



HAL
open science

Apport des mesures des compteurs Linky pour la connaissance des charges du réseau de distribution

Guilherme Ramos Milis

► **To cite this version:**

Guilherme Ramos Milis. Apport des mesures des compteurs Linky pour la connaissance des charges du réseau de distribution. Energie électrique. Université Grenoble Alpes [2020-..], 2024. Français. NNT : 2024GRALT021 . tel-04626658

HAL Id: tel-04626658

<https://theses.hal.science/tel-04626658>

Submitted on 27 Jun 2024

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

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : EEATS - Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)

Spécialité : Génie électrique

Unité de recherche : Laboratoire de Génie Electrique

Apport des mesures des compteurs Linky pour la connaissance des charges du réseau de distribution

Contribution of Linky meter measurements to the knowledge of the loads on distribution grid

Présentée par :

Guilherme RAMOS MILIS

Direction de thèse :

Raphael CAIRE

MAITRE DE CONFERENCES, Université Grenoble Alpes

Directeur de thèse

Marie-Cécile ALVAREZ-HERAULT

MAITRE DE CONFERENCES, Université Grenoble Alpes

Co-directrice de thèse

Rapporteurs :

Bruno FRANÇOIS

PROFESSEUR DES UNIVERSITES, Ecole Centrale de Lille

Georges KARINIOTAKIS

PROFESSEURE DES UNIVERSITES, MINES PARIS PSL

Thèse soutenue publiquement le **8 mars 2024**, devant le jury composé de :

Adeline LECLERCQ-SAMSON,

PROFESSEURE DES UNIVERSITES, Université Grenoble Alpes

Présidente

Bruno FRANÇOIS,

PROFESSEUR DES UNIVERSITES, Ecole Centrale de Lille

Rapporteur

Georges KARINIOTAKIS,

PROFESSEURE DES UNIVERSITES, MINES PARIS PSL

Rapporteur

Arturo SUMAN-BRETAS,

DIRECTEUR DE RECHERCHE, CNRS DELEGATION ALPES

Examinateur

Invités :

Christophe Gay

INGENIEUR DOCTEUR, Enedis

Cecile Donde

INGENIEURE, Enedis



Table des matières

Remerciements	iv
Liste des Figures	v
Liste des Tableaux	vii
Liste des Acronymes.....	viii
1 Contexte de la thèse.....	1
1.1 Évolution des usages du réseau	2
1.2 Évolution digitale du réseau.....	10
1.3 Vision générale et contributions de la thèse.....	14
1.3.1 Contributions.....	15
1.3.2 Structure.....	16
2 Cartographie d'utilisation des données	17
2.1 Utilisation des données des compteurs intelligents par les différents acteurs mondiaux ...	18
2.1.1 GRD mondiaux.....	18
2.1.2 Thématiques de recherche scientifique liés aux compteurs communicants.....	22
2.2 Cartographie de cas d'usage à partir des données des compteurs intelligents.....	36
2.3 Conclusion	39
3 Foisonnement des charges BT.....	41
3.1 Le foisonnement.....	42
3.1.1 Définition et importance du foisonnement	42
3.1.2 Historique du diversity factor.....	45
3.2 Périmètre des données	51
3.2.1 Construction des indicateurs.....	55
3.3 Modèle de prédiction	56
3.3.1 Modèle de forêt aléatoire	56
3.3.2 Analyse des indicateurs	60
3.3.3 Actualisation du modèle.....	63
3.4 Comparaison avec des méthodes existantes	65
3.5 Conclusions.....	69

4	Modélisation des charges BT à partir de la désagrégation	70
4.1	Méthode de désagrégation : concept et état de l'art	71
4.1.1	Revue Bibliographique	72
4.1.2	Méthode de désagrégation en 2 étapes	77
4.1.3	Données.....	78
4.2	Etape 1 : Segmentation des clients	79
4.2.1	Analyse en Composantes Principales	80
4.2.2	Méthodes de Segmentation et choix de K nombre de clients similaires.....	81
4.3	Etape 2 : Désagrégation	84
4.3.1	Modèle purement mathématique	86
4.3.2	Ajout de propriétés électriques au modèle purement mathématique.....	91
4.4	Analyse des résultats de l'application du modèle sur la totalité des postes à disposition .	101
4.5	Conclusions.....	107
5	Conclusion et Perspectives.....	109
5.1	Synthèse des travaux effectués.....	110
5.2	Perspectives.....	112
5.2.1	Court terme	112
5.2.2	Long terme	113
	Références.....	114
6	Annexes	127
6.1	Annexes 1 : Relatives au chapitre 3.....	127
6.1.1	Métriques d'évaluation des erreurs utilisées dans des modèles de prévision	127
6.1.2	Charte de chaleur avec tous les indicateurs.....	128
6.2	Annexes 2 : Relatives au chapitre 4.....	129
6.2.1	Les histogrammes des occurrences de puissances maximales et les points de maxima du Poste Exemple pour la saison d'hiver.....	129
6.2.2	Les erreurs du modèle 4 avec différentes valeurs de Lambda.....	131

Remerciements

Je voudrais commencer par remercier mes grands-parents Amauri et Ninita qui ne sont plus là pour voir ce moment. Merci beaucoup pour tout le soutien et l'affection que vous m'avez apportés au fil de ces années. Surtout pour m'avoir toujours encouragé à poursuivre mes rêves. J'espère que vous pouvez sourire en ce moment.

Je ne peux pas oublier de remercier certaines personnes qui ont fait partie de mon parcours ici en France. Merci Patrick de m'avoir fait confiance à l'époque pour participer au double diplôme UFSC-ENSE3. Un grand merci à Marie-Cécile de m'avoir donné l'opportunité de faire un stage au G2Elab et de m'avoir recommandé pour le stage de PFE chez Enedis. C'est à partir de ce PFE que mon projet de thèse a été développé. Merci Thomas de ne pas avoir ménagé d'efforts pendant la période de Covid-19 pour que le projet de thèse soit validé et que je puisse commencer cette aventure.

À la fin de ces 3 années de thèse, je voudrais remercier le G2Elab et Innovative Analytics (Enedis) de m'avoir accueilli. Un merci spécial à mes directeurs de thèse Raphael et Marie-Cécile, ainsi qu'à mon encadrant-Enedis Christophe. Merci pour les excellents conseils, les orientations et la patience à corriger mes erreurs de français.

Un merci à mes amis Gabriel et Vinicius qui ont partagé avec moi de bons rires et de bonnes bières tout au long de ces 3 années.

Un grand merci à ma famille qui, même de l'autre côté de l'océan, m'a toujours apporté un soutien total. Mes chers parents, Renato et Rossana. Mon frère, Renatinho. Merci beaucoup pour tout !

Enfin, un merci très spécial à mon amour, Mylena. Merci beaucoup d'être ma partenaire de vie ! D'être avec moi à tous les moments. De me donner tout le soutien et la motivation pour sourire. Je t'aime !

Liste des Figures

Figure 1-1 : Négociations internationales sur les politiques climatiques.....	2
Figure 1-2 : Production d'électricité par filière en France en 2012 (Total de 537 TWh) et 2019 (Total de 534 TWh). La production par la filière nucléaire est respectivement de 75,2 % en 2012 et 71 % en 2019 et n'est pas représentée ici.	4
Figure 1-3 : Chemin « classique » de l'électricité entre production et consommation – figure extraite d'une source interne Enedis.....	5
Figure 1-4 : Émissions de gaz à effet serre par activité en France entre 2008 (total de 526,1 Mt d'équivalent CO ₂) et 2021 (total de 418,2 Mt d'équivalent CO ₂)	6
Figure 1-5 : Répartition de la consommation énergétique résidentielle par usage en France en 2010 (Total de 528 TWh PCI) et en 2021 (Total de 486 TWh PCI).....	8
Figure 1-6 : Répartition des chauffages résidentiels selon le vecteur énergétique (où chauff. élec. est une abréviation de chauffage électrique).....	8
Figure 1-7 : Représentation schématique du système Linky	12
Figure 2-1 : Requête recherchée dans WoS	22
Figure 2-2 : Nombre de publications trouvées par WoS pour la requête TS	23
Figure 2-3 : Division des articles par thème	24
Figure 2-4 : Cartographie de cas d'usages des données des compteurs intelligents.....	38
Figure 3-1 : Illustration d'une colonne horizontale	44
Figure 3-2 : La figure extraite de l'article [135] et traduite en français, qui illustre la tendance probable en relation entre le Coincidence Factor de groupe et le nombre de consommateurs	47
Figure 3-3 : La figure extraite de l'article [136] et traduite en français, qui illustre l'ajustement de courbe pour tracer le DF en fonction uniquement du nombre de clients	48
Figure 3-4 : Valeurs de DF dans la période pour les deux poste choisis au hasard – Vision divisée par jour de la semaine (Jour ouvré ou jour weeked)	54
Figure 3-5 : Valeurs de DF dans la période pour les deux poste choisis au hasard – Vision divisée par saison.....	54
Figure 3-6 : Diagramme de flux des étapes.....	56
Figure 3-7 : Illustration de le modèle de forêt aléatoire	57
Figure 3-8 : Nuage de points – DF prédit x DF réel.....	59
Figure 3-9 : Carte de chaleur des indicateurs des catégories 1 à 4.....	60
Figure 3-10 : Illustatrion de la impureté de Gini	62
Figure 3-11 : Contributions de la température moyenne pour les variations du DF	64
Figure 3-12 : Contribution des clients HP/HC pour les variations de DF.....	65
Figure 3-13: Comparaison des résultats pour les 34 postes HTA/BT	67
Figure 3-14 : Histogramme Moyenne de SMAPE par jeu de test.....	68
Figure 4-1 : Exemple d'obtention de la courbe de charge d'un poste avec le modèle BT actuel	71
Figure 4-2 : Exemple d'obtention des courbes à la maille client à partir du modèle de désagrégation.....	72
Figure 4-3 : Carte représentant les localisations des postes HTA/BT choisis pour l'étude.....	79

Figure 4-4 : Exemple du fonctionnement du clustering Kmeans	83
Figure 4-5 : L'application de la méthode du Coude pour le Poste Exemple	84
Figure 4-6 : Format de la courbe des profils résidentiels tarif de base et des profils professionnels tarif de base	86
Figure 4-7 : Mesure Pinky du Poste Exemple pour une journée au hasard	87
Figure 4-8 : Résultat de l'ajustment de courbe avec l'équation 4-5 – courbes par cluster	88
Figure 4-9 : Comparaison entre le résultat de l'ajustement de courbe avec l'équation 4-5 et la courbe Pinky du poste Exemple dans la journée d'hiver	89
Figure 4-10 : Résultat de l'ajustment de courbe avec l'équation 4-6 – courbes par cluster	90
Figure 4-11 : Comparaison entre le résultat de l'ajustement de courbe avec l'équation 4-6 et la courbe Pinky du poste Exemple dans la journée d'hiver	90
Figure 4-12 : Occurrences de puissance maximale de 5 clients au cours de la saison	92
Figure 4-13 : Occurrences de la puissance maximale par les clients des cluster 0 et 5	92
Figure 4-14 : Points de maxima de l'histogramme des occurrences de puissance maximale	94
Figure 4-15 : Comparaison entre le résultat de l'ajustement de courbe avec l'équation 4-7 et la courbe Pinky du poste Exemple dans la journée avec le plus faible erreur SMAPE de 1,6 %	96
Figure 4-16 : Courbes résultants de la désagrégation par cluster au jour où le SMAPE est minimal ...	97
Figure 4-17 : Nombre de postes par tranche d'erreur	102
Figure 4-18 : Histogramme du nombre de clients pour les deux groupes des postes.....	103
Figure 4-19 : Histogramme du nombre de clients résidence secondaire pour les deux groupes des postes	104
Figure 4-20 : Histogramme de l'énergie moyenne pour les deux groupes des postes.....	104
Figure 4-21 : Histogramme des valeurs maximales en module de la deuxième dérivée pour les deux groupes des postes.....	105
Figure 4-22 : Courbe de charge d'une journée au hasard pour une poste à forte erreur	106
Figure 6-1 : Charte de chaleur	128
Figure 6-2 : Occurances de la puissance maximale pour le clients du Poste Exemple dans la saison d'hiver.....	129
Figure 6-3 : Points de maxima des clusters du Poste Exemple dans la saison d'hiver.....	130

Liste des Tableaux

Tableau 1-1 : Parc de production raccordée au RPD Enedis en nombre de producteurs.....	5
Tableau 1-2 : Nombre d'infrastructures de recharge de VE raccordées au RPD Enedis.....	7
Tableau 1-3 : Principaux démonstrateurs des réseaux intelligents en France	13
Tableau 1-4 : Démonstrateurs des réseaux intelligents dans le monde	14
Tableau 2-1 : Classement des références suivant les thematiques correspondantes	37
Tableau 3-1 : Indicateurs	55
Tableau 3-2 : Indicateurs retenus après analyse.....	62
Tableau 3-3 : Résultats des modèles.....	63
Tableau 3-4 : Valeurs de CP par tranche de nombre de clients	66
Tableau 4-1 : Résultats de la segmentation pour le Poste Exemple avec K = 6.....	84
Tableau 4-2 : Erreurs SMAPE d'ajustement de courbes pour la saison d'hiver pour les trois modèles proposés	95
Tableau 4-3 : Nouvelles erreurs SMAPE de l'ajustement de courbes pour la saison d'hiver.....	100
Tableau 4-4 : Les erreurs SMAPE des énergies du modèle 4	100
Tableau 4-5 : Moyenne des SMAPE pour les 48 postes HTA/BT	101
Tableau 4-6 : Forces et faiblesses des modèles de désagregation	106
Tableau 6-1 : Métriques d'évaluation des erreurs	127
Tableau 6-2 : Les erreurs du modèle 4 pour le Poste Exemple avec différents valeurs de lamda	131

Liste des Acronymes

ACB : Analyse Coûts-Bénéfices

BT : Basse Tension (avec $U_n = 400 \text{ V}$)

BTSUP : Basse Tension Supérieure

CCMA : Chaîne Communicante Marché d’Affaires

CPL : Courant Porteur en Ligne

DF : « Diversity Factor »

ENR : Énergies Renouvelables

GRD : Gestionnaire du Réseau de Distribution

GRT : Gestionnaire du Réseau de Transport

HC : Heure Creuse

HP : Heure Pleine

HTB : Haute Tension de niveau B (avec $20 \text{ kV} < U_n < 400 \text{ kV}$)

HTA : Haute Tension de niveau A (avec $1 \text{ kV} < U_n < 20 \text{ kV}$)

INSEE : Institut National de la Statistique et des Études Économiques

IRVE : Infrastructure de Recharge de Véhicules Électriques

NILM : « Non-intrusive Load Monitoring »

PCI : Pouvoir Calorifique Inférieur

PV : Photovoltaïque

RGPD : Règlement Général sur la Protection des Données

RPD : Réseau Public de Distribution

SI : Système d’Information

ToU : « Time-of-use »

TURPE : Tarif d’Utilisation des Réseaux Publics d’Électricité

VE : Véhicules Électriques

V2G : « Vehicle-To-Grid »

1 Contexte de la thèse

Ce chapitre présente l'évolution des usages du réseau, en plus de comprendre la digitalisation qu'il traverse. Les contributions scientifiques de ce manuscrit ainsi que sa structure y sont décrites.

1.1 Évolution des usages du réseau

L'être humain commence graduellement à accepter que ses actions aient des impacts sur notre planète. Sa quête incessante de l'accumulation de richesses et d'amélioration de la qualité de vie l'a conduit à l'exploitation excessive de ressources naturelles (en particulier les combustibles fossiles) sans intégrer les conséquences que cela pourrait entraîner. Et aujourd'hui, de nombreux rapports scientifiques démontrent que ces impacts sont directement liés à l'augmentation des émissions de CO₂ ainsi qu'à d'autres conséquences environnementales [1-2]. Ces impacts environnementaux deviennent si impactant pour nos sociétés modernes qu'en 1987 le concept de développement durable a été créé [3]. En 1992, la Convention Cadre des Nations Unies sur les Changements Climatiques, premier traité international visant à prévenir les impacts sur le climat, a été signée. Depuis cette date, les pays membres des Nations Unies se réunissent régulièrement pour discuter des moyens d'atténuer nos impacts sur l'environnement, comme l'illustre partiellement la Figure 1-1 [4] :

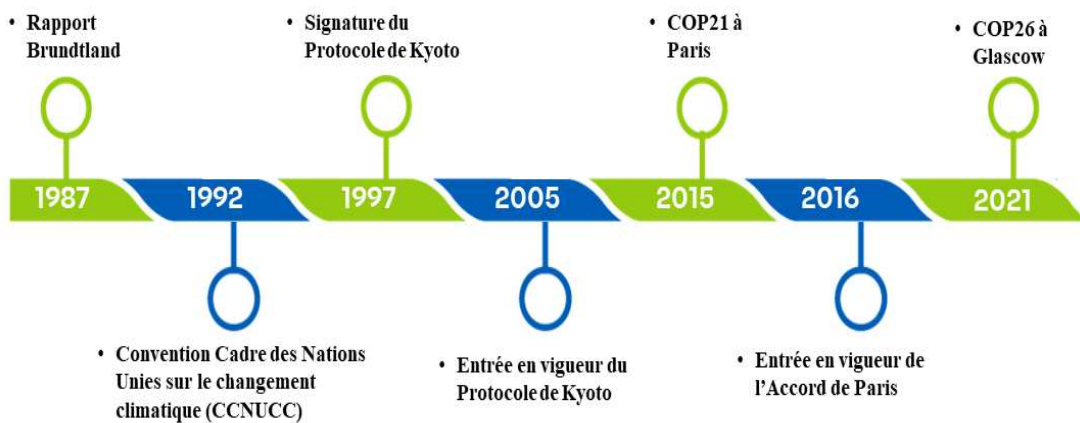


Figure 1-1 : Négociations internationales sur les politiques climatiques Figure 1-1

Une des manières d'atténuer nos impacts environnementaux passe par la réduction et, encore mieux, la non-utilisation des combustibles fossiles. Ceci peut être atteint partiellement grâce au système électrique. Selon [5], en 2022, 40 % des émissions mondiales de dioxyde de carbone proviennent de la production totale d'électricité et 73 % de ces émissions proviennent des centrales à charbon. Changer la façon dont notre énergie électrique est produite est une partie fondamentale de la réduction de nos impacts environnementaux. Cela tend à être de plus en plus impactant avec l'augmentation de la consommation d'électricité. Entre les années 1974 et 2019, la consommation finale d'énergie électrique est passée de 5268 TWh à 22847 TWh [6] et tend à augmenter encore davantage. Comme présenté dans le

rapport RTE sur le futur énergétique en 2050 en France [7], la part de la consommation finale d'énergie liée à l'électricité devrait passer de 25 % en 2021 à 55 % en 2050.

Ces changements de production et consommation de l'électricité entraînent des changements dans la manière dont le réseau de distribution d'électricité est utilisé. Ces changements vont de la répartition de flux entre la génération et la distribution, jusqu'à la façon dont les clients utilisent cette énergie et les équipements qui y sont connectés. Ces mutations des réseaux de distribution font partie d'un phénomène plus large appelé « transition énergétique » [8]. Pour mieux comprendre ces changements dans l'utilisation du réseau de distribution en général, nous allons nous focaliser sur l'exemple du réseau de distribution français.

