influence on the peak value (black lines with value added to the initial peak value, per 50 kW) by adding an office category load (Y axis) and a special-tariff residential load (X axis). We also depict the evolution of the peak hour (black dashed line). Adding offices contributes to increasing the 12:10 peak, whereas the residential load increases the 23:00 peak, which corresponds to the start of the special-tariff period.

This is an illustration of an application of the method that can for example help decision-makers to choose between two projects (offices or

a new residential area) and quantify the impact on the existing feeder demand.

### 4.2 Evaluation, comparison of the method and category relevancy

#### 4.2.1 Simulation evaluation

Thanks to the computed profiles, the aggregated demand of a feeder can be simulated. Each category profile is multiplied by the consumption share of the category. The category distribution is the only information required for the simulation; there is no need for historical demand recordings. We show a simulation example on Figure 5. Demand is simulated with only two categories: residential (green area) and tertiary (orange area). We sum the two profiles multiplied by their respective share (here 75% residential and 25% tertiary consumption). The measured consumption of a feeder with a 75/25 proportion is superimposed in black. The respective contribution of the two categories at each time step is clearly observable on the aggregated demand.

To assess the quality of the model, we use the Root-Mean-Square Error (RMSE) index. For each region and for each year, we compute profiles for 2, 8, 9 and 12 overall categories and use them to simulate new feeders. We then compare the simulation with actual demand with a leave- $k$ -out approach ( $k = 50$ ). This means that a subset of  $k$  feeders (that are not used in the training stage) is simulated. An RMSE for each of these feeder subsets is obtained and the average value is computed. This process is repeated 100 times to remove the volatility effect caused by the random subset of a 50-feeder selection. Computation takes roughly 16 hours for every region and every year on a 3.50 GHz machine.

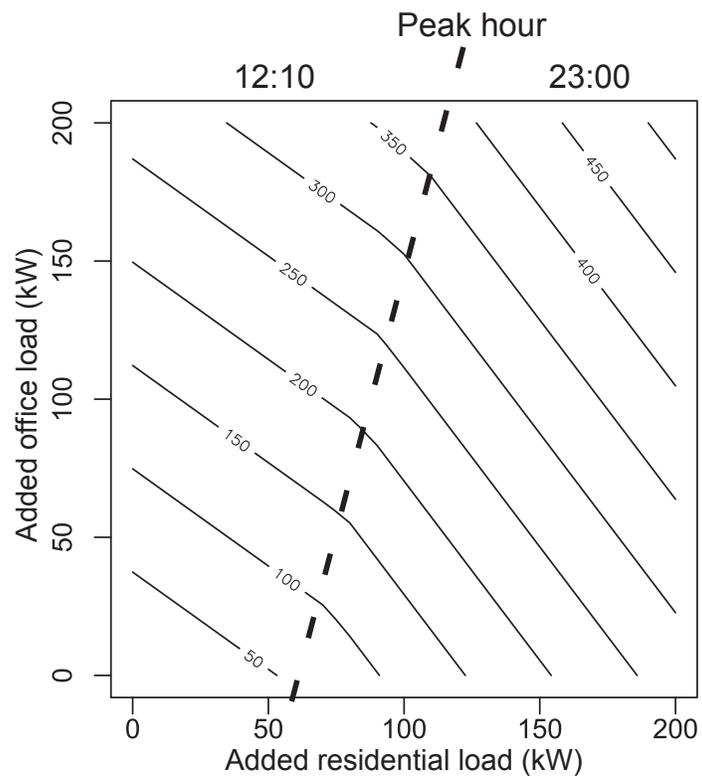


Figure 4: Peak change with a new load in a given feeder.