Scénario Français

En France, il est possible d'observer un fort engagement vis-à-vis la transition écologique et surtout énergétique, grâce aux lois Grenelle 1 [9] et 2 [10] promulguées respectivement en 2008 et 2010, ainsi qu'à la loi de transition énergétique pour la croissance verte [11] de 2015. Parmi les objectifs de ces lois, nous pouvons souligner : la promotion du développement des énergies renouvelables et la réduction des émissions de gaz à effet de serre dans les différents secteurs économiques. Examinons en détail comment ces deux points ont évolué au fil des ans et quels impacts ils ont eu sur le réseau de distribution français géré par Enedis. Enedis est le gestionnaire du Réseau Public de Distribution (RPD) sur 95 % du territoire métropolitain continental. Il développe, exploite et modernise ce réseau électrique et en gère les données associées. Ce réseau compte 1,4 million de kilomètres de circuits de Haute Tension de niveau A (HTA, niveaux de tension entre phase U_n comprises entre 1 et 50 kV) et de Basse tension de niveau A (BTA – nous utiliserons par la suite BT, niveaux de tension entre phase U_n comprises entre 50 et 500 V) pour environ 37 millions de clients (particuliers, professionnels et industriels). Enedis exerce, dans les conditions fixées par la loi et les contrats de concession signés avec chacune des autorités concédantes (collectivités locales), les missions de gestion du réseau public de distribution.

Dans le cadre de la promotion des énergies renouvelables, l'industrie de l'énergie subit, quant à elle, une transformation majeure grâce à l'augmentation de l'intégration de la production d'énergies renouvelables. Nous pouvons observer ce phénomène grâce aux données de RTE sur la production d'énergie par filière [12], présentées dans la Figure 1-2.

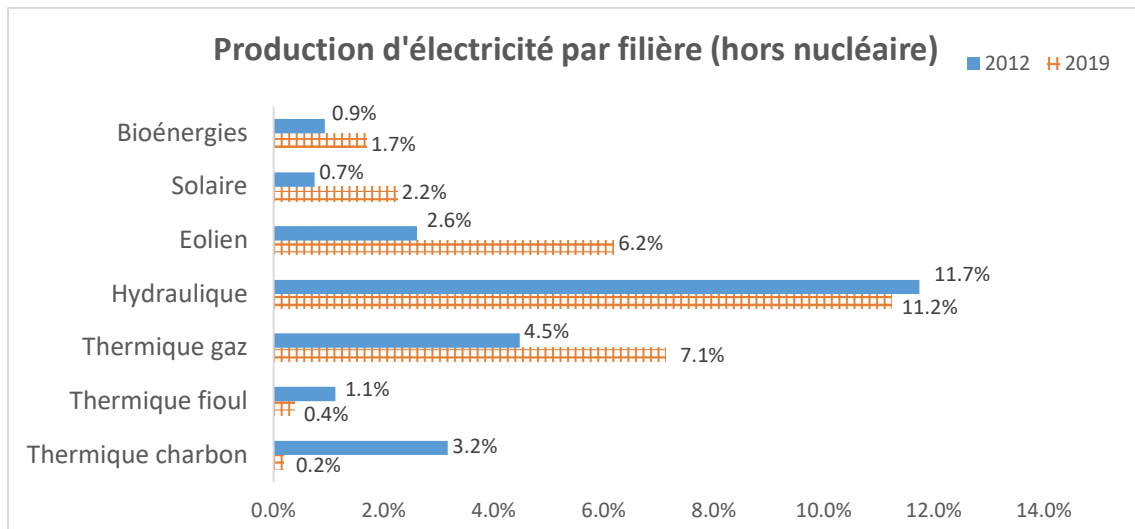


Figure 1-2 : Production d'électricité par filière en France en 2012 (Total de 537 TWh) et 2019 (Total de 534 TWh). La production par la filière nucléaire est respectivement de 75,2 % en 2012 et 71 % en 2019 et n'est pas représentée ici.

Entre 2012 et 2020, il y a eu une légère réduction de la production de la filière thermique (charbon et fioul) et une augmentation de la production thermique au gaz. Respectivement, ces sources de production d'énergie émettent 820 grammes et 490 grammes de CO₂ par kWh produit. Nous remarquons également une augmentation significative des filières renouvelables, notamment la production éolienne qui émet 11 grammes de CO₂ par kWh produit. Ces chiffres illustrent comment un changement dans les parcs de production d'électricité joue un rôle fondamental dans la réduction des émissions de gaz à effet de serre. Cette transformation impacte le chemin que l'énergie parcourt de son lieu de production à son lieu de consommation. Quand nous pensons au chemin de l'électricité depuis la production jusqu'à son utilisation par le client final, nous avons classiquement à l'esprit le schéma de la Figure 1-3.

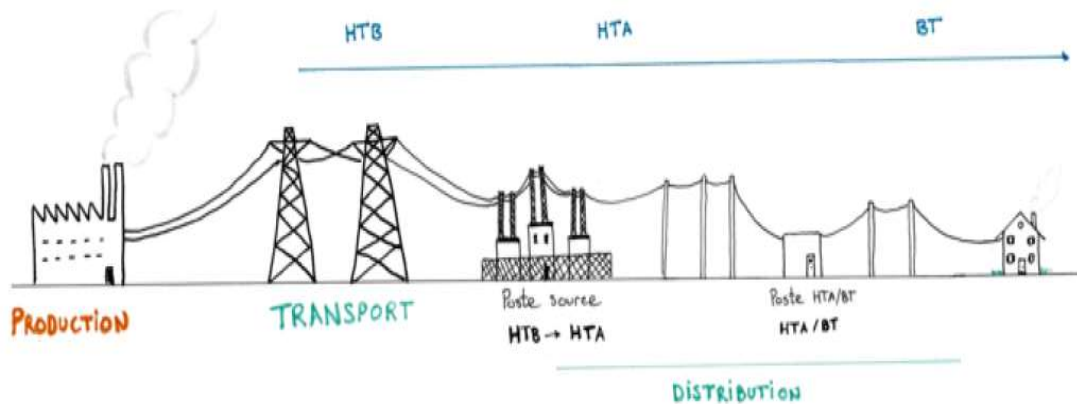


Figure 1-3 : Chemin « classique » de l'électricité entre production et consommation – figure extraite d'une source interne Enedis

Cependant, la filière des énergies renouvelables a profondément remis en question ce chemin classique pour diverses raisons, comme les économies sur les pertes d'énergie dans les réseaux de transmission et de distribution [154]. De cette manière, une partie de la production d'énergie est raccordée directement au RPD. Nous pouvons observer cela à travers les données sur l'évolution du parc de production raccordé au RPD géré par Enedis [13], en nombre de producteurs, qui sont représentées dans le Tableau 1-1.

Filière	Année 2012 (T1)	Année 2021 (T1)	Année 2023 (T3)
Photovoltaïque	238 311	486 213	779 979
Éolien	982	1838	2120
Hydraulique	1792	2317	2401
Biomasse	254	954	1116
Cogénération	537	860	915
Stockage	0	49	316
Autres	71	183	178
Total	241 947	492 414	787 025

Tableau 1-1 : Parc de production raccordée au RPD Enedis en nombre de producteurs

La Tableau 1-1 nous montre que le parc de productions raccordées au RPD a augmenté d'un facteur 3 entre le premier trimestre de 2012 et le premier trimestre de 2023. Nous pouvons observer que la filière qui a connu la plus forte croissance est la filière photovoltaïque, pour laquelle la grande majorité des installations sont inférieures à 36 kW et donc logiquement raccordées en BT [13]. En revanche, la filière éolienne, qui a doublé son nombre d'installations sur la période, est presque entièrement raccordée au réseau HTA [13]. Cette croissance va

dans le sens de la croissance de la production renouvelable présentée précédemment par la Figure 1-2. Cette nouvelle dynamique entre la production et la distribution d'électricité engendre de nouvelles contraintes pour le réseau de distribution, telles que par exemple des variations de tension haute et la bidirectionnalité du flux d'électricité [14]. Ces contraintes n'étaient pas anticipées lors du développement et de la construction du réseau de distribution. Cela signifie que les responsables de la planification et de l'exploitation du réseau doivent trouver des solutions pour éviter et gérer ces contraintes.

Cette analyse de développement massif des énergies renouvelables nous a montré la nécessaire évolution du réseau électrique de distribution. Maintenant, regardons la réduction des émissions de gaz à effet de serre dans les autres secteurs d'activité. Les impacts causés par les lois citées précédemment peuvent être observés à travers les données d'émission de gaz à effet de serre mises à disposition par l'Institut National de la Statistique et des Études Économiques (L'INSEE) et présentées dans la Figure 1-4 [15]. Il est possible de constater qu'entre 2008, date de promulgation de la loi Grenelle 1, et 2021, données les plus récentes disponibles, nous pouvons observer une réduction globale de 20,5 % des émissions de gaz à effet de serre.

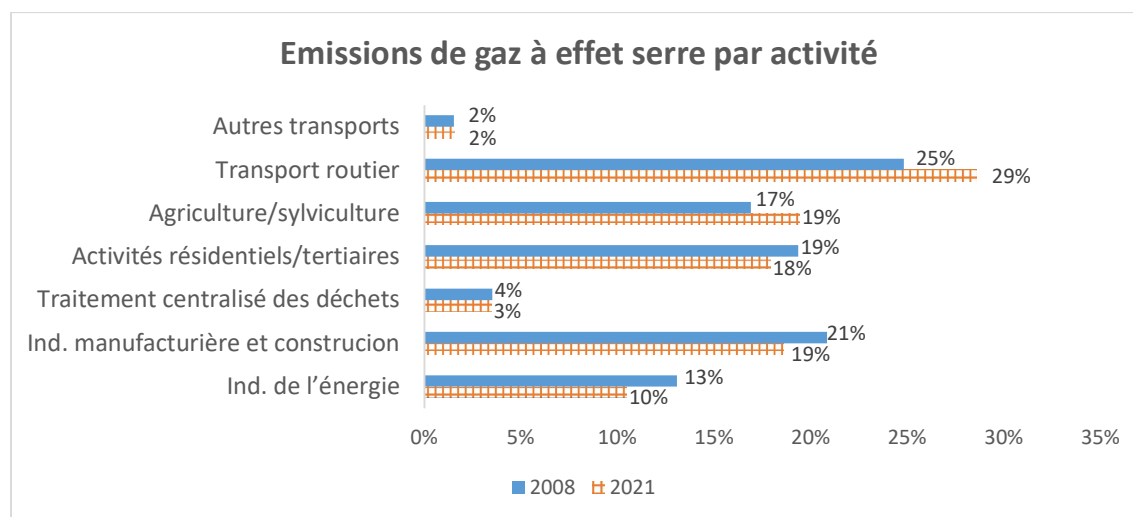


Figure 1-4 : Émissions de gaz à effet serre par activité en France entre 2008 (total de 526,1 Mt d'équivalent CO2) et 2021 (total de 418,2 Mt d'équivalent CO2)

Même avec une augmentation de 4 % sur la période par rapport aux autres secteurs d'activité, le transport routier a subi une réduction de 10,9 millions de tonnes d'équivalent CO2. Le secteur des activités résidentielles/tertiaires a été le secteur à subir la plus grande réduction

des émissions de gaz à effet de serre, avec 26,9 millions de tonnes d'équivalent CO2 entre 2008 et 2021. Concentrons-nous maintenant sur le secteur des transports. Le transport routier a connu une réduction, entre autres, grâce à la popularisation et à l'utilisation des véhicules 100 % électriques et hybrides. Le rapport de l'association nationale pour le développement de la mobilité électrique, Avere [16], montre l'évolution des immatriculations sur le parc de véhicules électriques français. En 2010, les français comptaient 102 véhicules électriques immatriculés, et en 2021, ce nombre est passé à 756 000 (sur un total d'un peu moins de 39 millions). Cette croissance peut également être observée du point de vue du GRD Enedis, à partir des données de l'évolution du nombre d'infrastructures de recharge de VE (2015, 2021 et 2023). Cela est représenté dans le Tableau 1-2.

Lieu de l'infrastructure	Année 2015 (T1)	Année 2021 (T1)	Année 2023 (T3)
En société	34 746	269 327	652 131
Particulier	25 638	307 174	906 942
Accessibles au public	8 478	37 491	109 856
Total	68 862	613 992	1 668 929

Tableau 1-2 : Nombre d'infrastructures de recharge de VE raccordées au RPD Enedis

Le Tableau 1-2 nous montre que le parc des infrastructures de recharge raccordées au RPD a augmenté d'un facteur 9 entre le premier trimestre de 2015 et le premier trimestre de 2021. La recharge des VE n'était pas une charge du réseau envisagée à l'époque du dimensionnement du RPD. En plus d'augmenter la charge du réseau électrique, les VE introduisent une nouvelle dynamique grâce aux perspectives apportées par le « vehicle-to-grid » (V2G). Le V2G est la possibilité d'utiliser les VE comme sources d'énergie pour le réseau. Cela peut apporter de nombreux avantages, mais cela représente également un défi pour trouver un cadre applicable local ou globalement, car la batterie d'un véhicule électrique est une ressource très petite qui, individuellement, a un impact pratiquement négligeable sur le réseau de distribution [17]. En plus, la majorité de ces points de recharge se trouve chez les particuliers, ce qui représente une évolution dans les habitudes de consommation traditionnelles des clients. Les véhicules électriques ne sont pas le seul changement dans la consommation électrique. Pour approfondir davantage les détails des changements de consommation traditionnelle des clients, nous pouvons examiner les données du Ministère de la Transition Écologique et de la Cohésion des Territoires sur la répartition de la consommation d'énergie par type d'utilisation [18]. Ils sont présentés dans la Figure 1-5 : Répartition de la consommation énergétique résidentielle par usage en France en 2010 (Total de 528 TWh PCI) et en 2021 (Total de 486 TWh PCI).

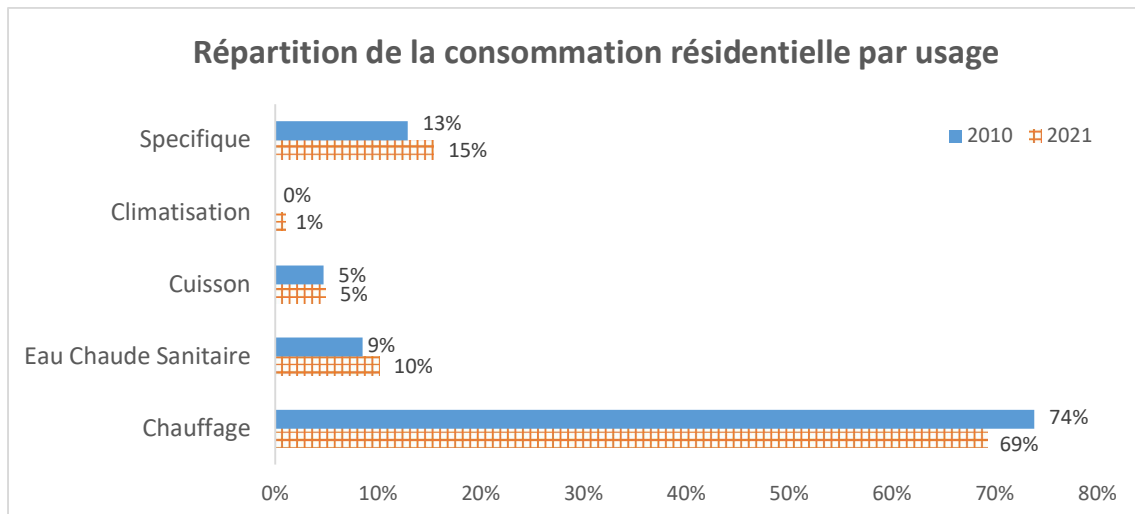


Figure 1-5 : Répartition de la consommation énergétique résidentielle par usage en France en 2010 (Total de 528 TWh PCI) et en 2021 (Total de 486 TWh PCI)

Ces données nous montrent que le chauffage est la principale source de consommation d'énergie dans les résidences françaises. Comme nous pouvons le voir à travers la Figure 1-5, de 2010 à 2021, le chauffage représente environ 70 % de la consommation. Même en ayant diminué de 5 % au cours de cette période, elle reste toujours l'usage résidentiel dominant. Le Ministère de la Transition Écologique fournit également des données sur la répartition de la source d'énergie utilisée pour chaque usage résidentiel [18] comme le montre la figure 6. Nous pouvons alors voir si cela est cohérent avec la réduction des émissions de gaz à l'effet serre du secteur résidentiel/tertiaire précédemment montrée.

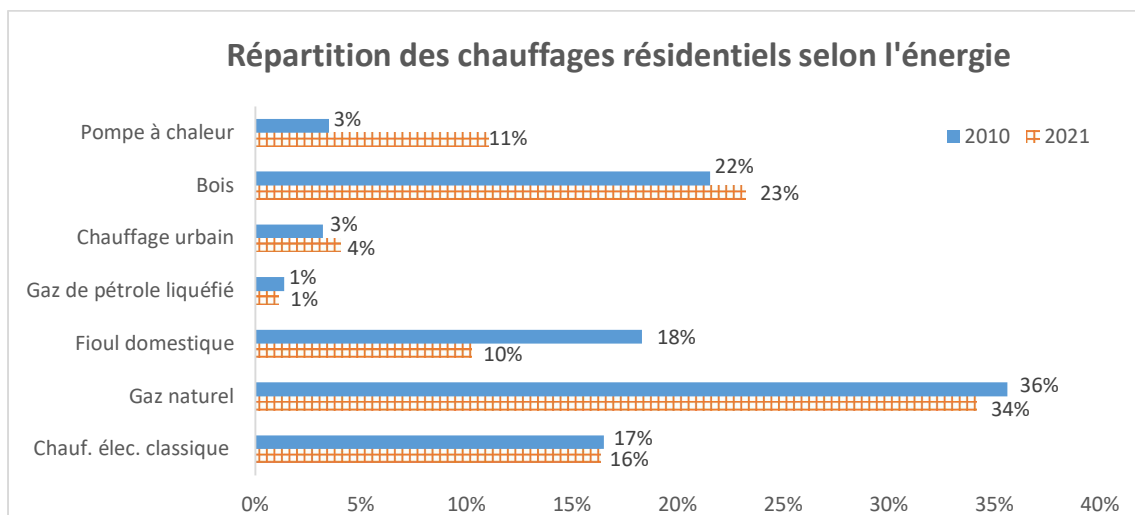


Figure 1-6 : Répartition des chauffages résidentiels selon le vecteur énergétique (où chauf. élec. est une abréviation de chauffage électrique)

On peut observer que de 2010 à 2021, il y a une réduction de l'utilisation des combustibles fossile (fioul domestique et gaz naturel) et l'augmentation de l'utilisation de la pompe à chaleur. Il y a encore beaucoup de marge pour réduire l'utilisation de sources qui contribuent aux émissions de gaz à effet de serre. Nous pouvons donc imaginer que le nombre de résidences avec un chauffage électrique ou avec pompe à chaleur tend à augmenter pour lutter contre ces émissions. Pour le GRD, cela signifie une augmentation de la consommation des clients résidentiels, ce qui pourrait entraîner la nécessité d'adapter les modèles de consommation et de mettre en place des mécanismes pour éviter des congestions au réseau.

Nous pouvons donc conclure que la lutte contre les impacts environnementaux induits par notre recherche d'amélioration du confort et de la qualité de vie, passe par un changement important dans les habitudes de production et consommation d'électricité. Les évolutions en France de cette dernière décennie nous ont permis de visualiser ces changements. Elles nous ont montré une accélération de la production d'énergie renouvelable, dont une partie est directement raccordée au RPD à différents niveaux de tension. Nous avons également constaté des changements dans les habitudes de consommation. Le nombre croissant de véhicules électriques en circulation au cours des dernières années constitue une charge supplémentaire nécessaire pour la décarbonation de notre mobilité, mais qui peut également offrir des services au réseau. De plus, nous pouvons voir que les consommateurs français se chauffent de manière plus respectueuse de l'environnement, ce qui implique directement ou indirectement une augmentation de la consommation d'électricité. Cela ne se produit pas exclusivement en France. Chaque pays, avec son rythme et ses particularités, tend à traverser ces changements dans sa quête pour lutter contre le changement climatique et réduire son impact environnemental. Ce qui rend cette discussion pertinente au-delà du contexte français, c'est qu'elle aborde des questions et des défis qui sont universels. Nous pouvons constater tout au long de cette section 1.1 que ces changements dans les habitudes de consommation et l'intégration des énergies renouvelables se produisent majoritairement à la maille du RPD. Nous pouvons donc imaginer qu'avoir un réseau de distribution solide, adaptable et stable joue un rôle essentiel dans l'accélération de ces changements, réduisant ainsi notre impact environnemental. Le RPD lui-même a connu des évolutions, et c'est ce dont nous allons discuter par la suite.

1.2 Évolution digitale du réseau

À mesure que les usages du réseau évoluent, le RPD évolue également pour accompagner ces changements. Avec de nouveaux types de consommation et un flux d'énergie devenu bidirectionnel du fait des énergies renouvelables connectées au RPD, le réseau doit être capable d'en optimiser son utilisation pour en assurer son bon fonctionnement. Cela passe par une bonne gestion de l'offre d'énergie (producteurs, acteurs du marché) et de sa demande (consommateurs), à différentes échelles temporelles. Grâce au développement des technologies de l'information et communication, les réseaux électriques peuvent devenir plus observables et pilotables, transformant notre réseau « classique » en un réseau « intelligent ». Selon l'Agence Internationale de l'Energie [19], « *les réseaux intelligents sont des réseaux électriques qui utilisent des technologies numériques, des capteurs et des logiciels pour mieux faire correspondre l'offre et la demande d'électricité en temps réel tout en minimisant les coûts et en maintenant la stabilité et la fiabilité du réseau* ». Nous pouvons tirer de cette définition que la digitalisation du réseau est une partie essentielle de la transition des réseaux intelligents depuis les réseaux classiques. Le pilier de cette digitalisation passe par le remplacement des compteurs d'électricité, passant des compteurs analogiques/numériques aux compteurs communicants. Les compteurs communicants sont des dispositifs électroniques qui enregistrent des informations telles que la consommation d'énergie électrique, puis communiquent certaines de ces informations au consommateur, aux fournisseurs d'électricité et au GRD. L'utilisation de ces informations rend possible la considération de qui consomme quand et donc quoi, permettant un lien plus direct entre le consommateur et les sources d'énergies sollicitées au travers des marchés de l'énergie. Les débuts de l'utilisation des compteurs communicants dans le réseau électrique ont eu lieu en Europe et est lié à la Directive 96/92/CE [20], concernant les règles communes pour le marché intérieur de l'électricité. La libéralisation du marché de l'électricité a ouvert de nouveaux défis où les entreprises doivent faire face à un marché concurrentiel. Un nouveau modèle tarifaire centré sur le client exige un système de mesure capable de fournir des informations sur la consommation réelle, à des échelles de temps quotidiennes ou plus courtes, pour tous les acteurs du marché (des clients aux producteurs). La tarification de l'électricité est ainsi transformée, en la basant non plus sur des estimations, ce qui apporte une plus grande dynamique aux prix. L'Italie a été le premier pays européen à investir dans le changement de son parc de compteurs. En 2001, ENEL a commencé le déploiement de la première génération de compteurs communicants à partir du projet *Telegestore* [21]. La Suède est parmi les premiers pays européens à adopter l'utilisation des compteurs communicants. En 2003, le parlement suédois a imposé que, avant 2009, tous les clients devraient recevoir leurs factures d'électricité fondées sur leur consommation réelle et non sur une estimation. C'est ainsi que

la première génération de compteurs communicants a été déployée [22]. Nous allons entrer davantage dans les détails des compteurs communicants Linky d'Enedis, car ils sont la source des données du réseau utilisées dans la suite de la thèse.

Déploiement des compteurs communicants sur le RPD Enedis

Le Programme de modernisation du parc de comptage français (Linky) a été lancé en 2006, par décision du président d'EDF et suite à l'injonction de l'Union Européenne [23]. Il consistait à remplacer les anciens compteurs BT, dont la puissance souscrite est inférieure ou égale à 36 kW, par une nouvelle génération de compteurs communicants permettant de les relever et de les opérer à distance (un exemple de télé-opération du compteur est le changement de la puissance souscrite). Les objectifs, derrière cette action, étaient de fournir aux clients une relève d'énergie précise et de réduire les coûts grâce à la relève à distance et la télé-opération. En outre, les compteurs communicants contribuent à la transition énergétique, car l'exploitation de ses données peuvent faciliter l'insertion des nouveaux usages et des ENR. Ce projet de déploiement a été divisé en trois grandes phases :

- Une première phase d'expérimentation a eu lieu entre 2008 et 2011 et comprenait la construction du SI (système d'information) Linky et le déploiement de 300 000 compteurs et 5 000 concentrateurs sur les zones géographiques tests ;
- Une deuxième phase de pré-généralisation a eu lieu entre 2012 et 2015 et visait à valider l'entrée dans une phase d'industrialisation du déploiement (signature des marchés avec les constructeurs, validation de l'investissement de la première tranche du déploiement massif) ;
- Une dernière phase de généralisation, qui a débuté le 1er décembre 2015, avait comme objectifs le déploiement de 3 millions de compteurs jusqu'en 2016, puis le déploiement massif du reste des compteurs.

En 2023, plus de 37 millions de compteurs Linky ont été déployés. À l'état actuel, le déploiement massif des compteurs est déjà terminé. Certains clients n'ont pas de compteur Linky pour deux raisons. La première est due à des problématiques où le CPL ne passe pas correctement dans la zone où se trouve le client. La deuxième concerne les clients qui refusent d'avoir un compteur Linky et préfèrent payer les coûts de la relève manuelle. En regardant de plus près le compteur Linky, il est conçu à partir des fonctionnalités du « compteur bleu électronique » et est enrichi de 7 attributs supplémentaires majeurs :

- Une horloge ;
- Un interrupteur ;

- Un logiciel évolué ;
- Un modem CPL ;
- Un module de chiffrement ;
- Huit contacteurs virtuels pour le piloter à distance ;
- Un module radio.

L'architecture de communication du système Linky est représentée par la Figure 1-7 :

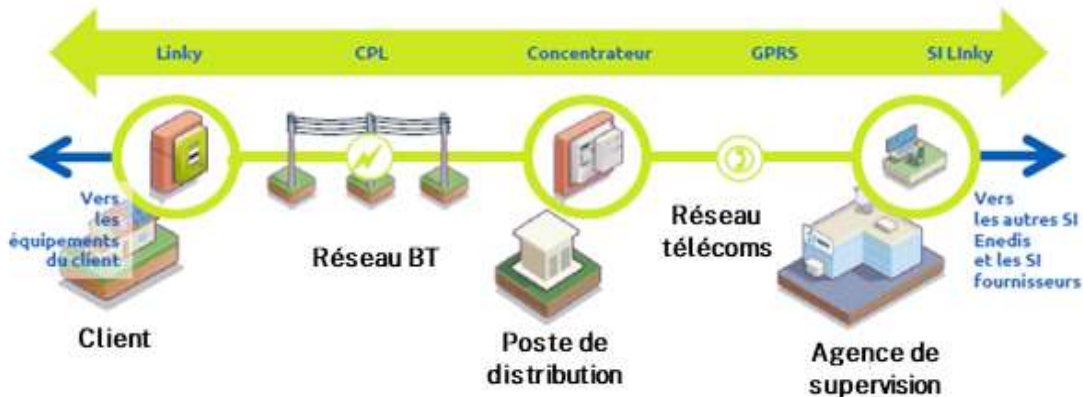


Figure 1-7 : Représentation schématique du système Linky

La Figure 1-7 montre la chaîne de transmission de l'information depuis le compteur jusqu'au SI Linky. La communication est assurée en Courant Porteurs de Ligne (CPL) entre le concentrateur situé dans le poste de distribution publique (avec le transformateur HTA/BT) et le compteur. La communication entre le poste et le SI Linky est faite par GPRS. Le système Linky est bidirectionnel. C'est-à-dire que le compteur peut recevoir des ordres en provenance du SI Linky mais peut aussi envoyer des données sans intervention physique d'un technicien. Cependant, le SI Linky n'a pas été développé dans le but de réaliser des opérations de pilotage en temps réel. Cela implique que le gestionnaire du réseau de distribution doit planifier ces opérations avec un certain délai. Le compteur Linky mesure :

- Les index de consommation et/ou de production journaliers. Avec l'accord de ses clients, Enedis peut relever la puissance de ces derniers avec un pas de temps fin (de 10 à 30 minutes) permettant ainsi d'obtenir une courbe de charge plus précise, sur des pas horaires définis ;
- La puissance maximale journalière soutirée ou injectée avec l'horaire de l'évènement ;
- Les évènements de qualité de fourniture comme les excursions de tension (enregistrement des écarts de tensions au-delà de la plage contractuelle) et les coupures sont aussi enregistrés et transmis.

Parallèlement au SI Linky, le SI Chaîne Communicante Marché d’Affaires (CCMA) permet d’avoir les mêmes fonctionnalités pour des clients BTSUP. Concrètement, les clients BTSUP sont les clients dont la puissance souscrite est supérieure à 36 kVA. Il s’agit généralement des « petites » installations du segment industriel tels que, par exemple, les boulangeries, les lycées, les exploitations agricoles entre autres. Tous les clients du SI CCMA sont relevés en courbe de charge, avec un pas de temps fin (de 10 à 30 minutes).

Enedis a aussi récemment développé des centrales de mesure Pinky, qui sont des compteurs communicants à la maille du poste de distribution HTA/BT, côté BT. Pinky permettent l’acquisition des données sur la qualité de la tension, le courant et les puissances actives et réactives à la maille du poste HTA/BT sur l’ensemble des phases du réseau et à un pas de temps de 10 minutes. A la date de 2023, 1000 compteurs Pinky ont été déployés sur réseau Enedis.

Les données des compteurs communicants constituent le pilier de la construction de réseaux intelligents. Grâce aux données du compteur Linky, plusieurs projets de démonstrateurs de réseaux intelligents ont été développés en France au cours des 10 dernières années [24]. Les principaux projets sont synthétisés dans le Tableau 1-3 :

Démonstrateur	Lancement	Thématiques d’intérêt
Nice Grid	2011	Conduite des réseaux ; Intégration des ENR ; Maîtrise de la demande
SoGrid	2011	Conduite des réseaux ; Maîtrise de la demande
Smart Electric Lyon	2012	Maîtrise de la demande ; Maisons intelligentes ; VE
GreenLys	2012	Conduite des réseaux ; Intégration des ENR ; Maîtrise de la demande

Tableau 1-3 : Principaux démonstrateurs des réseaux intelligents en France

À partir de l’exemple français, nous pouvons constater à quel point le réseau de distribution classique évolue et se transforme en un réseau intelligent. Le remplacement des anciens compteurs par des compteurs communicants constitue le pilier de cette transformation. Les compteurs communicants permettent de collecter, dans le respect de la vie privée des clients (consommateurs, producteurs, consomm’acteur), les informations nécessaires au pilotage du

système. Pour étendre davantage notre vision sur l'importance de l'utilisation des données des compteurs intelligents pour le développement des réseaux intelligents, le Tableau 1-4, regroupant les informations contenues dans [149], présente les autres démonstrateurs dans le monde.

Démonstrateur	Pays	Lancement	Thématiques d'intérêt
CSIRO	Australie	2018	Intégration des ENR
LEAFS	Autriche	2015	Flexibilités de la demande
Charge the North	Canada	2016	VE
EnergyLab Nordhavn	Danemark	2015	Pilotage des réseaux
Decentralized FLISR	Etats-Unis	2016	Micro réseaux
Smart Otaniemi	Finlande	2018	Flexibilités
Smart Grid CESC	Inde	2014	Intégration des ENR ; Maitrise de la demande ; Conduite des réseaux
OSMOSE	Italie	2018	Flexibilités
FME CINELDI	Norvège	2017	Conduite des réseaux ; Micro réseaux ; Flexibilités
HIBREAK	Suède	2017	Haute tension courant continu (HVDC)
Moixa's ERIC projec	Royaume Uni	2015	Autoconsommation ; Stockage résidentiel

Tableau 1-4 : Démonstrateurs des réseaux intelligents dans le monde

1.3 Vision générale et contributions de la thèse

Tout au long de ce chapitre, nous avons observé que le réseau de distribution était en train de passer par de profonds changements. Nous avons pu mieux comprendre les transformations dans l'utilisation du réseau qui sont en cours et une partie de leurs raisons. Nous avons, non seulement vu quels sont ces changements, mais aussi comment ils sont perceptibles dans le réseau de distribution, à travers l'exemple du RPD géré par Enedis. Nous assistons à une prolifération de la production d'électricité décentralisée grâce aux sources d'énergie renouvelable, ce qui entraîne un flux bidirectionnel dans le réseau. Nous constatons un changement dans les habitudes de consommation des clients, fortement lié à la réduction des émissions de gaz à effet de serre. Cela se traduit par de nouveaux usages tels que les VE et l'utilisation d'équipements à faible émission de carbone (électriques) dans un contexte de l'électrification accrue. En parallèle de ces changements, nous avons un réseau de distribution

qui évolue en passant par des transformations visant à le rendre plus intelligent. Ces transformations, ayant comme pilier le changement du parc de compteurs, fournissent au distributeur de plus en plus d'informations sur le réseau. Comme cela a été vu dans l'exemple français, les compteurs communicants apportent de nouvelles données (qu'il s'agisse de mesures auxquelles on n'avait pas accès auparavant ou de la volumétrie de données) au GRD, ce qui lui permet d'innover et d'améliorer ses solutions de distribution. Dans ce contexte, ce manuscrit de thèse vise à discuter comment ces nouvelles données de mesure du réseau de distribution BT peuvent améliorer la connaissance des charges du réseau BT.

1.3.1 Contributions

La discussion de la thèse est centrée sur deux thèmes et apportera trois contributions scientifiques majeures. Ces deux thèmes sont les suivants : l'identification des différents domaines où les données des compteurs communicants peuvent apporter des avantages, et l'utilisation de ces données pour la modélisation des charges BT. Les contributions à ces thèmes seront les suivantes : C1 – Nouvelle cartographie des utilisations des données, C2 – Estimation et analyse du foisonnement des charges du réseau BT, C3 - Modélisation des courbes de charges des clients du réseau BT.

C1 – Nouvelle cartographie des utilisations des données

Avec les nouvelles données sur le réseau de distribution et une volumétrie accrue, se pose la question de leur utilisation. Cette question est valable pour tous les acteurs du secteur, des GRD aux chercheurs du domaine scientifique. Nous cherchons donc à y répondre et la synthétisons grâce à une cartographie. Celle-ci peut guider les GRD en phase de changement de leur parc de comptage pour utiliser les données afin de répondre aux enjeux de leur réseau. Elle peut également aider les chercheurs à trouver plus facilement des références pour leurs sujets de recherche.

C2 – Estimation et analyse du foisonnement des charges du réseau BT

Le coefficient de foisonnement nous indique à quel point la consommation d'un groupe de clients donné est hétérogène. Avec ces données, nous étudierons à quel point la consommation des clients des réseaux BT réels est diversifiée et quels sont les facteurs ayant le plus d'impact sur la variation de ce coefficient. De plus, nous développerons un modèle pour prédire ce coefficient et comparerons les résultats de ce modèle avec des méthodes

classiques. La bonne connaissance du foisonnement permet de mieux cibler les investissements, planifier l'évolution des réseaux et estimer sa capacité d'accueil.

C3 - Modélisation des courbes de charges des clients du réseau BT

Les données de comptage d'Enedis nous permettent d'explorer une autre manière de modéliser les clients du réseau BT. Le modèle actuel d'Enedis, Bagheera, date des années 90 et repose sur un panel de 2 000 clients représentatifs de la consommation BT. Avec les données des compteurs communicants, nous explorerons une approche « disruptive » pour modéliser les courbes de charge des clients BT afin de faciliter le suivi des évolutions de l'utilisation du réseau BT. De plus, ce modèle est conçu de manière à être reproductible avec des données de différents pays, facilitant ainsi son utilisation par la recherche. Une meilleure modélisation des courbes de charge permet une meilleure estimation des contraintes sur le réseau et donc une meilleure gestion prévisionnelle. De plus, dans un contexte de déploiement de solutions de flexibilité, cela permettrait de mieux en évaluer le gisement. Enfin, de nouveaux comportements de consommation faisant leur apparition comme l'autoconsommation et le V2G, avoir des outils de modélisation de charges BT performants devient nécessaire afin de mieux anticiper l'évolution des réseaux.

1.3.2 Structure

Ce mémoire de thèse est structuré en 4 chapitres organisés de la manière suivante :

- Le chapitre 2 aborde la thématique de l'identification des différents domaines où les données des compteurs communicants peuvent apporter des avantages et nous y trouverons la contribution C1 ;
- Le chapitre 3 aborde la thématique du foisonnement des charges du réseau BT et nous y trouvons la contribution C2 ;
- Le chapitre 4 aborde la thématique de la modélisation des courbes de charges du réseau BT et nous y trouvons la contribution C3 ;
- Le chapitre 5 aborde la conclusion de ce manuscrit de thèse ainsi que les perspectives, les axes d'amélioration du travail développé.

2 Cartographie d'utilisation des données

Une analyse de l'utilisation des données des compteurs est réalisée. Elle repose sur deux acteurs principaux : l'utilisation des données par les GRD et l'utilisation des données dans la recherche scientifique. Sur la base de cette analyse, une nouvelle cartographie de l'utilisation des données des compteurs est réalisée.

2.1 Utilisation des données des compteurs intelligents par les différents acteurs mondiaux

Dans cette section, nous allons examiner comment les différents acteurs mondiaux utilisent les données collectées par les compteurs intelligents. Pour cela, nous allons diviser cette analyse en fonction du type d'acteur, qu'il s'agisse d'un GRD ou de la recherche académique. Notre objectif final sera de la compréhension de l'utilisation des données pour améliorer les problèmes du réseau et de les cartographier.

2.1.1 GRD mondiaux

En 2014, la Commission Européenne a présenté un rapport d'analyse comparative de l'état d'avancement du déploiement des compteurs communicants dans l'Union Européenne. Il visait à fournir un aperçu des analyses coûts-bénéfices nationales que les États membres avaient été invités à mener à la suite de l'adoption de la directive électricité 2009/72/CE [25]. Afin de promouvoir l'efficacité énergétique et parmi les États membres, cette directive indique dans l'article 3, paragraphe 11, l'utilisation des compteurs « intelligents » comme une alternative possible. La décision de procéder au déploiement des compteurs reste au niveau des États membres, à partir des évaluations économiques de long terme, comme l'ACB (analyse coût-bénéfice). Lorsque l'évaluation est jugée positive au niveau d'un état membre, la directive indique qu'au moins 80 % des consommateurs devraient être équipés de compteurs « intelligents » [26]. Le niveau de déploiement des compteurs intelligents varie d'un pays à l'autre. Cela influe sur la manière dont chaque pays envisage l'utilisation de ces données, en plus de ses objectifs spécifiques. Pour limiter le périmètre à investiguer, les GRD européens analysés dans cette section ont été choisis dans le but d'obtenir une diversité de niveaux de déploiement de compteurs et parce que les informations sont accessibles. Ces GRD sont situés en Italie [27-29], Suède [30-33], Portugal [34-39], Espagne [40-45], Royaume-Uni [46-48] et Pays-Bas [49-52]. L'Italie a été le premier pays européen à moderniser son parc de compteurs. En 2017, ils ont commencé le déploiement d'une nouvelle génération de compteurs intelligents. La Suède et l'Espagne n'étaient pas loin derrière et avaient entamé leurs premières discussions sur le sujet avant les années 2010, terminant leur modernisation avant 2020. Les Pays-Bas et le Royaume-Uni avaient fixé des objectifs pour achever le déploiement en 2020. Bien que l'objectif de 100 % de compteurs intelligents ne soit prévu qu'en 2024, le Portugal est pionnier depuis 2010 dans l'étude et le développement de villes intelligentes. La ville d'Évora a été la première ville intelligente européenne, et aujourd'hui, l'objectif est de

transformer les principales villes portugaises en centres d'innovation et de recherche pour résoudre les défis de la transition énergétique.