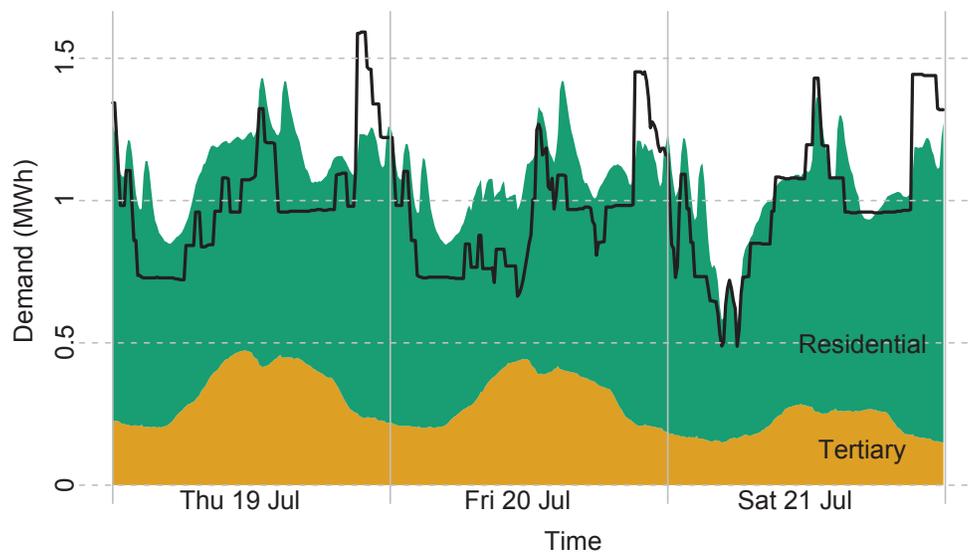


Figure 5: Simulation for one feeder. The profiles were obtained using demand data from Blois for 2012. The black line represents the actual consumption of the unknown feeder (not used in the training dataset). Our algorithm obtained two profiles: the orange part represents the tertiary demand and the green part the residential demand.

Table 1 reports the average RMSE and its deviation for the Blois, Lyon and Rennes during the 4 years for different numbers of categories. As a reminder, with consumption normalization, average consumption is dimensionless and equal to 1 (see Section 3.2). Hence, the RMSE reported is also dimensionless, and can be expressed as a percentage.

#### 4.2.2 Category relevancy

Average RMSE is 22.59% for Blois, 18.16% for Lyon and 22.42% for Rennes with 9 categories. The errors are highly dependent on the regions, meaning that some regions are less predictable than others. Increasing the number of categories improves the overall model quality. The 8 category scheme almost always outperforms the 2 category one (by 2.5%). The 9 category scheme slightly improves results compared to the 8 version (by 1%), and so dividing customers into basic and special tariffs is meaningful. However, splitting small categories into even smaller categories is not recommended, as can be seen by the poor results of the 12 category scheme. A first reason may come from the use of CIS for classification: previous works have stated that using directly the CIS classification does not necessarily lead to the best profiles [12].

Another reason can come from the inter-group variability. As in any blind source separation task, a class is easy to recover and predict if it is distinctly separated from the other classes, and if it is observed in many different configurations. The statistics literature proposes many different separation metrics, but the simplest is a ratio between an inter-group variability measure and a total variability. In this context, since the variable of interest is a vector or even a curve it is not obvious to define the variability. We propose

to define an inter-group variability measure with the weighted distance between  $d_k$  and  $d^f$

$$V_{inter} = \sum_{f,k} p_k^f \|d_k - d^f\|_2^2,$$

and a total variance by  $V_{tot} = \sum_f \|d^f\|_2^2$ , where  $\|x\|_2^2$  is the sum of the square of a vector  $x$ . The ratio between inter-groups and the total variance should be as high as possible. Measuring the diversity of configurations in which the final signal is observed can be related to the variance  $\sigma_k^2$  and mean  $m_k$  of  $p_k^f$  among the feeders, the larger this variance and mean the more accurate the estimation will be. These separations and variability measures can be used to evaluate the value of adding categories. The inter-variance requires the computation of the  $d_k$  but  $\sigma_k^2$  and  $m_k$  can be computed before any estimation.

#### 4.2.3 Comparison to other models

Errors are higher than for middle-term forecasting methods, which can be around 7 to 10% of RMSE (see *e.g.* [8], [17]). However, our problem is different, and the relationship between the demand for a feeder  $f_1$  for a given year  $y_0$  and the demand for a feeder  $f_1$  for the next year  $y_1$  is much stronger than the relationship between the consumption of a feeder  $f_1$  and the consumption of feeder  $f_2$  for the same year  $y_0$ .

Framework of Andersen *et al.* is more similar to ours [5]. This presents “a model calculating local consumption by categories of customer with specific consumption profiles and different weights in local areas”. Unlike us, their profiles are obtained by clustering representative smart-meter measurements, *i.e.* a bottom-up method. Their results from simulating local areas without using past measurements are expressed with  $R^2$

value and are between 0.95 and 0.56 (their mean  $R^2$  is 0.84). In their case study, the mean consumption of areas is 55.3 MW while in our case, for a given feeder it is between 0.5 and 7 MW. In order to compare our method with their method, we aggregated our areas to obtain similar average power levels and computed the  $R^2$  between prediction and measurements. The results are shown in Table 2.

The performances of our method are a little higher than Andersen *et al.*'s method in the Lyon and Rennes case studies, and similar in the Blois study.