En Amérique du Nord, les pays analysés sont les États-Unis et le Canada. Selon les données de l'Energy Information Administration, en 2021, les entreprises de services publics d'électricité aux États-Unis possédaient environ 111 millions de compteurs intelligents déployés, ce qui représente environ 69 % du total des compteurs électriques du pays [53]. Ces compteurs fournissent des données à la fois à l'entreprise de services publics et au client au moins une fois par jour. Ils peuvent varier des compteurs de communicants simples avec des relevés horaires aux compteurs en temps réel avec une communication bidirectionnelle intégrée, capables d'enregistrer et de transmettre des données instantanées. Selon l'Agence Internationale de l'Energie, plus de 82 % des compteurs installés au Canada sont intelligents. Toutes ses provinces et territoires ont déjà commencé le déploiement, mais le nombre total de clients desservis varie d'une province à l'autre. La Colombie-Britannique et l'Ontario ayant déjà terminé leur déploiement de compteurs intelligents, nous nous concentrerons donc sur ces provinces [64]. Les Canadiens ont achevé leur transition vers les compteurs intelligents au début des années 2010. Les Américains PG&E et FPL ont la majorité de leurs clients équipés de compteurs intelligents, tandis que ConEdison a commencé le déploiement de ses compteurs en 2017. Au total, cinq GRD ont été analysés, à savoir les américains PG&E [54-56], FPL [57-60], ConEdison [61-63] et les canadiens BC Hydro [65] et Hydro One [66].

En Asie-Pacifique, on trouve des pays à différents stades de maturité en ce qui concerne leur parc de compteurs intelligents. Six pays ont été choisis pour cette analyse : l'Australie [67-72], la Nouvelle-Zélande [73], le Japon [74-75], la Corée du Sud [76-78], la Thaïlande [79-80] et l'Indonésie [81-83]. La Nouvelle-Zélande a commencé le déploiement de son parc en 2008 et compte aujourd'hui plus de 90 % de compteurs intelligents. Le Japon a commencé les discussions sur les compteurs intelligents en 2014 et avait pour objectif de déployer 80 millions d'unités d'ici 2020. La Corée du Sud a commencé le remplacement de son parc de compteurs en 2012 et visait à atteindre 22,5 millions de compteurs intelligents au début des années 2020. L'Australie présente un cas particulier, car la région de Victoria a été pionnière dans le pays en rendant obligatoire l'utilisation de compteurs intelligents en 2011. Cependant, les autres régions du pays affichaient des taux de déploiement compris entre 25 % et 35 % en 2021. En Thaïlande, le principal GRD du pays prévoit de déployer un million de compteurs intelligents dans la zone appelée « corridor économique du pays ». Tandis qu'en Indonésie, son GRD a lancé un plan d'installation de 79 millions de compteurs dans tout le pays d'ici 2027.

Chacun de ces pays a une vision propre des avantages que les données des compteurs intelligents peuvent apporter et qui peuvent être classés en 4 catégories :

- Avantages pour les clients finaux ;
- Avantages pour le marché ;
- Avantages pour le réseau ;
- Avantages pour l'intégration des nouveaux usages.

Avantages pour les clients finaux

Les avantages pour les clients finaux commencent par le fait que les compteurs intelligents leur permettent d'avoir une facturation d'électricité précise et régulière, et non plus des estimations entre les relèves. Cette vision est partagée par tous les GRD. Un autre point simple mais important mentionné par la Suède, le Royaume-Uni, le Canada, le Japon et la Thaïlande est que les compteurs intelligents permettent d'ajuster la puissance souscrite, changer de type de contrat, ainsi que de connecter et déconnecter le client à distance. Ceci réduit le besoin d'interventions techniques coûteuses pour ces entreprises. Tous les GRD soulignent le fait que les données des compteurs aident les clients à avoir plus de contrôle et à mieux gérer leur consommation. Cela peut se faire grâce à l'affichage direct de la consommation sur les compteurs. Certains GRD, comme PG&E et FPL aux États-Unis, ont mis en place des applications conçues pour la visualisation de la consommation sous la forme d'un journal de bord avec différentes fonctionnalités, comme par exemple l'affichage graphique de la consommation à différentes échelles temporelles et la projection de la valeur de la facture. Le Royaume-Uni évoque comment les compteurs intelligents ont le potentiel de transformer les domaines de la santé et des services sociaux. Ce sujet est approfondi dans les domaines des soins aux personnes âgées et de l'identification des familles en situation de précarité énergétique.

Avantages pour le marché

Concernant les avantages pour le marché, il est largement reconnu par ces GRD que les données des compteurs peuvent aider à la création de nouveaux tarifs d'énergie, mieux adaptées aux clients et au réseau. L'Italie, l'Espagne, les Pays-Bas, le Royaume-Uni, le Canada, la Thaïlande et l'Indonésie évoque la création de tarifs multi-horaires, quotidiens, hebdomadaires et saisonniers. La Corée du Sud évoque des tarifs différents selon les saisons, car les consommations varient beaucoup d'une saison à l'autre. L'Australie, la Nouvelle-Zélande et les Pays-Bas parlent de tarifs flexibles, dans un mode de tarification dynamique. Les Australiens donnent l'exemple de tarifs qui encouragent la consommation pendant les

heures d'abondance de production solaire. La Thaïlande et le Royaume-Uni, croient également en des tarifications reposant sur des contrats prépayés. Ces derniers peuvent effectivement être aussi mis en place grâce aux compteurs intelligents. Le Royaume-Uni mentionne des tarifs spécifiques pour les véhicules électriques. Les canadiens de BC Hydro évoquent la possibilité de la fourniture d'incitations pour que leurs clients disposent de thermostats programmables et de logiciels de gestion de l'énergie. La Nouvelle-Zélande considère également que les données des compteurs peuvent contribuer à réduire les coûts de production d'énergie et à atteindre les objectifs de sa politique énergétique nationale.

Avantages techniques pour le réseau

Les avantages techniques pour le réseau électrique sont variés. La Suède, l'Italie, l'Espagne, le Portugal, les Etats-Unis, le Canada, l'Australie, la Nouvelle-Zélande, la Corée du Sud et la Thaïlande mentionnent comment les données des compteurs peuvent aider à identifier de manière plus efficace les défauts, interruptions et autres événements sur le réseau, ce qui réduit le temps moyen d'interruption pour les clients. Le Portugal, l'Italie, le Canada, la Nouvelle-Zélande, la Thaïlande, l'Indonésie et la Suède considèrent que les données des compteurs peuvent contribuer à réduire les pertes techniques et non techniques. La Suède souligne également que les données des compteurs améliorent la visibilité du réseau BT, ce qui permet une meilleure gestion de la qualité de l'énergie. L'Espagne évoque le potentiel de l'utilisation des véhicules pour le réseau (V2G) ouvert par les compteurs intelligents. Les canadiens de BC Hydro évoquent également l'utilisation des données pour l'optimisation de la tension. Les Japonais estiment que les données des compteurs intelligents les aideront à dimensionner plus précisément les différentes infrastructures du réseau.

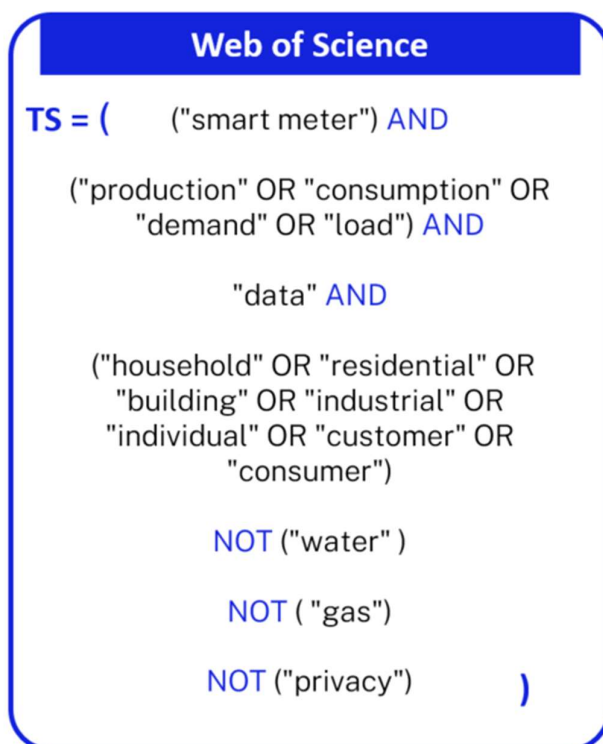
Avantages pour l'intégration des nouveaux usages

Tous les GRD comprennent que les compteurs intelligents représentent une aide essentielle pour l'intégration des énergies renouvelables dans le réseau. L'Espagne mentionne également l'avantage de produire sa propre énergie en autoconsommation photovoltaïque sans perdre les avantages d'être connecté au réseau. Le Royaume-Uni évoque également l'insertion de nouvelles technologies en général, telles que les appareils électroménagers intelligents et les véhicules électriques. L'Australie évoque la création de scénarios où des centrales virtuelles pourraient être mises en place pour permettre aux quartiers de commercer localement la

production solaire locale. La Corée du Sud aborde également l'augmentation des systèmes de stockage d'énergie et le développement de la recharge des véhicules électriques.

2.1.2 Thématiques de recherche scientifique liés aux compteurs communicants

L'insertion des compteurs communicants parmi les différents réseaux de distribution mondiaux fait que des données sont de plus en plus accessibles pour le distributeur et plus disponibles pour la recherche scientifique. Pour fournir une vue d'ensemble des articles utilisant des données de compteurs intelligents dans leurs recherches, une analyse bibliométrique a été réalisée le 17 octobre 2023 en utilisant la célèbre base de données Web of Science (WoS). La requête recherchée dans WoS est présentée par la Figure 2-1 :

A blue-bordered box with a blue header containing the text 'Web of Science'. Below the header, a search query is displayed in blue text. The query is: 'TS = (("smart meter") AND ("production" OR "consumption" OR "demand" OR "load") AND "data" AND ("household" OR "residential" OR "building" OR "industrial" OR "individual" OR "customer" OR "consumer") NOT ("water") NOT ("gas") NOT ("privacy"))'.

```
Web of Science
TS = ( ("smart meter") AND
("production" OR "consumption" OR
"demand" OR "load") AND
"data" AND
("household" OR "residential" OR
"building" OR "industrial" OR
"individual" OR "customer" OR
"consumer")
NOT ("water")
NOT ("gas")
NOT ("privacy") )
```

Figure 2-1 : Requête recherchée dans WoS

Avec cette requête, nous pouvons nous concentrer exclusivement sur les articles utilisant des données provenant de compteurs électriques intelligents. Cela nous permet d'éviter les articles fondés sur des campagnes de mesures manuelles. Cette requête nous permet également d'exclure les recherches liées aux compteurs d'eau et de gaz ainsi que les articles

liés à la confidentialité des données, car en dehors du périmètre de la thèse. Un total de 904 publications a été trouvées par WoS entre les années 2009 et 2022. Comme l'année 2023 n'était pas complète lors de la rédaction de ce manuscrit, il a été décidé de ne pas l'inclure dans ce travail de recherche. Cela signifie que les articles écrits cette année-là ne sont pas pris en compte dans cette analyse. L'évolution du nombre de recherches par an est présentée dans la Figure 2-2 :

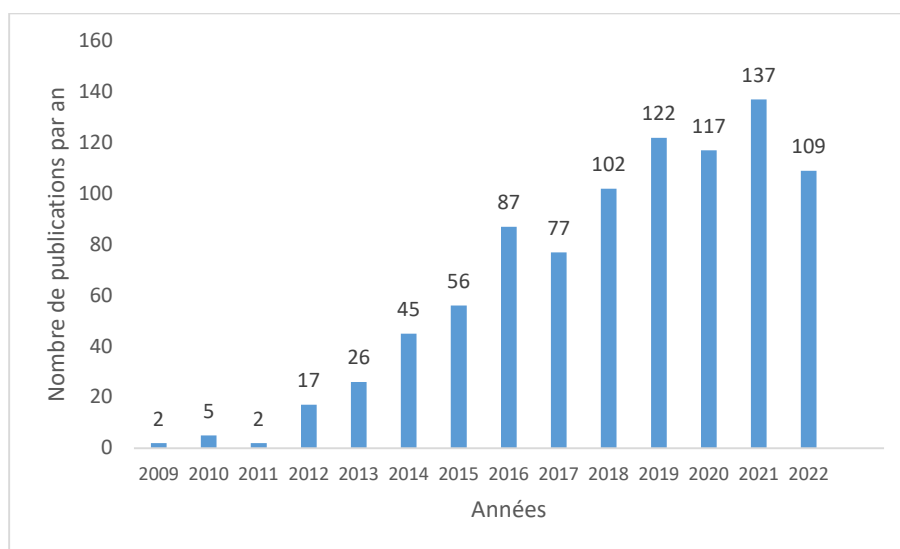


Figure 2-2 : Nombre de publications trouvées par WoS pour la requête TS

Nous pouvons observer une augmentation franche du nombre de publications par année. Un facteur qui pourrait expliquer cela est l'augmentation des jeux de données publics mis à disposition de la recherche scientifique, comprenant des données anonymisées de compteurs intelligents. Parmi ces jeux de données, nous pouvons mentionner : Pecan Street [84], Ausgrid Residents [85], Customer Behavior Trials [86], Low London Carbon [87] et Building Data Genome [88]. Un autre facteur qui contribue à cette augmentation est le déploiement croissant des compteurs, atteignant des chiffres plus élevés au fil des ans, comme cela peut être observé dans les autres sections de ce chapitre. L'objectif de cette analyse n'est pas d'entrer dans les sujets abordés par les 904 articles, mais plutôt de découvrir les cas d'usages des données de compteurs intelligents. La méthodologie appliquée est présentée dans le paragraphe suivant.

Méthodologie

À partir de la lecture du trio d'informations *Article Title, Author Keywords, Abstract* présents dans le résultat de la requête effectuée dans *WoS* une l'analyse a été fait manuellement. Elle consistait à classer les articles dans les catégories d'utilisation des données des compteurs par les GRD présentées précédemment. Pour les articles qui ne correspondaient à aucune des catégories présentées auparavant, ils ont été regroupés un par un de manière à ce que le thème central auquel ces articles appartenaient soit clairement interprété et mis en évidence. Comme mentionné précédemment, cette analyse n'a pas pour objectif de se plonger dans les articles eux-mêmes et de fournir des détails précis sur leurs recherches. Au total, 6 grands thèmes ont été identifiés :

- *Analyse des comportements de consommation ;*
- *Aspects techniques des réseaux électriques ;*
- *Ecosystème des compteurs intelligents ;*
- *Évolution des usages ;*
- *Tarifcation et marché de l'électricité ;*
- *Services pour les clients finaux.*

Dans la suite, nous allons approfondir les détails de chacun de ces thèmes. La division des articles par thème est présentée par la Figure 2-3 :

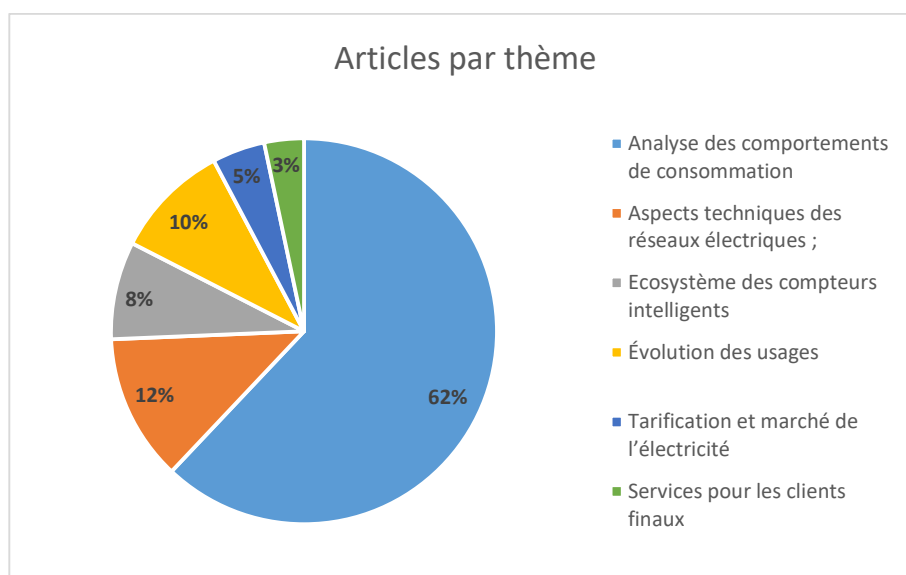


Figure 2-3 : Division des articles par thème

Le contenu de chaque thème sera mis en évidence à partir des articles les plus cités selon le WoS. Ainsi, le lecteur peut prendre connaissance des articles clés de chaque thème pour avoir une vue d'ensemble des sujets les plus discutés. Cela est enrichissant car de cette manière, il est possible d'avoir une idée de ce qui se fait actuellement dans la recherche et quelles sont les zones qui ont encore de l'espace à explorer.

2.1.2.1 Analyse des comportements de consommation

Au total, sur les 904 articles, 561 articles correspondent à ce thème, ce qui équivaut à 62 %. Il englobe l'analyse de la charge, la segmentation et la caractérisation des comportements. De plus, il explore la désagrégation et l'agrégation de la charge, la prévision à différentes échelles temporelles, en passant par les différentes modélisations et l'ensemble du monitoring.

Au cours de la dernière décennie, la pénétration des compteurs intelligents a connu une augmentation significative, avec des entreprises mondiales du secteur privé ainsi que par les GRD. Les chercheurs ont réagi à cette tendance par une augmentation des études explorant diverses méthodes et techniques d'utilisation des données des compteurs intelligents [89]. La littérature couvre un large éventail d'applications, notamment la prévision, le regroupement, la classification et l'optimisation [89]. Une revue approfondie de la littérature révèle des écarts notables entre les études, rendant les comparaisons et les analyses approfondies difficiles [89].

Un aspect crucial des applications des réseaux intelligents est le développement de modèles de charge au niveau de l'appareil [90]. Contrairement aux approches de surveillance directe des appareils, l'exploitation des données des compteurs intelligents pour générer des modèles de charge au niveau de l'appareil de manière non intrusive et généralisable à tous les foyers équipés de compteurs intelligents est considérée comme une approche flexible et pratique [90]. De tels modèles peuvent être cruciaux pour la désagrégation non intrusive de la charge, la prévision de la charge et le soutien à la réponse à la demande [91]. Des infrastructures et des algorithmes utilisant des compteurs intelligents et des smartphones ont été proposés pour permettre de nouveaux services d'efficacité énergétique, tels que des factures d'électricité détaillées par appareils [92]. La désagrégation, l'extraction des données de fin d'utilisation et de l'appareil à partir de signaux énergétiques agrégés, est positionnée comme

une solution rentable et évolutive pour réaliser des réductions d'énergie significatives dans les secteurs résidentiels et commerciaux [93]. Les techniques de *non-intrusive load monitoring* (NILM) jouent un rôle crucial en fournissant des informations détaillées sur les modèles de consommation d'énergie électrique résidentielle au niveau de chaque appareil [94]. Les algorithmes de désagrégation itérative fondés sur la signature hybride offrent des solutions efficaces pour décomposer les charges électriques agrégées en contributions individuelles de chaque utilisation finale [94]. Des approches de NILM fondées sur le filtrage de particules ont été proposées pour identifier les appareils en fonction de leurs caractéristiques électriques dans la consommation totale d'énergie du foyer [95].

Les efforts visant à améliorer la prévision de la charge intra-journalière au niveau du système ont avancé avec le déploiement de l'infrastructure de comptage [96]. Le regroupement est appliqué pour identifier des groupes de clients avec des schémas de consommation similaires à partir des compteurs intelligents avant de réaliser une prévision de la charge [96]. Ce changement par rapport à la prévision traditionnelle agrégée au niveau du système présente de nouvelles perspectives pour la prévision de la charge à différents niveaux du système électrique [96]. La proposition d'inférer les états d'occupation à partir de données de séries temporelles de consommation à l'aide d'un modèle de Markov caché offre des aperçus sur le comportement et les caractéristiques des utilisateurs [97]. Les comportements de consommation dans les données peuvent prédire avec une bonne précision les caractéristiques des utilisateurs, mettant en évidence la prédictibilité individuelle potentielle de la consommation des utilisateurs au niveau de la population [97]. Dans le contexte de la consommation résidentielle, le déploiement des compteurs aux États-Unis a fourni un accès aux données de consommation d'énergie horaires [98]. Ces données ont été utilisées pour des méthodologies de segmentation de l'électricité domestique à partir des formes de charges prétraitées [98]. Les méthodes de regroupement appliquées aux données des compteurs intelligents résidentiels offrent des opportunités pour les GRD [99]. Identifier les candidats à la réponse à la demande et améliorer la modélisation du profil énergétique sont des avantages potentiels du regroupement pour les GRD [99].

Des méthodes probabilistes de prévision de la charge, telles que les réseaux neuronaux profonds et les modèles de régression par quantiles, sont proposées pour traiter la variabilité et l'incertitude des profils de charge futurs [100]. Ces modèles démontrent des performances supérieures dans l'estimation de la demande individuelle et agrégée par rapport aux méthodes traditionnelles [100]. Les applications d'apprentissage profond pour la prévision de la charge domestique introduisent une nouvelle approche fondée sur un réseau neuronal

récurrent profond [101]. Cette méthode adresse directement la volatilité élevée et l'incertitude des profils de charge, surpassant les techniques traditionnelles dans la prévision de la charge domestique [101]. Des modèles intelligents d'exploration de données sont proposés pour analyser, prévoir et visualiser des séries temporelles d'énergie afin de découvrir divers motifs temporels de consommation d'énergie [102]. Le modèle proposé combine le regroupement non supervisé des données, l'analyse d'extraction de comportements et la prédiction fondée sur les réseaux bayésiens pour la prévision de l'utilisation de l'énergie [102]. Les méthodes d'apprentissage par transfert, telles que le modèle proposé, visent à améliorer la précision de la prédiction de l'énergie pour les nouveaux bâtiments avec des données limitées en utilisant des ensembles de données de bâtiments similaires [103]. Des modèles prédictifs fondés sur les données pour l'utilisation d'énergie des appareils sont présentés, utilisant des mesures de réseaux sans fil, des données météorologiques et l'utilisation enregistrée de l'énergie des luminaires [104]. Des modèles statistiques, y compris la régression linéaire multiple, la machine à vecteurs de support, la forêt aléatoire et les machines de renforcement du gradient, sont comparés et évalués pour leur capacité à décomposer la demande totale en énergie mesurée par un compteur intelligent en celles utilisées par chaque appareil individuel [104]. Les variations de l'utilisation de l'électricité à des échelles temporelles finement résolues sont explorées à l'aide de données mesurées sur la consommation d'électricité des résidences [105]. Le regroupement des résidences en groupes fondés sur des comportements similaires d'utilisation horaire de l'électricité révèle des influences saisonnières sur la consommation d'énergie [105]. Un sondage a été fait pour identifier des corrélations entre les caractéristiques des propriétaires et les comportements de consommation d'électricité [105]. Les résultats indiquent que des variables telles que le fait de travailler à domicile, le nombre d'heures de télévision regardées par semaine et le niveau d'éducation présentent des corrélations significatives avec la forme moyenne du profil.

Cette vue des articles les plus cités met en évidence la diversité des applications et des méthodologies émergeant de l'analyse de la consommation d'électricité résidentielle et des applications des données des compteurs intelligents. Nous pouvons observer que la majorité des articles se concentrent sur la modélisation de la consommation des appareils et sur son agrégation pour modéliser la consommation résidentielle, ainsi que sur la segmentation des clients en fonction de leurs comportements de consommation. Cela est justifié par le fait que les données des compteurs intelligents permettent une vision précise à échelle du consommateur, et les jeux de données accessibles sont souvent de la consommation résidentielle. Cependant, il n'est pas clair comment une modélisation de la consommation résidentielle fondée sur la consommation des appareils peut être utilisée dans les situations d'exploitation et de planification du réseau. Il existe tout un aspect à explorer sur la manière

dont la modélisation et la segmentation de la consommation résidentielle peuvent être réalisées de manière à être utilisées pour répondre aux problématiques du réseau du point de vue du GRD.

2.1.2.2 Aspects techniques des réseaux électriques

Ce thème implique les aspects techniques du réseau de distribution. Au total, sur les 904 articles, 111 articles correspondent à ce thème, ce qui équivaut à 12 %. Il couvre la qualité de l'énergie, les pertes, les analyses des postes, la planification et les opérations, ainsi que la gestion de la demande.

Les articles de ce thème explorent en profondeur les vastes domaines du réseau en utilisant l'abondance de données fournies par les compteurs intelligents pour améliorer la gestion du réseau et lutter contre le vol d'électricité. L'avènement de la technologie des réseaux intelligents a entraîné un changement de paradigme dans la gestion de l'énergie, mettant l'accent sur les méthodologies fondées sur les données. Une avancée significative dans cette direction est évidente dans [106], où une méthodologie scalable pour le ciblage des programmes de réponse à la demande est proposée. En utilisant des données de compteurs intelligents au niveau individuel, cette approche aborde les défis liés à la sélection des clients, transformant le scénario traditionnel de la réponse à la demande [106]. Une optimisation efficace des programmes de réponse à la demande dans les systèmes de gestion d'énergie domestique est explorée dans [107]. L'article introduit une approche pour résoudre des problèmes d'optimisation non convexes, démontrant sa pertinence pour une implémentation dans des dispositifs intégrés tels que les compteurs intelligents [107]. L'utilisation des données des compteurs intelligents pour estimer le potentiel de réponse à la demande est un thème central dans [108]. L'étude utilise une régression linéaire et une classification non supervisée pour déterminer la corrélation entre l'activité de la climatisation résidentielle, les changements de température et le potentiel de réponse à la demande. Les résultats suggèrent un potentiel significatif de réduction à court terme et mettent en évidence le ciblage des gros consommateurs pour maximiser la rentabilité.

En explorant le domaine de la sécurisation du réseau, l'article [109] présente un système de détection d'intrusion conçu pour lutter contre le vol d'énergie. Ce système fusionne des données provenant des compteurs intelligents et des modèles de consommation, démontrant une précision accrue dans la distinction entre les profils de charge normaux et anormaux. Ce

thème d'amélioration de la sécurité se poursuit dans [110], qui propose une méthode d'analyse des données pour détecter divers types de vols d'électricité. En combinant le coefficient d'information maximal et la technique de regroupement, l'étude offre une solution complète au défi croissant du vol d'électricité. Les pertes non techniques est un problème abordé dans [111]. Cet article introduit une méthodologie utilisant l'apprentissage supervisé pour identifier les pertes d'électricité dues à des anomalies ou à des fraudes [111]. Dans le contexte des régions confrontées à des défis de distribution d'électricité [112], l'étude est fondée sur l'analyse et la prévention des pertes non techniques à l'aide des compteurs intelligents. Cette analyse englobe les pertes financières, la réduction de la fiabilité du réseau et les stratégies préventives.

Ces études soulignent l'impact des méthodologies fondées sur les données sur les aspects techniques des réseaux. La définition des stratégies de réponse à la demande, d'amélioration de la sécurité contre le vol d'énergie et à la résolution des pertes non techniques contribuent collectivement à un écosystème énergétique plus sûr et efficace. Cependant, de nombreux autres aspects techniques peuvent être davantage explorés avec l'aide des données des compteurs intelligents. On peut souligner des sujets tels que le calcul des indicateurs tels que le *system average interruption duration index* (SAIDI) et *system average interruption frequency index* (SAIFI), la maintenance prédictive du réseau, l'optimisation de la configuration topologique du réseau. Nous pouvons également mentionner le sujet du vieillissement du matériel du réseau, où nous pouvons trouver des articles sur la dégradation des câbles, comme le [155]. Cela indique la diversité des thèmes qui peuvent être approfondis grâce aux données précises du réseau BT.

2.1.2.3 *Ecosystème des compteurs intelligents*

Ce thème englobe une variété de sujets liés aux compteurs intelligents. Sur les 904 articles au total, 74 articles correspondent à ce sujet, soit 8 %. Il aborde notamment la conception et la construction de ces compteurs, leur déploiement, leur adoption et leur mise en œuvre. De plus, il explore les tendances à venir, les défis qui les accompagnent, ainsi que les scénarios potentiels. Enfin, il examine également les aspects liés à la cyber-sécurité et les éléments relatifs à la communication et à ses protocoles.

Les compteurs intelligents gagnent en importance en tant qu'étape fondamentale pour les pays cherchant un développement durable et une efficacité énergétique [113]. En Indonésie,

les compteurs intelligents sont considérés comme une technologie prometteuse pour la préservation des ressources énergétiques limitées. Malgré les avantages potentiels, il existe une sensibilisation limitée des consommateurs et une compréhension limitée de ces dispositifs, entraînant un scepticisme [113]. Pour résoudre ce problème, une étude a été menée pour développer un indice mesurant la propension des consommateurs à adopter des compteurs intelligents dans les bâtiments résidentiels. Aussi, il identifie les facteurs influençant l'acceptation des consommateurs. Les résultats améliorent la compréhension des perceptions des consommateurs, aidant les décideurs et les compagnies d'énergie à élaborer des politiques efficaces [113]. Une revue inter-pays explore davantage l'adoption des réseaux intelligents dans les bâtiments résidentiels [114]. Comprendre les perceptions des consommateurs et l'influence des valeurs et normes nationales sur le comportement d'adoption des compteurs intelligents est crucial pour les décideurs politiques et les investisseurs [114]. Une étude comparative entre Taiwan, la Corée du Sud, l'Indonésie et le Vietnam fournit des informations sur les différences régionales dans l'adoption des compteurs intelligents par les consommateurs, facilitant la prise de décision en matière d'investissement et le développement des réseaux intelligents [114].

Dans la partie des communications, l'introduction de compteurs intelligents avec des interfaces radio cellulaires ajoute une nouvelle charge aux réseaux cellulaires sans fil [115]. À mesure que les ressources énergétiques distribuées augmentent, il y a un passage à des volumes de données et des taux plus élevés par dispositif, posant des défis aux systèmes cellulaires actuels [115]. L'étude souligne la nécessité de mettre à niveau les normes cellulaires pour accommoder des compteurs intelligents [115]. La fourniture ubiquitaire de communications dans les villes intelligentes pour les systèmes d'énergie intelligents est essentielle [116]. Le manque d'interopérabilité nécessaire et l'intégration des normes de communication entravent le déploiement efficace [116]. Une architecture hiérarchique, intégrant le fog et le cloud computing, ainsi qu'un algorithme de sélection de réseau, adresse ce problème, soutenant l'infrastructure et les services des systèmes d'énergie intelligents dans les villes intelligentes [116].

Cette structure organisée offre un aperçu complet de l'adoption des compteurs intelligents, des stratégies de déploiement. Ce sont des informations très utiles pour un pays qui n'a pas encore commencé la transition de ses parcs de comptage afin de prendre en compte les difficultés et les avantages possibles. Le grand enjeu de ce thème dans les années à venir sera sans aucun doute la cybersécurité. Comme le réseau électrique est une partie vitale du fonctionnement des villes et des pays, garantir la sécurité du système sera toujours une

priorité. De plus, assurer la sécurité des données personnelles des clients du réseau est d'une importance extrême.

2.1.2.4 Evolution des usages

Ce thème est lié aux nouveaux usages du réseau de distribution. Au total, sur les 904 articles, 88 articles correspondent à ce thème, ce qui équivaut à 10 %. Il aborde tout ce qui concerne les énergies renouvelables, les véhicules électriques, le stockage, les clients dit consom'acteur et les flexibilités.

Dans le domaine des consom'acteurs, un modèle d'optimisation linéaire explore l'effet de la résolution temporelle sur les taux d'autoconsommation et la dimension optimale des systèmes PV et PV-batterie [117]. Utilisant des données de compteur intelligent à haute résolution provenant de résidences allemandes, l'étude met en évidence le rôle crucial de la résolution temporelle dans les profils de charge pour une dimension précise du système [117]. Les résultats apportent des perspectives précieuses pour les décideurs politiques (policy-makers), les GRD et les résidences envisageant des investissements dans des systèmes PV-batterie [117]. L'évaluation économique joue un rôle pivot dans l'adoption des systèmes de batterie photovoltaïque [118]. En analysant des profils de charge d'électricité du monde réel, un modèle de simulation techno-économique optimise les configurations de systèmes photovoltaïques-batteries (PVB) pour les résidences [118]. L'étude révèle la rentabilité hétérogène des systèmes PVB parmi les résidences, soulignant l'importance d'identifier précisément les installations rentables [118]. De plus, un algorithme d'apprentissage automatique est développé pour prédire la rentabilité du système, fournissant des perspectives précieuses tant pour les résidences que pour les GRD [118]. L'intégration des systèmes photovoltaïques et de batteries offre une solution potentielle pour aligner la production solaire avec la demande résidentielle [119]. En utilisant un an de données de compteur intelligent, une étude calcule les avantages économiques à la fois des systèmes PV et des systèmes couplés PV-batterie, en tenant compte de divers schémas de tarification de l'électricité [119]. Alors que le PV s'avère rentable pour la majorité des consommateurs, les systèmes PV-batterie améliorent l'autosuffisance, avec une rentabilité potentielle soumise aux futurs prix de l'électricité et aux réductions de coûts [119].

Dans le domaine de l'intégration des VE, une approche probabiliste tire parti des données de compteur intelligent et de charge de VE pour évaluer l'impact sur les réseaux de distribution

[120]. L'étude utilise des ensembles de données divers représentant des réseaux de distribution britanniques urbains, ruraux et fortement chargés, révélant des variations dans la capacité des réseaux à accueillir les VE [120]. En considérant la diversité spatiale et temporelle de la demande de charge des VE, la recherche suggère des efforts collaboratifs entre les GRD et les nouveaux acteurs du marché pour améliorer la robustesse de l'infrastructure et réduire les incertitudes de planification [120]. Dans [121], une méthode de surveillance de charge non intrusive (NILM) utilise un réseau neuronal profond multi-tâches pour analyser les données de compteur intelligent, extrayant des informations détaillées sur la consommation et l'état de fonctionnement des appareils avec une grande précision [121]. Les résultats permettent le développement d'un système de gestion d'énergie NILM efficace pour les microgrids résidentiels, optimisant la répartition des centrales de génération distribuée et des charges contrôlables tout en privilégiant la satisfaction du client [121]. La prévision de la flexibilité de la demande devient cruciale pour la gestion des charges résidentielles en utilisant les données de compteur intelligent [122]. Une méthodologie à deux niveaux est proposée pour décomposer la charge totale prévue des résidences, en incorporant des mesures de compteurs intelligents avec réseaux neuronaux artificiels [122]. La méthodologie, validée sur des données réelles de site pilote, soulignant son applicabilité pour les applications de gestion de la demande [122].

Grâce aux données des compteurs intelligents, de nombreux articles liés à l'évolution des usages du réseau émergent et devraient continuer à émerger. Les sujets se concentrent principalement sur les systèmes photovoltaïques et les véhicules électriques car ils sont directement liés à l'usage résidentiel, qui est l'échelle à laquelle les données sont mesurées. Parmi les articles cités, le sujet lié à la capacité d'accueil du réseau BT n'est pas abordé. Cependant, ce sujet est important pour le GRD afin de mieux planifier le réseau et d'accueillir ces nouveaux usages de manière plus efficace.

2.1.2.5 Tarification et marché de l'électricité

Ce thème englobe tout ce qui concerne la tarification de l'électricité, allant de l'analyse des tarifs existants à la proposition de création de nouveaux tarifs. Mais au-delà de cela, ce thème englobe tout ce qui concerne les marchés et les politiques énergétiques. Sur les 904 articles au total, 40 articles correspondent à ce thème, ce qui équivaut à 4 %.

L'augmentation de la disponibilité des données de compteur intelligent de haute qualité a ouvert la voie à des approches innovantes dans le paysage des tarifs d'électricité. Dans le domaine de la vente au détail, l'article [123] explore le potentiel de personnalisation des tarifs selon les heures creuses fondée sur des données de compteur intelligent historiques étendues. Cette approche, ancrée dans un cadre axé sur les données, respecte la capacité des consommateurs à choisir librement et leur volonté de s'engager. Le modèle capture la relation de Stackelberg entre les détaillants cherchant à maximiser leurs profits et les consommateurs stratégiques, présentant une nouvelle solution pour des tarifs incitatifs [123]. Pour valider ce modèle, l'étude utilise un ensemble de données de compteur intelligent de la Commission de régulation de l'énergie en Irlande [123]. La difficulté d'analyser l'impact des politiques tarifaires au milieu de facteurs perturbateurs tels que la météo est abordée dans [124]. L'étude introduit une méthodologie pour modéliser l'effet de la météo sur la demande d'électricité résidentielle, soulignant le besoin d'analyses d'impact claires et reproductibles [124]. En utilisant un ensemble de données de compteur intelligent dans le sud-ouest de l'Ontario, au Canada, l'étude évalue l'impact des tarifs obligatoires ToU, révélant des effets nuancés sur la conservation de l'électricité selon les périodes de tarification [124]. L'article [125] se penche sur le comportement des consommateurs concernant les tarifs intelligents en heures creuses. Une expérience de sondage menée sur un échantillon représentatif au niveau national des payeurs de factures énergétiques britanniques révèle que plus d'un tiers des répondants sont favorables à la transition vers un tarif intelligent. Cependant, l'étude [125] met en évidence des variations dans la volonté de changer, motivées par des facteurs tels que l'aversion à la perte et la possession d'appareils demandeurs-flexibles. Des campagnes marketing ciblées envers les propriétaires de véhicules électriques émergent comme une avenue potentielle pour stimuler l'adoption [125].

Explorant les effets des tarifs en heures creuses et des tarifs d'exportation dans les zones résidentielles avec stockage d'énergie, l'article [126] étudie l'impact sur les flux de pointe dans les réseaux de distribution. En simulant le fonctionnement de la batterie fondé sur des données au niveau des résidences, l'étude souligne la complexité des effets tarifaires, surtout dans des zones avec un stockage d'énergie, des panneaux solaires PV et des pompes à chaleur diversifiés [126]. Les résultats soulignent l'importance d'incitations dédiées à la réduction des pointes et de stratégies de contrôle de stockage plus intelligentes [126]. En se concentrant sur les défis des structures tarifaires existantes, [127] examine les propositions de réforme des tarifs électriques à taux fixe par défaut. En utilisant des données de compteur intelligent de la région victorienne du marché national de l'électricité en Australie, l'étude révèle les effets néfastes des structures à taux fixe sur les résidences en difficulté financière. Même après un rééquilibrage des tarifs réseau, le modèle met en évidence les avantages de passer à des

structures tarifaires plus réfléchies sur le coût, en particulier lors de l'incorporation de la réponse à la demande [127].

Ces études contribuent collectivement à notre compréhension du scénario en constante évolution des tarifs d'électricité. Il est possible de constater que les données des compteurs intelligents peuvent aider au développement de tarifs plus adaptés aux différents types de consommation. De plus, on peut observer que les tarifs peuvent servir d'incitation à la consommation en dehors des heures de pointe électrique. Un sujet de recherche qui devrait se développer est l'étude de tarifs dynamiques liés à la production d'énergies renouvelables, afin d'encourager la consommation aux moments de forte production d'énergie.

2.1.2.6 Services pour les clients finaux

Ce thème est exclusivement lié au consommateur final d'énergie. Au total, sur les 904 articles, 31 articles correspondent à ce thème, ce qui équivaut à 3 %. Il englobe depuis la prise de connaissance du client sur sa facture jusqu'aux actions générées à partir de cette connaissance, telles que la consommation responsable et les économies d'énergie. Ainsi qu'aux différents services qui peuvent être imaginés pour bénéficier au client.