Area	Avg. demand 2010 (MW)	$R^2$
Blois	31.5	0.82
Lyon	46.2	0.88
Rennes	37.4	0.87

Table 2: Coefficient of determination  $R^2$  for different areas showing the predictive performance of our method with a 9-category breakdown. The prediction of a group of 20 feeders is compared to the measured demand of the 20 feeders. We also report the average demands, which are comparable to the areas described by Andersen *et al.* [5] with similar  $R^2$  values : on average they found  $R^2$  of 0.84 for predicting different areas with an average demand of 55.3 MW.

## 5 Conclusions

Our paper has proposed a novel method to estimate elementary profiles. The main assumption of the method relies on feeder demands that aggregate various shares of elementary profiles associated with different customer categories. The profiles are optimally found by minimizing pre-

diction errors in a new algorithm relying on the augmented Lagrangian method.

Unlike bottom-up methods that require individual load curves, our method only requires several feeder demand curves and a description of customers. One of the advantages of using aggregated measurements on a set of individual load curves is that they can be updated regularly and are fully representative. In the meantime, we have shown that our method performs similarly or better than a bottom-up method in the literature to predict a new local area.

The method has been applied in a case study comprising three zones in France, with around 300 available feeder measurements over 4 years per zone. The result is a load profile for each customer category. We have shown that each load profile gathers intrinsic features of the given category.

A first application using the resulting profiles was presented for planning the expansion of a new area at DSO level. The resulting profiles allow for different quantification and forecasting of the contribution made by the new set of customers to peak load demand. This was illustrated by a case study on a specific feeder where the evolution of peak demand in the case of adding two share categories was discussed. A second application of the profiles is to simulate the electricity demand of the new unmeasured areas. This can be used to test the relevancy of various types of categorization (2, 8, 9 or 12 groups were tested). By analyzing forecasting errors, we observe that using more categories does not necessarily lead to more efficient models, several causes are discussed.

Further research could investigate the creation of an automatic way to create categories, *e.g.* by maximizing entropy information, to create the best profiles and minimize prediction errors.

Socio-demographic statistics might be efficient to accurately describe categories. Information such as mean household area and building age are very meaningful in electricity demand forecasting, and are thus areas for further research.

## 6 Acknowledgments

The authors would like to thank Enedis for supplying data to make this work possible, and particularly Nicolas Kong from the Direction Technique, Politiques et Stratégie group, for his precious expertise on measurements, CIS data, and the Bagheera planning model.

## References

- [1] I. E. AGENCY, *World Energy Outlook*, IEA Publishing, Paris, 2015.
- [2] —, *Co2 emissions from fuel combustion highlights 2016*, IEA Publishing, Paris, 2016, ch. Key trend in CO2 emission from fuel combustion, p. 12.
- [3] —, *Key World Energy Statistics*, IEA Publishing, Paris, 2016.
- [4] F. ANDERSEN, H. LARSEN, AND T. BOOMSMA, *Long-term forecasting of hourly electricity load: Identification of consumption profiles and segmentation of customers*, *Energy Conversion and Management*, 68 (2013), pp. 244 – 252.
- [5] F. ANDERSEN, H. LARSEN, AND R. GAARDESTRUP, *Long term forecasting of hourly electricity consumption in local areas in Denmark*, *Applied Energy*, 110 (2013), pp. 147 – 162.
- [6] F. ANDERSEN, H. LARSEN, N. JUUL, AND R. GAARDESTRUP, *Differentiated long term projections of the hourly electricity consumption in local areas. the case of Denmark West*, *Applied Energy*, 135 (2014), pp. 523 – 538.
- [7] N. BASSAMZADEH AND R. GHANEM, *Multiscale stochastic prediction of electricity demand in smart grids using bayesian networks*, *Applied Energy*, 193 (2017), pp. 369 – 380.
- [8] K. G. BOROOJENI, M. H. AMINI, S. BAHRAMI, S. IYENGAR, A. I. SARWAT, AND O. KARABASOGLU, *A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon*, *Electric Power Systems Research*, 142 (2017), pp. 58 – 73.
- [9] K. G. BOROOJENI, S. MOKHTARI, M. H. AMINI, AND S. S. IYENGAR, *Optimal two-tier forecasting power generation model in smart grids*, *CoRR*, abs/1502.00530 (2015).
- [10] S. BOYD, N. PARIKH, E. CHU, B. PELLEATO, AND J. ECKSTEIN, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, *Foundations and Trends® in Machine Learning*, 3 (2011), pp. 1–122.
- [11] J.-F. CARDOSO, *Blind signal separation: statistical principles*, *Proceedings of the IEEE*, 86 (1998), pp. 2009–2025.
- [12] G. CHICCO, R. NAPOLI, AND F. PIGLIONE, *Comparisons among clustering techniques for electricity customer classification*, *IEEE Transactions on Power Systems*, 21 (2006), pp. 933–940.