Les compteurs intelligents jouent un nouveau rôle dans l'équilibrage de la production-demande d'électricité, permettant des services énergétiques en temps réel et de nouvelles solutions de facturation [128]. En Italie, une nouvelle génération de compteurs intelligents est déployée, permettant des services tels que la sensibilisation à la consommation de l'énergie et l'automatisation domestique des usages. Cet article examine la situation actuelle en Europe concernant les services en temps réel pour les clients finaux et discute des options architecturales et technologiques pour le déploiement des compteurs intelligents en Italie [128].

Les campagnes d'efficacité énergétique personnalisées qui utilisent des informations spécifiques aux résidences ont le potentiel de déclencher des économies d'énergie substantielles dans le secteur résidentiel [129]. En extrayant ces informations des données de compteur intelligent à l'aide de l'apprentissage automatique, les GRD peuvent dériver des caractéristiques et reconnaître des catégories de résidences similaires pour améliorer l'efficacité énergétique [129]. La gestion de la demande est un aspect clé des réseaux

intelligents, et l'intégration de l'informatique de service dans ce domaine propose un système de recommandation personnalisé pour les appareils économes en énergie [130]. Le système utilise une surveillance non intrusive de la charge des appareils pour déduire les intérêts et besoins des utilisateurs, recommandant des appareils économes en énergie en fonction des profils d'utilisateurs et des publicités d'appareils. Des expériences valident l'efficacité du système proposé dans l'optimisation de l'exploitation du réseau et la promotion des économies d'énergie [130]. La croissance rapide des données dans les réseaux intelligents offre des opportunités aux GRD pour optimiser l'exploitation du réseau grâce à l'analyse des données et à la prise de décision [131]. Un système de recommandation de plan tarifaire d'électricité personnalisé est proposé pour les utilisateurs résidentiels. Le système analyse les habitudes de consommation d'énergie des utilisateurs, recommandant le plan tarifaire le plus adapté en fonction des techniques de filtrage collaboratif. Des expériences fondées sur des données de compteurs intelligents réelles valident l'efficacité du système [131]. Quantifier les économies d'énergie et les potentiels de réduction de la demande de pointe dans les bâtiments lors d'une période de réponse à la demande est essentielle pour des allocations optimales des ressources [132]. Les résultats de simulation fondés sur un modèle de bâtiment et des données de compteur intelligent révèlent l'impact des changements de consigne des climatisations sur les économies d'énergie et la réduction de la demande de pointe [132]. Les informations d'utilisation en temps réel ont un impact significatif sur la consommation d'électricité résidentielle, comme le montre une expérience naturelle à grande échelle en Irlande du Nord [133]. Le remplacement des compteurs prépayés par des compteurs intelligents a permis aux consommateurs de suivre l'utilisation en temps réel, entraînant une baisse substantielle de la consommation d'électricité. Cette réduction est robuste selon différentes spécifications et méthodes de correction de biais de sélection, démontrant des réductions rentables des émissions de dioxyde de carbone [133].

L'intégration des compteurs intelligents dans le secteur résidentiel offre de grandes promesses pour répondre à la demande croissante d'électricité. En comprenant et en incorporant des facteurs tels que la connaissance de l'économie d'électricité et la sensibilisation à l'environnement, ces avancées technologiques améliorent non seulement l'efficacité énergétique, mais contribuent également à une approche plus durable et informée de la consommation d'énergie. Alors que les réseaux intelligents continuent d'évoluer, la mise en œuvre réussie de recommandations personnalisées renforce le rôle des compteurs intelligents dans le façonnement d'un scénario énergétique plus résilient pour l'avenir.

2.2 Cartographie de cas d'usage à partir des données des compteurs intelligents

Notre analyse de l'utilisation des compteurs intelligents par les différents acteurs mondiaux a éclairé diverses questions autour de ce sujet. La première chose à souligner est que le déploiement de compteurs par les GRD a toujours eu pour objectif principal d'automatiser le relevé des mesures de consommation. En Europe le déploiement est ensuite relatif à l'ouverture des marchés de l'énergie. Cette automatisation permet aux GRD de ne plus envoyer d'agents pour effectuer la collecte de mesures. Il est à souligner que la facture d'énergie des consommateurs devient alors fondée sur leur consommation réelle et non estimée. En outre, ces données ouvrent la voie à une meilleure gestion du réseau. Comme mentionné dans la section 2.1.1, les GRD considèrent comme avantage de l'utilisation des données une meilleure identification et résolution des incidents, tels que les coupures. Cela constitue donc un autre cas d'économie pour le GRD, offrant aux clients une meilleure qualité de distribution d'énergie. Les clients, en plus de bénéficier de toutes les améliorations que ces données apportent au réseau, profitent directement de la vision que les données leur offrent sur leur manière de consommer de l'énergie. Souligné par tous les GRD, les données des compteurs intelligents aident à fournir des services qui permettent au client d'obtenir des retours sur sa consommation, fournissent des informations sur les moyens de réduire la consommation et favorisent ainsi une consommation d'énergie consciente et responsable. Un autre point à souligner est l'utilisation de ces données pour la tarification. Elles permettent une meilleure compréhension de la manière dont différents types de tarifs affectent la consommation. De plus, elles permettent la création de nouvelles offres mieux adaptées aux différentes habitudes de consommation. Elles peuvent aussi être un incitateur pour que les clients modifient leurs habitudes de consommation afin de consommer aux moments moins critiques pour le réseau. L'utilisation des données des compteurs intelligents peut favoriser l'intégration des sources d'énergie renouvelable dans le réseau de distribution comme mentionné par les GRD observés dans la section 2.1.1. Avec la croissance rapide de ce domaine, que ce soit par la création de parcs d'énergies renouvelables ou par leur installation par des consommateurs résidentiels, les GRD sont confrontés à de nombreuses questions techniques liées au réseau, telles que la surtension. Les données collectées peuvent les aider à trouver des solutions pour gérer cela et favoriser l'intégration de ces sources d'énergie.

Ces avantages et applications des données des compteurs intelligents, mis en évidence lors de leur déploiement par les GRD, sont également observés dans la recherche scientifique. L'analyse bibliométrique que nous avons réalisée conduit à des résultats qui met en évidence cette réalité. La classification des articles dans les 6 thèmes précédemment mis en avant

souligne ce fait. Le tableau suivant synthétise dans quel thème se trouve chacune des références :

Thème	References
<i>Analyse des comportements de consommation</i>	89 – 105
<i>Aspects techniques des réseaux électriques</i>	106 – 112
<i>Ecosystème des compteurs intelligents</i>	113 – 116
<i>Évolution des usages</i>	117 – 122
<i>Tarification et marché de l'électricité</i>	123 – 127
<i>Services pour les clients finaux</i>	128 – 133

Tableau 2-1 : Classement des références suivant les thématiques correspondantes

Pour les lecteurs souhaitant approfondir l'ensemble des résultats trouvés par la requête WoS, l'ensemble des 904 articles résultants est disponible en ligne dans un référentiel GitHub en accès libre (<https://github.com/these-guilherme/DataCartographie>). En se basant sur l'ensemble de cette recherche sur l'utilisation des données des compteurs, une cartographie de leurs différentes utilisations est proposée dans la Figure 2-4 :

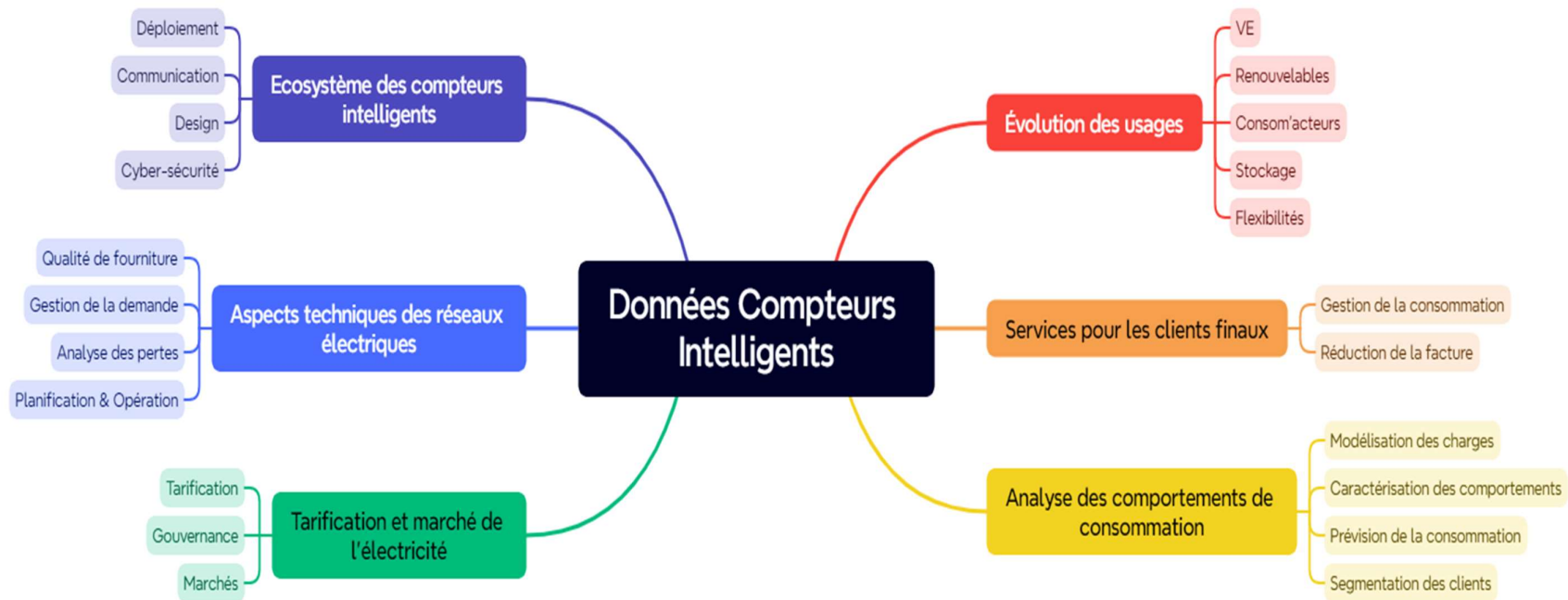


Figure 2-4 : Cartographie de cas d'usages des données des compteurs intelligents

2.3 Conclusion

Tout au long de ce chapitre, nous avons analysé l'utilisation des compteurs intelligents par différents acteurs à l'échelle mondiale. Il a été possible d'observer les raisons qui ont poussé les GRD à changer leur parc de comptage, en plus de comprendre les avantages qu'ils envisageaient avec l'utilisation des nouvelles données. Nous avons également pu comprendre que la recherche scientifique est un grand contributeur à ce sujet. En effectuant une étude bibliométrique dans la base de données WoS sur les différents articles utilisant les données des compteurs communicants, nous avons pu constater l'immensité des articles qui ont été développés à partir de ces données de compteurs. Nous pouvons observer les sujets de recherche centraux tout en mettant en évidence les domaines à explorer. Grâce à ce travail, nous avons été en mesure de créer une cartographie de l'utilisation des données. Il s'agit de la première contribution scientifique (C1) de ce travail de thèse. Son objectif est d'aider la recherche en mettant en évidence les domaines explorés et, par conséquent, de faciliter l'exploration de nouveaux sujets de recherche. De plus, elle peut être utile aux GRD du monde entier qui n'ont pas encore entamé le déploiement de compteurs intelligents, en les aidant à comprendre les domaines dans lesquels ils peuvent bénéficier de ces avantages, clarifiant ainsi le débat et la planification de ce changement.

Cette cartographie et tout le travail associé pour la développer, qui sont présentés tout au long du chapitre, renforcent l'intérêt des prochains chapitres de cette thèse et par conséquent, des contributions (C2) et (C3). Nous avons observé que parmi les articles de référence de notre recherche bibliographique, le sujet de la capacité d'accueil du réseau n'est pas exploré. Afin de mettre en lumière ce sujet, dans le chapitre 3, nous aborderons le thème du foisonnement, car le coefficient de foisonnement est utilisé dans le dimensionnement du réseau. Ainsi, une meilleure compréhension et estimation de celui-ci peuvent être utiles pour la discussion de la capacité d'accueil. Nous avons vu dans la section 2.1.2.1 que la modélisation de la consommation du réseau à partir des données des compteurs est principalement abordée par les articles de référence à partir de la modélisation des équipements. Cependant, il n'est pas clair comment une modélisation de la consommation résidentielle fondée sur la consommation des appareils peut être utilisée dans les situations d'exploitation et de planification du réseau. Afin de répondre à cette nécessité, dans le chapitre 4, une nouvelle méthode de modélisation des courbes de charge des clients consommateurs du réseau BT sera proposée.

On peut constater que le centre de cette cartographie est constitué par les données des compteurs intelligents, à partir desquelles émergent les 6 thèmes. Cette conception permet d'intégrer de nouveaux thèmes à mesure que la recherche progresse. Chaque thème se subdivise en sujets spécifiques qui le composent, préservant ainsi la possibilité d'ajouter de

nouveaux sujets et de faire évoluer la cartographie au fil du temps. Ainsi, la première contribution scientifique (C1) relative à nouvelle cartographie d'utilisation des données a été présentée de ce travail de thèse.

3 Foisonnement des charges BT

Dans ce chapitre, nous allons plonger dans la thématique du foisonnement des charges en BT. Nous commencerons par expliquer le sens du foisonnement des charges, son importance pour le réseau, et une revue bibliographique sur le sujet. Ensuite, nous analyserons les facteurs qui influent sur la variation de ce coefficient et proposerons un modèle de prévision. Notre modèle présente des résultats 50 % meilleurs en moyenne par rapport à une méthodologie classique d'estimation pour notre jeu de données.

3.1 Le foisonnement

3.1.1 Définition et importance du foisonnement

Pour commencer ce chapitre, nous devons parfaitement définir ce qu'est le foisonnement. Le foisonnement est un coefficient qui traduit le fait que les clients du réseau électrique ne consomment pas à leur puissance maximale simultanément. En d'autres termes, nous pouvons dire que ce coefficient mesure la diversité de la demande. Plus la demande est diversifiée, plus la pointe de la demande des clients est espacée dans le temps, et moins elle est diversifiée, moins la pointe de la demande des clients est espacée dans le temps. Ce concept a été défini par *W. Lackie* en 1909 et nommé comme *Diversity Factor*. Mathématiquement, il est formulé par l'équation 3-1 :

$$Diversity\ Factor = \frac{\sum(Pointes\ individuelles)}{Pointe\ du\ groupe} \quad \text{Équation 3-1}$$

L'équation 3-1 met en évidence comment le concept mesure la diversité de la demande. Plus le *Diversity Factor (DF)* est élevé, plus le ratio entre la somme des pointes individuelles de consommation des clients par rapport à pointe totale de ce groupe de clients est élevé, ce qui signifie que la demande est plus diversifiée. Comme il sera discuté dans les paragraphes à venir, dans la littérature, nous trouvons des auteurs parlant du *Coincidence Factor*, qui mesure à quel point la demande est coïncidente et, par conséquent, à quel point les clients consomment leur maximum au même moment. Ce concept est l'exact opposé du *Diversity Factor*, comme le montre l'équation 3-2 :

$$Coincidence\ Factor = \frac{1}{Diversity\ Factor} \quad \text{Équation 3-2}$$

Faire cette mise au point à ce stade vise à rendre la compréhension du lecteur aussi claire que possible, et a été principalement conçue pour le lecteur français qui a déjà entendu le terme Coefficient de Foisonnement (CF), qui est compris de la même manière que le *Coincidence Factor*. Dans la suite de ce chapitre, nous utiliserons le *Diversity Factor* car, il est plus compréhensible de parler de la diversité de la demande que de sa coïncidence. Si nous adaptons l'équation 3-1 au réseau BT, nous obtenons l'équation 3-3 :

$$DF = \frac{\sum(\text{pointes individuelles des clients en aval du Poste})}{\text{Pointe à maille du Poste HTA/BT}} \quad \text{Équation 3-3}$$

Il devient encore plus évident de comprendre que plus la demande est diversifiée, plus la pointe de consommation à maille du poste HTA/BT est basse. En ce qui concerne la planification du réseau, avoir une connaissance plus précise de la pointe de consommation sur le poste implique un dimensionnement plus juste du réseau. De cette manière, la personne en charge du dimensionnement du réseau peut quantifier à quel point les infrastructures du réseau sont rentabilisées, mais aussi leur capacité à accueillir de nouveaux consommateurs, quelles seront les besoins en renforcement du réseau avec l'évolution de la demande, parmi de nombreuses autres considérations. Si nous examinons cela en conjonction avec les transformations du réseau, nous pouvons nous interroger sur la manière dont les nouveaux usages, en plus de l'évolution des comportements des clients, affectent le pic de consommation. Si nous abordons cette question dans un contexte français, nous pouvons penser à l'exemple des véhicules électriques. Au début du premier chapitre, nous avons examiné les données montrant l'augmentation du nombre de véhicules électriques et hybrides immatriculés en France, ainsi que l'augmentation du nombre de points de recharge pour ces véhicules. Ces chiffres, à eux seuls, sont déjà très significatifs et témoignent de la tendance à la croissance de ce nouvel usage. Pour renforcer davantage cette tendance, le parlement européen est parvenu à un consensus sur une nouvelle législation concernant la fin de la vente de véhicules thermiques dans toute l'Union européenne. Cette nouvelle législation stipule qu'à partir de 2035, la vente de nouveaux véhicules thermiques sera interdite. Cette mesure a été adoptée dans le but d'atteindre les objectifs climatiques de neutralité carbone d'ici 2050. Une des raisons pour lesquelles cette mesure entrera en vigueur en 2035 est que la durée de vie moyenne de ces véhicules est d'environ 15 ans, de sorte que d'ici 2050, tous les véhicules en circulation seront électrifiés. Selon les prévisions d'Enedis, la moitié du parc automobile français sera convertie en véhicules électriques d'ici 2035. Comment cela affecte-t-il le GRD ? Cela nous amène rapidement à réfléchir à deux impacts potentiels majeurs. Le premier est le raccordement massif des infrastructures de recharge déployées dans les voies publiques des villes, sur les routes et dans les résidences collectives (habitées par 44 % de la population française). Si on se focalise sur le résidentiel collectif, la loi Climat et Résilience de 2021 met en place de nouvelles dispositions pour accélérer le déploiement d'IRVE (infrastructure de recharge de véhicules électriques) dans l'habitat collectif, grâce un mécanisme de préfinancement. Si l'infrastructure collective est publique le financement peut être « avancé » par Enedis qui est, pour cela, couvert par le TURPE sur la durée de vie de l'ouvrage. Un exemple d'infrastructure publique de bornes de recharge en

parking de résidence collective, également appelée « colonne horizontale » est illustrée par la Figure 3-1.