- [13] E. COMMISSION, *Benchmarking smart metering deployment in the EU-27 with a focus on electricity*, European Commission, Brussels, June 2014.
- [14] R. DE TRANSPORT D'ÉLECTRICITÉ, *Bilan prévisionnel de l'équilibre offre-demande d'électricité en france*, 2016, ch. Consommation d'électricité en France, p. 40.
- [15] J. DICKERT AND P. SCHEGNER, *Residential load models for network planning purposes*, in 2010 Modern Electric Power Systems, Sept 2010, pp. 1–6.
- [16] N. DING, *Load models for operation and planning of electricity distribution networks with metering data*, theses, Université de Grenoble, Nov. 2012.
- [17] Y. GOUDE, R. NEDELLEC, AND N. KONG, *Local short and middle term electricity load forecasting with semi-parametric additive models*, IEEE transactions on smart grid, 5 (2014), pp. 440–446.
- [18] M. JIN, W. FENG, P. LIU, C. MARNAY, AND C. SPANOS, *Mod-dr: Microgrid optimal dispatch with demand response*, Applied Energy, 187 (2017), pp. 758 – 776.
- [19] K. LE ZHOU, S. LIN YANG, AND C. SHEN, *A review of electric load classification in smart grid environment*, Renewable and Sustainable Energy Reviews, 24 (2013), pp. 103 – 110.
- [20] V. LEFIEUX, *Modèles semi-paramétriques appliqués à la prévision des séries temporelles. Cas de la consommation d'électricité.*, PhD thesis, Université Rennes 2, 2007.
- [21] E. MCKENNA, I. RICHARDSON, AND M. THOMSON, *Smart meter data: Balancing consumer privacy concerns with legitimate applications*, Energy Policy, 41 (2012), pp. 807–814.
- [22] F. MCLOUGHLIN, A. DUFFY, AND M. CONLON, *A clustering approach to domestic electricity load profile characterisation using smart metering data*, Applied Energy, 141 (2015), pp. 190 – 199.
- [23] A. MUTANEN, M. RUSKA, S. REPO, AND P. JARVENTAUSTA, *Customer classification and load profiling method for distribution systems*, IEEE Transactions on Power Delivery, 26 (2011), pp. 1755–1763.
- [24] R CORE TEAM, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015.
- [25] J. D. RHODES, W. J. COLE, C. R. UPSHAW, T. F. EDGAR, AND M. E. WEBBER, *Clustering analysis of residential electricity demand profiles*, Applied Energy, 135 (2014), pp. 461 – 471.
- [26] T. RÄSÄNEN, D. VOUKANTSIS, H. NISKA, K. KARATZAS, AND M. KOLEHMAINEN, *Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data*, Applied Energy, 87 (2010), pp. 3538 – 3545.
- [27] A. SEPPÄLÄ, *Load research and load estimation in electricity distribution*, theses, Technical research center of Finland, VTT Publications, Jan. 1996.

- [28] E. P. R. SERVICE, *Smart electricity grids and meters in the EU Member States*, European Parliament, Brussels, September 2015.
- [29] Z. SHAO, F. CHAO, S.-L. YANG, AND K.-L. ZHOU, *A review of the decomposition methodology for extracting and identifying the fluctuation characteristics in electricity demand forecasting*, Renewable and Sustainable Energy Reviews, (2016), pp. –.
- [30] Z. SHAO, F. GAO, Q. ZHANG, AND S.-L. YANG, *Multivariate statistical and similarity measure based semiparametric modeling of the probability distribution: A novel approach to the case study of mid-long term electricity consumption forecasting in China*, Applied Energy, 156 (2015), pp. 502 – 518.
- [31] J. L. VIEGAS, S. M. VIEIRA, R. MELÍCIO, V. MENDES, AND J. M. SOUSA, *Classification of new electricity customers based on surveys and smart metering data*, Energy, 107 (2016), pp. 804 – 817.

Region	Year	2 categories	8 categories	9 categories	12 categories
BLOIS	2010	24.36 (2.32)	<b>23.90 (2.75)</b>	24.04 (2.99)	26.78 (2.91)
	2011	23.87 (1.62)	<b>22.79 (1.42)</b>	22.91 (1.16)	24.78 (2.09)
	2012	22.84 (1.26)	22.54 (1.17)	<b>22.09 (1.24)</b>	24.09 (2.17)
	2013	22.34 (2.06)	22.31 (1.98)	<b>21.32 (1.96)</b>	23.34 (2.04)
	Average	23.35 (1.86)	22.89 (1.93)	<b>22.59 (1.98)</b>	24.75 (2.33)
LYON	2010	19.05 (2.39)	19.42 (2.71)	<b>18.29 (2.24)</b>	19.23 (1.94)
	2011	19.28 (1.24)	<b>18.06 (1.55)</b>	18.56 (1.20)	18.46 (1.42)
	2012	19.07 (1.35)	<b>18.21 (1.72)</b>	18.30 (1.34)	19.00 (1.86)
	2013	18.06 (1.03)	17.92 (2.03)	<b>17.49 (1.12)</b>	18.68 (1.91)
	Average	18.87 (1.59)	18.40 (2.05)	<b>18.16 (1.58)</b>	18.84 (1.79)
RENNES	2010	22.57 (0.96)	21.67 (1.23)	<b>21.59 (1.04)</b>	22.70 (1.57)
	2011	22.62 (1.22)	<b>21.54 (1.48)</b>	21.57 (1.06)	22.10 (1.08)
	2012	22.75 (1.11)	22.96 (0.99)	<b>22.39 (0.98)</b>	22.61 (0.84)
	2013	24.94 (1.03)	<b>23.99 (1.37)</b>	24.14 (1.26)	24.08 (1.37)
	Average	23.22 (1.08)	22.54 (1.28)	<b>22.42 (1.09)</b>	22.87 (1.25)

Table 1: RMSE (in %) of the models for the 3 different zones over the 4 years with a different number of categories. The simulation is run 100 times. We reported the average RMSE and its standard deviation between parentheses. The best results over the 4 numbers of categories are written in bold.





## Résumé

L'évolution de la consommation électrique est un point clé pour les choix à venir, tant pour les moyens de production d'électricité, que pour le dimensionnement du réseau à toutes ses échelles. Aujourd'hui, ce sont majoritairement des modèles statistiques basés sur les consommations passées et des tendances démographiques ou économétriques qui permettent de prédire cette consommation. Dans le contexte de la transition énergétique, des changements importants sont en cours et à venir, et la consommation future ne sera certainement pas une continuation des tendances passées. Modéliser ces changements nécessite une modélisation fine de type bottom-up de chaque contributeur de la consommation électrique. Ce type de modèle présente des challenges de modélisation, car il nécessite un grand nombre de paramètres d'entrée qui peuvent difficilement être renseignés de façon réaliste à grande échelle. En même temps, les données et informations de tout type n'ont jamais été autant disponibles. Cela représente à la fois un atout pour la modélisation, mais aussi une difficulté importante notamment à cause de l'hétérogénéité des données. Dans ce contexte, cette thèse présente une démarche de construction d'un simulateur de consommation électrique bottom-up capable de simuler différentes alternatives énergétiques à l'échelle de la France. Un travail de recensement, de classification et d'association des bases de données pour expliquer la consommation électrique a d'abord été mené. Ensuite, le modèle de consommation électrique a été présenté ; il a été validé et calibré sur une grande quantité de mesures de consommation électrique des départs HTA fournies par Enedis. Ce modèle a enfin pu être utilisé pour simuler différentes alternatives énergétiques afin d'aider au dimensionnement du réseau de distribution.

## Mots Clés

consommation électrique, grandes masses de données, réseau de distribution, modélisation bottom-up, alternatives énergétiques, courbe de charge

## Abstract

Future trend of electricity demand is a key point for sizing both the electricity network and the power plants. In order to forecast future electricity demand, current models mostly use statistical approaches based on past demand measurements and on demographic and economic trends. Because of current context of energy transition which comes along with important changes, future electricity demand is not expected to be similar to past trends. Modeling these changes requires a bottom-up modeling of each contributor to electricity demand. This kind of model is challenging because of the large number of input data required. At the same time, data and information are more and more available. Such availability can be considered both as an asset for modeling and as an important issue because of data heterogeneity. In this context, this dissertation offers an approach to build a bottom-up load curve simulator which enables to simulate prospective scenarios at the scale of France country. Firstly, an assessment, classification, and matching of the large databases explaining the electricity demand have been performed. Then, the electricity demand model has been presented. It has been validated and calibrated on Enedis' large volumes of electricity demand measurements of medium voltage feeders. Finally, this model has been used to simulate several prospective scenarios in order to improve the electricity distribution network sizing.

## Keywords

electricity demand, big data, distribution network, bottom-up modeling, prospective scenarios, load curve