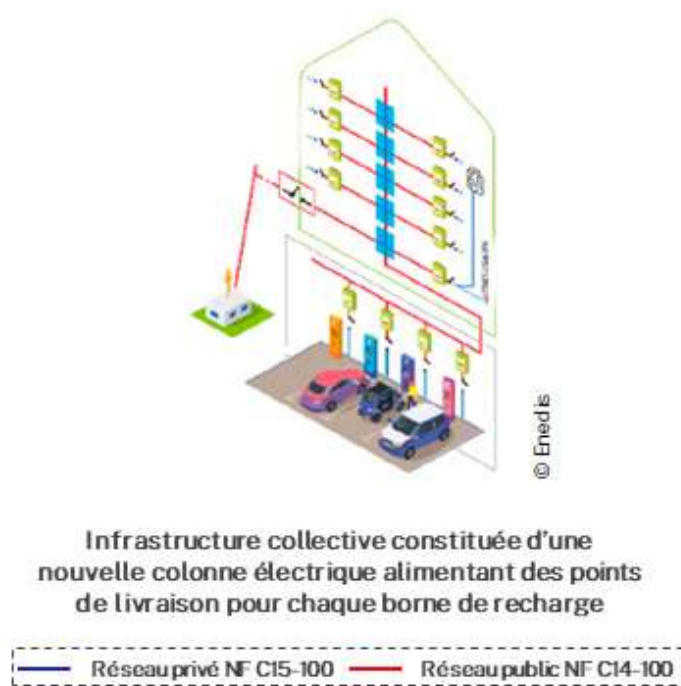


Figure 3-1 : Illustration d'une colonne horizontale

Cela génère le besoin d'études pour comprendre s'il est possible de raccorder ces structures au réseau existant, avec ou sans la nécessité de renforcements, ou s'il est nécessaire de dimensionner de nouvelles installations. Le deuxième impact potentiel concerne la présence d'une nouvelle charge ayant une demande de puissance élevée, avec la pointe de consommation des clients de plus en plus lié aux moments de recharge des véhicules, réduisant ainsi la diversité de la demande. Dans les deux cas, une meilleure connaissance du *Diversity Factor* entraîne une meilleure estimation de la pointe sur le poste HTA/BT, ce qui permet au planificateur d'avoir une meilleure compréhension de la capacité d'accueil du réseau, et par conséquent, de mieux orienter les investissements. Maintenant que nous avons une compréhension de ce qu'est le *Diversity Factor* et de son importance pour le réseau de distribution, nous pouvons plonger dans la littérature scientifique qui traite de ce sujet.

3.1.2 Historique du diversity factor

Comme mentionné au début de ce chapitre, le concept du *Diversity Factor* a été introduit par W. Lackie en 1909. En 1910, H.B. Gear a écrit l'article intitulé : *Diversity Factor* [134]. Il commence son article en nous expliquant que dans la distribution d'électricité dans les grandes villes, la demande maximale d'un jour à l'autre varie en fonction du jour de la semaine et du mois de l'année. Il justifie cela en raison du fait que la durée de la période diurne varie en raison des différences saisonnières, de la compréhension que la caractéristique (commerciale, résidentielle ou manufacturière) du quartier desservi change le modèle de consommation, comme le montrent les graphiques de l'article, et en fonction de la caractéristique de la population desservie. Selon lui, les habitudes des habitants d'un quartier résidentiel d'ouvriers diffèrent de celles des habitants des boulevards, tout comme les besoins des personnes vivant dans des appartements diffèrent de ceux des personnes vivant dans des maisons. Ainsi, selon lui, la combinaison de ces effets fait que :

- la somme des demandes maximales des clients est supérieure à celle des transformateurs,
- la somme des demandes maximales des transformateurs est supérieure à celle du départ moyenne tension,
- la somme des demandes maximales des feeders est supérieure à celle des Postes Source,
- la somme des demandes maximales des Postes Sources est supérieure à celle de la centrale électrique.

Ce phénomène a alors été appelé le *Diversity Factor*. Pour lui, l'étude du DF est d'une grande importance à la fois d'un point de vue commercial et technique. En effet, l'investissement nécessaire par le DSO pour chaque kilowatt de demande maximale déterminait les frais fixes nécessaires lors de la création du système tarifaire. Son étude couvrait un système de courant alternatif qui englobait la génération d'énergie jusqu'au consommateur final. Grâce à cette étude, il a créé un tableau du *Diversity Factor* pour différentes catégories de quartiers (commerciaux, résidentiels, manufacturiers) et pour différents réseaux. Il a observé que le DF était plus faible dans les zones résidentielles où il y avait moins de 10 clients et plus élevé dans les zones résidentielles où il y avait plus de 30 clients.

En 1945, C. Bary, dans son article [135], explore les principaux facteurs qui influencent le foisonnement dans l'utilisation du service électrique. Il commence par expliquer que la manière dont les consommateurs individuels utilisent le réseau et le degré de coïncidence de

cette utilisation avec un groupe de consommateurs sont déterminés par de nombreux facteurs. Parmi ceux-ci, il cite les principaux : les habitudes de la population et les pratiques commerciales et communautaires, le climat et d'autres facteurs climatiques, la conception des équipements et les méthodes utilisées pour contrôler l'utilisation du service. Il discute de manière générale chacun de ces sujets. En ce qui concerne les habitudes de la population, il explique que celles-ci exercent une puissante influence sur le degré de coïncidence de l'utilisation du service électrique par les clients. Il donne des exemples de la façon dont le comportement des personnes dans des situations triviales, telles que l'heure du réveil, jusqu'à des situations extrêmes, comme en période de guerre, affectent le *Coincidence Factor*. Il précise que ces conditions ne sont pas sujettes à une formulation mathématique précise et que leurs effets peuvent être différents pour différentes communautés, soumises à des conditions climatiques, politiques et sociales différentes. De la même manière, il affirme que le climat joue également un rôle puissant dans ce degré de coïncidence, en citant des exemples de la manière dont la consommation change en fonction de différents types d'événements climatiques. Selon lui, ces conditions ne sont pas non plus sujettes à des formulations mathématiques exactes, mais elles sont susceptibles d'être évaluées en fonction de la disponibilité de données météorologiques. Après avoir abordé ce sujet, il aborde la discussion de la relation entre le *Coincidence Factor* et le nombre de consommateurs. Il formule que, pour des consommateurs de taille égale, n'ayant aucune restriction d'utilisation du service ou d'autres caractéristiques anormales d'utilisation du service : « Le degré de coïncidence des demandes maximales d'énergie d'un groupe de consommateurs diminue le long d'une courbe qui ressemble à une hyperbole rectangulaire, commençant à l'unité pour un consommateur dans un groupe et se rapprochant des valeurs asymptotiques de l'hyperbole rectangulaire pour un nombre infini de consommateurs dans un groupe. » En d'autres termes, à mesure que le nombre de consommateurs dans un groupe augmente, le degré de coïncidence des demandes maximales d'énergie diminue, mais il tend vers une valeur limite lorsque le nombre de consommateurs devient infini dans ce groupe. D'après sa formulation, il montre que des résultats statistiquement stables du *Coincidence Factor* peuvent être attendus pour des groupes d'environ 100 consommateurs, et que des résultats présentant une stabilité suffisante sur le plan technique peuvent être trouvés pour des groupes de 30 à 50 consommateurs.

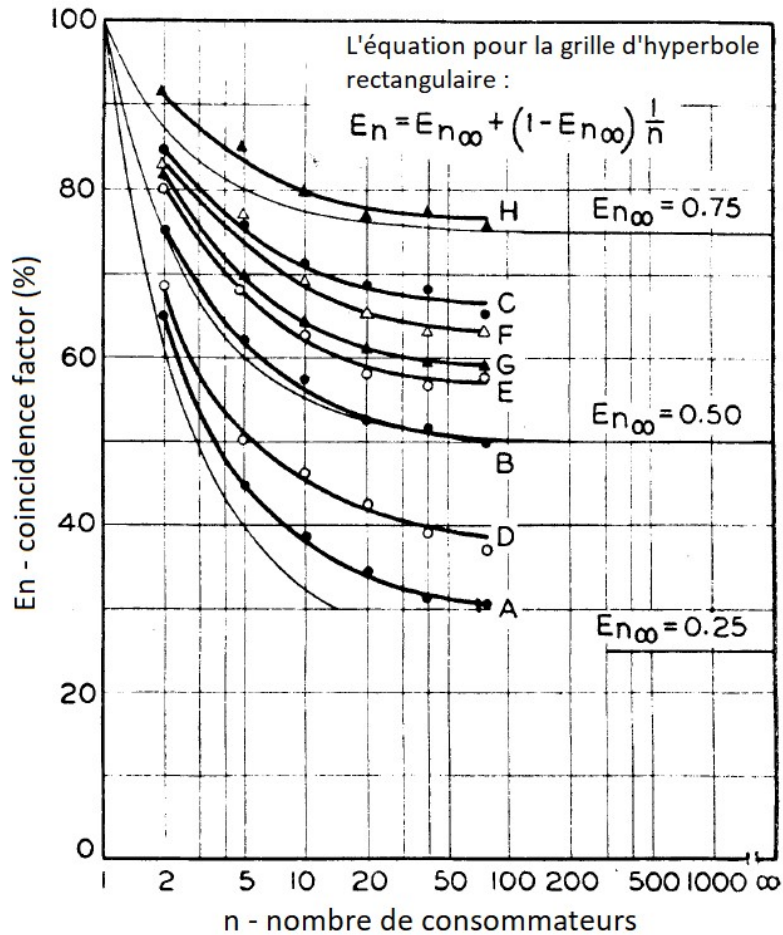


Figure 3-2 : La figure extraite de l'article [135] et traduite en français, qui illustre la tendance probable en relation entre le Coincidence Factor de groupe et le nombre de consommateurs

En 1994, Robert P. Broadwarter [136], utilise les données de la charge horaire annuelle de 299 clients résidentiels pour calculer le *Diversity Factor*. Ces clients ont été regroupés en deux classes en fonction du type de système de chauffage, qu'il soit électrique ou non. Les données ont été organisées en fonction du mois de l'année et du type de jour (jour de semaine ou week-end). Le DF est calculé en fonction du nombre de clients dans le groupe. A partir de ces informations et des données, il formule que le DF moyen dépend de la classe de clients du groupe, de la quantité de clients dans le groupe, du mois et du type de jour. Il explique que ces DF moyens ont été représentés graphiquement en fonction du nombre de consommateurs pour chacune des combinaisons de classe, de mois et de type de jour. Ainsi, un ajustement de courbes (*curve fitting*) a été réalisé pour obtenir les courbes qui expliquent les valeurs de DF calculées en fonction du nombre de clients. La Figure 3-3 est une extraction de l'article [136] illustrant cet ajustement de courbe pour les mois d'août et de février :

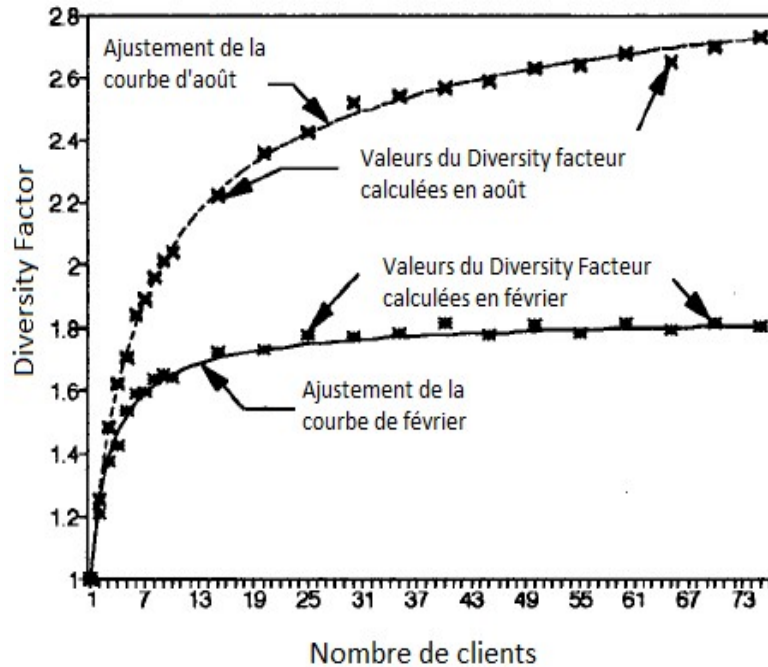


Figure 3-3 : La figure extraite de l'article [136] et traduite en français, qui illustre l'ajustement de courbe pour tracer le DF en fonction uniquement du nombre de clients

En 1998, J. Nazarko a écrit [137]. L'objectif était d'identifier les propriétés statistiques du *Diversity Factor* à partir d'une base de données de clients résidentiels d'*Arkansas Power and Light Company*. Tout d'abord, il a calculé le test d'indépendance des échantillons pour le DF. Les résultats informatiques présentés indiquent que les échantillons utilisés pour calculer le DF mensuel pour différents mois sont indépendants, résultat également trouvé pour le DF quotidien. La deuxième propriété statistique étudiée était le test de normalité, dans le but de déterminer si une distribution normale pouvait approximer la distribution des valeurs du DF pour différents groupes de clients (en ce qui concerne la quantité de clients). Les résultats obtenus montrent qu'il n'y a aucune raison de rejeter l'hypothèse selon laquelle la distribution de DF peut être approximée par une distribution normale.

En 2007, V. P. Chatlani [138], étudie les propriétés statistiques du *Diversity Factor* pour évaluer la charge probabiliste des transformateurs de distribution. Le jeu de données comprenait 685 charges résidentielles, à des intervalles de 15 minutes, provenant d'un projet à Phoenix. Il n'est pas clairement indiqué si toutes les charges résidentielles sont sous le même transformateur de distribution. L'auteur reconnaît également que le DF dépend de plusieurs facteurs, tels que la classe des clients, qu'il varie selon le mois et la saison de l'année, et qu'il n'est pas le même pour différentes régions d'un même pays, mais aucun de ses aspects n'est

approfondi. Dans son analyse, il se concentre sur la saison estivale, car selon lui, c'est la saison qui génère les pics les plus importants dans son jeu de données. Pour calculer le DF, ils ont d'abord identifié la pointe individuelle de ces charges, puis pour calculer la pointe du groupe, ils ont réalisé la somme des profils de charge. Ensuite, un test d'indépendance des échantillons a été effectué pour déterminer s'il était possible d'attribuer une certaine distribution aux DF calculés. Les résultats ont indiqué que les observations étaient aléatoires et indépendantes, de la même manière que dans [137]. Tout comme dans [137], ils ont testé si une distribution normale pouvait bien représenter la distribution du DF pour différentes tailles de groupes de clients. Leurs résultats montrent qu'une distribution Gamma représente mieux la distribution du DF que la distribution normale. A partir de là, l'article se concentre sur la combinaison de la distribution statistique de la demande prévue avec la distribution statistique de la capacité de charge du transformateur afin d'obtenir une évaluation statistique quantifiable de la charge du transformateur et du risque associé à cette charge.

L'article [139], proposé par G.T. Heydt en 2023, vise à discuter des modèles probabilistes et du *Diversity Factor* pour les charges des transformateurs de distribution. Il revisite les résultats exposés dans [138] et, à partir de ses analyses, conclut que le DF de la charge du transformateur de distribution peut être modélisé en tant que quantité probabiliste, et que le DF de la charge du transformateur de distribution peut être modélisé comme le rapport de variables aléatoires normales bivariées corrélées. (Cela fait référence à une paire de variables aléatoires suivant une distribution normale multivariée, ce qui signifie qu'elles sont toutes deux normalement distribuées et qu'il existe une corrélation entre elles. La distribution normale multivariée est caractérisée par une matrice de covariance qui capture à la fois les variances des variables individuelles et les covariances entre elles.) Il étudie également les impacts des recharges de véhicules électriques sur la diversité de la charge. Il nous montre que la prolifération des bornes de recharge des véhicules électriques entraîne un DF individuel du transformateur de distribution plus bas, il faudra donc en tenir compte dans le dimensionnement des transformateurs de distribution.

En 2022, J. Lee [140], a développé un modèle pour estimer la charge de la pointe en utilisant la distribution Généralisée de Valeurs Extrêmes pour modéliser la charge de pointe en tant que variable aléatoire conditionnelle à la consommation moyenne d'énergie. La Distribution Généralisée de Valeurs Extrêmes est une famille de distributions de probabilité utilisée pour modéliser les valeurs extrêmes d'une variable aléatoire. Elle est couramment utilisée dans le domaine de l'analyse des risques, notamment pour estimer les queues de distribution des événements extrêmes, tels que les crues, les tempêtes, les tremblements de terre. L'auteur

de l'article démontre cette modélisation, qui suppose que l'espérance et la variance de la charge maximale évoluent, il a une intuition théorique et donne de bons résultats dans des simulations sur des groupes de charges similaires. De plus, leurs résultats montrent que la distribution Généralisée de Valeurs Extrêmes capture la queue de la distribution mieux qu'une distribution gaussienne.

En 2010, J. Dickert [141] se focalise sur l'investigation des modèles de charge résidentielle pour la planification du réseau. Il mentionne différentes caractéristiques des clients consommateurs, parmi lesquelles la diversité de la demande résidentielle, expliquée par [134]. Ensuite, il approfondit les modèles de charge résidentielle qui estiment le pic de la demande, car c'est le cas dimensionnant en planification des réseaux. Il commence par parler du *Coincidence Factor* et suppose que les charges des clients individuels aux moments de leur pointe suivent une distribution normale. A partir de là, il formule mathématiquement une manière de calculer le *Coincidence Factor*, en supposant que tous les clients ont la même demande maximale, et présente d'autres formulations de calcul plus simples. Ensuite, il présente la formule de Velander représentée par l'équation 3-4, décrite par Velander en 1947 [142-143], qui est largement utilisée dans les pays scandinaves.

$$P_{max1} = k_1 * E + k_2 * \sqrt{E} \quad \text{Équation 3-4}$$

Cette méthode transforme l'énergie annuelle (E) en demande maximale d'un client (P_{max1}). Pour cela, la formule utilise des coefficients (k_1, k_2) calculés empiriquement. Dans l'article [141], l'auteur fournit un tableau avec des exemples de ces coefficients.

A travers ces articles, nous pouvons à nouveau comprendre l'importance du *Diversity Factor* pour la planification des réseaux. Il nous permet d'accéder à la demande maximale d'un ensemble de charges. En transposant cela au réseau BT, notre ensemble de charges est l'ensemble des clients en aval du poste du réseau de distribution, nous permettant ainsi de connaître la demande maximale à maille du poste. Grâce à cela, le planificateur est capable de mieux dimensionner le réseau, comme le montrent à partir d'exemple les articles [138, 139] dans la modélisation de la charge et le choix du transformateur HTA/BT. Nous pouvons également observer que depuis la première description du *Diversity Factor*, les auteurs comprennent et mettent en évidence que la demande maximale des clients dépend de multiples facteurs, dont les principaux sont liés au comportement des clients, à leurs habitudes de consommation, à la catégorie de consommation à laquelle ils appartiennent

(résidentielle, professionnelle ou industrielle) et aux conditions sociales des clients. Ils comprennent que cette demande varie de jour en jour, de mois en mois, de saison en saison, et qu'elle varie en fonction des effets climatiques en général. Cependant, malgré cette compréhension, peu de recherches ont été menées ou de modélisations ont été effectuées sur ce sujet. En général, l'attention s'est principalement concentrée sur la manière dont le DF change en fonction du nombre de clients, sur la modélisation de cet effet et sur la recherche du type de distribution qui représente le mieux ces variations par groupes de clients. Ceci est en partie compréhensible, car il manquait de données suffisamment représentatives pour d'autres phénomènes comme par exemple la température et saisonnalité, et en général, les données les plus facilement accessibles sont le nombre de clients. Cependant, avec l'avancée des compteurs intelligents et la collecte de données plus exhaustive sur d'autres phénomènes, une opportunité s'ouvre désormais pour explorer cette discussion encore peu abordée. C'est pourquoi ce chapitre vise à apporter une contribution scientifique à partir de l'exploitation d'autres variables pour la compréhension et la modélisation du *Diversity Factor*. Pour ce faire, il sera utilisé les données disponibles grâce aux compteurs Linky et Pinky. Avec des données provenant de réseaux réels, il sera possible d'étudier le DF d'une manière non artificielle, étant donné que les groupes de clients en question appartiendront à des réseaux électriques existants. De plus, une autre contribution consistera à développer un modèle de prédiction du DF, dont l'efficacité sera testée par rapport à d'autres modèles de prévision de la pointe à la maille du poste HTA/BT.

Les travaux réalisés dans cette étude ont abouti à la publication d'un article [144] lors de la conférence des distributeurs d'électricité (CIRED), qui a également été présenté sous forme de poster lors de la conférence à Rome en 2023. Comme au moment de la rédaction de ce chapitre, plus de données étaient disponibles, ce chapitre va au-delà de l'article susnommé.

3.2 Périmètre des données

Les données collectées pour cette étude couvrent la période allant du début de juillet 2021 à la fin de juin 2023. Cela nous permet d'obtenir une meilleure représentation des variations météorologiques, ce qui enrichit considérablement notre analyse, car la charge du réseau français dépend largement de la température. Toutes les données utilisées dans cette étude proviennent des bases de données d'Enedis et ne sont pas accessibles au public. Ils sont divisés en 3 catégories : données électriques, caractéristique des clients et données climatiques. Il est nécessaire de préciser que les données électriques et les caractéristiques des clients utilisées dans cette étude sont associées à un panel de clients qui s'appelle « panel réseaux ». Ce sont

des clients ayant donné leur consentement pour l'utilisation de ces données dans le cadre des études sur le réseau, auxquelles cette thèse est intégrée, conformément au règlement général sur la protection des données (RGPD).

Données électriques

La pointe récupérée par les compteurs Pinky de 169 postes HTA/BT situés dans différentes régions de la France, couvrant ainsi la diversité climatique du pays. Ces postes desservent 21186 clients, dont les pointes de consommation et d'énergie ont été récupérées à partir des mesures Linky.

Caractéristique des clients

La catégorie de consommation de ces 21186 clients a été collectée, ainsi que la valeur de puissance souscrite contractuelle. Les catégories de consommation sont les suivantes : Résidentiel Base, Résidentiel HP/HC, Professionnel Base, Professionnel HP/HC, Éclairage Public et Industriels. Les valeurs de puissance souscrite contractuelle sont les suivantes : 3kW, 6kW, 9kW, 12 kW, 15kW et supérieur à 15kW.

Données climatiques

Des données de température et de nébulosité provenant des stations Météo-France les plus proches des postes HTA/BT ont été recueillies. De plus, la zone climatique établie selon la norme RT2012 et la classification du type de territoire (Urbain, Rural, Semi-urbain et Dense) établie par Enedis ont également été collectées.

A partir des données électriques, nous pouvons calculer le DF quotidiennement pour toute la période en utilisant la formule présentée par équation 3-3. Nous pouvons reformuler l'équation de la manière suivante, comme montre l'équation 3-5.

$$DF = \frac{\sum(\text{pointes individuelles des CL et CNL})}{\text{Pointe Pinky}} \quad \text{Équation 3-5}$$

Où *CL* et *CNL* sont respectivement les abréviations des clients Linky et client Non-Linky. À chaque client Non-Linky, un client Linky « miroir » qui fait partie du panel est attribué. Ce

client Linky « miroir » a la même puissance souscrite et la même catégorie de consommation que le client Non-Linky. De cette manière, il est possible d'attribuer une valeur de pointe aux clients Non-Linky.

Étant donné que ces données proviennent des réseaux électriques existants (ensemble de mesures à maille du poste et mesures des clients en aval), il est possible d'étudier comment différents indicateurs, tels que la température, la géographie locale et le comportement des clients, affectent le DF. La Figure 3-4 et 3-5 montre le résultat de l'application de l'équation 3-5 (à des postes choisis au hasard dans l'ensemble de données sur toute la période. On peut observer que la valeur du DF n'est pas constante au fil du temps, présentant des variations saisonnières marquées ainsi que des variations journalières. Cela met en évidence que d'autres facteurs sont essentiels pour expliquer la variation du DF, qui vont au-delà du nombre de clients en aval du poste. Comprendre ces variations et être capable de les prédire peut offrir au planificateur une nouvelle vision du réseau c'est à dire celle liée à ce qui se passe au quotidien, en plus de la vision classique des situations extrêmes. L'utilité de cette vision peut être illustrée à travers des exemples. Prenons le cas où un nouveau client souhaite se raccorder au réseau électrique. Son raccordement doit minimiser la somme des coûts de réalisation des ouvrages de raccordement, tout en respectant les seuils de contraintes électriques et la dimension technico-économique des travaux. Dans un premier temps, si le raccordement du client au réseau existant entraîne une contrainte, des travaux seront réalisés. Les contraintes sont calculées dans les hypothèses de planification d'Enedis pour le raccordement d'un client consommateur dans une situation de 100 % de consommation du réseau et de 0 % de production, où la consommation est la valeur calculée par les modèles de référence. Connaître l'information du DF du poste où le client sera raccordé peut permettre au planificateur d'envisager d'autres solutions pour le raccordement, en plus de la réalisation de travaux. Comme par exemple proposer le raccordement du client avec la possibilité de délestage de puissance en cas de contrainte.

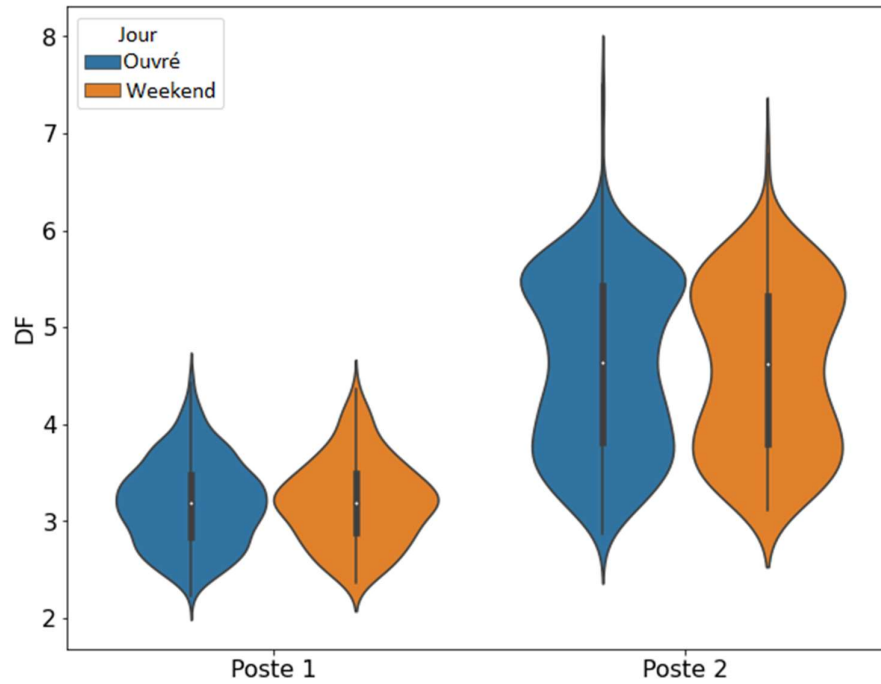


Figure 3-4 : Valeurs de DF dans la période pour les deux postes choisis au hasard – Vision divisée par jour de la semaine (Jour ouvré ou jour weekend)

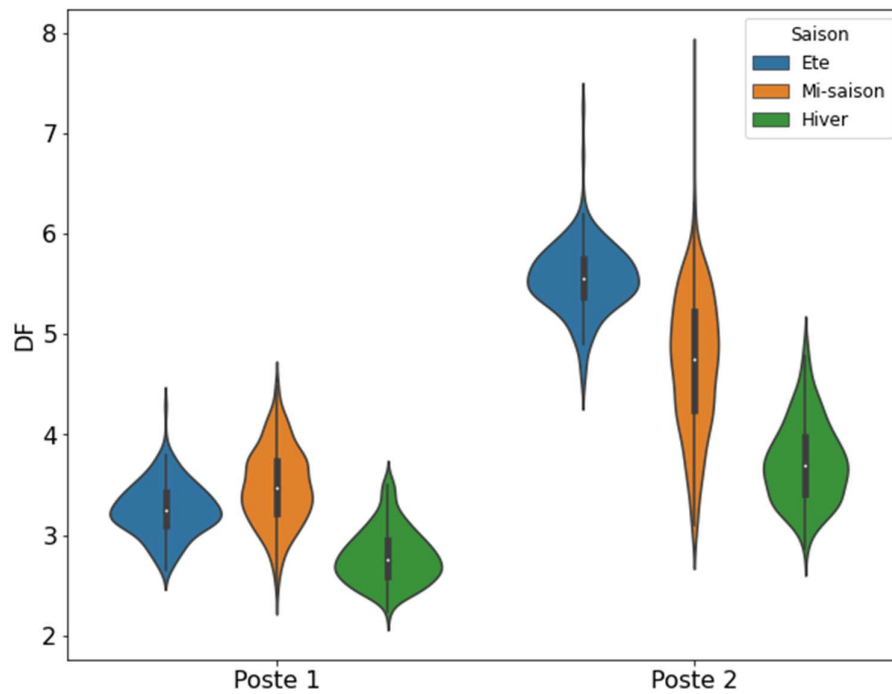


Figure 3-5 : Valeurs de DF dans la période pour les deux postes choisis au hasard – Vision divisée par saison

3.2.1 Construction des indicateurs

A partir des données collectées énumérées ci-dessus, une série d'indicateurs a été construite pour chacun des postes HTA/BT, dans le but de traduire les différences entre les postes. Ce jeu d'indicateurs sera ultérieurement utilisé pour alimenter un modèle de prévision du DF. Ces indicateurs sont décrits dans le Tableau 3-1 :

Catégorie	Indicateurs							
1 - Taux de type de clients	RES Base	RES HP/HC	PRO Base	PRO HP/HC	Eclairage	ENT	-	-
2 - Taux Puissance Souscrite	P3kW	P6kW	P9kW	P12kW	P15kW	P>15kW	-	-
3 - Taux Pointe par type de client	Pmax RES Base	Pmax RES HP/HC	Pmax PRO Base	Pmax PRO HP/HC	Pmax Eclairage	Pmax ENT	-	-
4 - Taux Pointe par Puissance Souscrite	Pmax P3kW	Pmax P6kW	Pmax P9kW	Pmax P12kW	Pmax P15kW	Pmax P>15kW	-	-
5 - Taux Energie par type de clients	Nrj RES Base	Nrj RES HP/HC	Nrj PRO Base	Nrj PRO HP/HC	Nrj Eclairage	Nrj ENT	-	-
6 - Taux Energie par Puissance Souscrite	Nrj P3kW	Nrj P6kW	Nrj P9kW	Nrj P12kW	Nrj P15kW	Nrj P>15kW	-	-
7 - Climat	tempMoy	tempMin	tempMax	nebuMoy	-	-	-	-
8 - Type territoire	Urbain	Semi-Urbain	Rural	Dense	-	-	-	-
9 - Zone Climatique RT2012	1A	1B	1C	2A	2B	2C	2D	3
10 - Jour de la semaine	Weekend	-	-	-	-	-	-	-

Tableau 3-1 : Indicateurs

En tout, 53 indicateurs ont été construits, répartis en 10 catégories, calculés par jour et par poste. Les 6 premières catégories présentent des indicateurs en pourcentage, où la somme des indicateurs pour chaque catégorie représente 100 %. Dans la catégorie 7, on trouve des

indicateurs climatiques, comprenant la température minimale moyenne, la température maximale moyenne et la nébulosité moyenne. La catégorie 9 indique à quelle zone climatique de norme RT2012 [145] appartient le poste. Enfin, la catégorie 10 indique le jour de la semaine de l'observation, si c'est un jour « weekend » ou un jour « ouvré ».

3.3 Modèle de prédiction

Trois étapes principales sont employées dans le développement du modèle et sont illustrées dans la Figure 3-6, à savoir :

- Former un modèle de prédiction en utilisant notre ensemble d'indicateurs (Modèle forêt aléatoire) ;
- Analyser le résultat pour comprendre l'importance des indicateurs pour notre prédiction (Analyse des indicateurs) ;
- Ré-entraîner le modèle en fonction de cette analyse (Modèle actualisé).

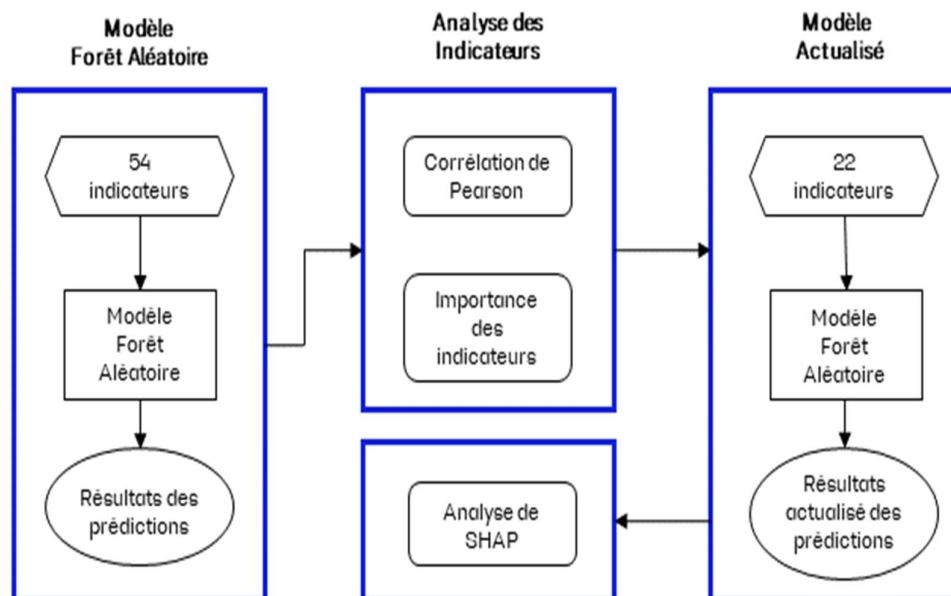


Figure 3-6 : Diagramme de flux des étapes

3.3.1 Modèle de forêt aléatoire

L'algorithme d'apprentissage des forêts aléatoires (en anglais : *random forest*) est la méthode d'apprentissage supervisé choisie pour prédire le DF à partir de notre ensemble sélectionné

d'indicateurs. Cet algorithme est fondé sur l'idée des arbres de décision et a été conçu pour améliorer leurs performances en réduisant le sur-apprentissage, en augmentant la stabilité et en fournissant une meilleure généralisation. Il construit une « forêt » d'arbres de décision. Chaque arbre de décision est formé sur un échantillon aléatoire de données et lors de la construction de chaque nœud de l'arbre, il sélectionne aléatoirement un sous-ensemble des variables disponibles pour prendre la décision. Cela introduit de la variabilité et réduit le risque de sur-apprentissage. Une fois que tous les arbres de décision sont construits, Random Forest effectue une agrégation des prédictions de chaque arbre. Pour la régression, la sortie finale est la moyenne des prédictions de tous les arbres [146]. Cela est montré par la Figure 3-7 :

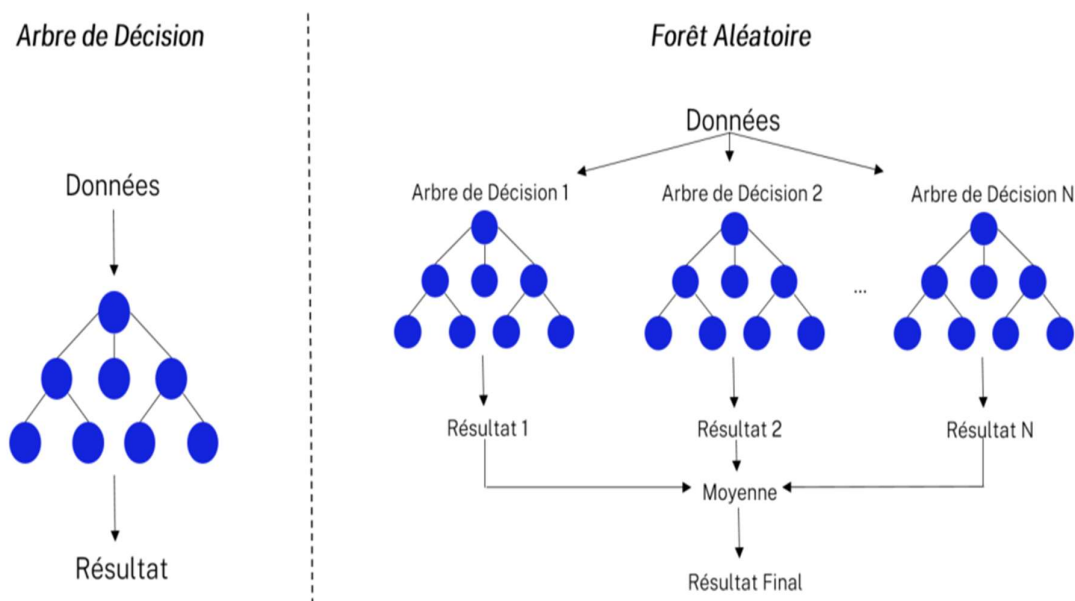


Figure 3-7 : Illustration de le modèle de forêt aléatoire

Cet algorithme est robuste aux valeurs aberrantes, à une réduction de sur-apprentissage par rapport aux arbres de décision individuels. En plus, il permet de calculer l'importance de chaque variable pour la sortie de la prédiction de manière simple. Comme l'objectif est d'utiliser un modèle puissant mais explicatif, pour montrer l'importance des indicateurs dans les variations du DF, la forêt aléatoire est bien adaptée à cette tâche.

Le jeu de données des indicateurs est classiquement divisé en 80 % comme jeu d'entraînement et 20 % comme jeu de test en fonction de la quantité de postes HTA/BT pour éviter tout biais de prédiction. Pour plus de clarté, si la division était fondée sur l'ensemble total des données,

il y aurait un risque que dans une répartition aléatoire entre jeu d'entraînement et jeu de test, le jeu d'entraînement n'inclurait pas l'une des saisons de l'année, ce qui pourrait entraîner un biais dans les prévisions. La division par poste faite aléatoirement génère 135 postes HTA/BT dans le jeu d'entraînement et 34 postes HTA/BT dans le jeu l'ensemble de test.

Pour évaluer les résultats du modèle, la métrique SMAPE (*Symmetric Mean Absolute Percentage Error*) a été choisie. SMAPE est une mesure couramment utilisée pour évaluer la précision des prévisions ou des modèles de prévision. Elle calcule la précision des prévisions en comparant les valeurs réelles et les valeurs prévues de manière symétrique, comme le montre l'équation 3-6 :

$$SMAPE = \frac{1}{N} \times \sum_{n=1}^N \frac{|P_n - R_n|}{(|R_n| + |P_n|)/2} \quad \text{Équation 3-6}$$

Où P est la valeur prévue et R et la valeur réelle. Donc, pour chaque point de données, elle prend la différence absolue entre la valeur réelle et la valeur prévue, puis la divise par la somme des valeurs réelles et prévues. Cette étape est répétée pour chaque point de données, et les résultats sont ensuite moyennés. Le résultat du SMAPE est généralement exprimé en pourcentage, ce qui signifie que plus le SMAPE est bas, plus les prévisions sont précises. Cette métrique d'évaluation des erreurs a été préférée par rapport à d'autres métriques (comme le MAE, le MSE, le RMSE, le MAPE qui sont décrites dans l'annexe 6.1.1) car l'erreur exprimée en pourcentage a été considérée plus facile à interpréter, étant donné que notre modèle cherche à prédire un coefficient qui n'a pas d'unité de mesure. Le SMAPE est robuste aux valeurs petites. En raison de sa symétrie, le SMAPE garantit que la pénalisation de l'erreur ne soit pas sensible aux grandes différences entre la prédiction et la valeur réelle. Cela est dû au fait que sa division est par la moyenne de ces valeurs, ce qui n'est pas le cas du MAPE. De plus, le SMAPE est une métrique robuste et adaptée aux valeurs aberrantes.

En utilisant tous les indicateurs présentés dans la section précédente, les prédictions du modèle ont un SMAPE de 16 %. L'importance de cette prévision réside dans les postes HTA/BT où il n'y a pas de compteurs installés dans la maille du poste. En étant capable de prédire la valeur de DF (à partir d'indicateurs liés à la part de la consommation proviennent des différents types clients, et aux caractéristiques météorologiques et géographiques de la région), nous pouvons estimer la puissance maximale consommée dans la maille du poste HTA/BT. Avec cette estimation, les GRD peuvent mieux orienter les investissements du réseau

pour mieux accueillir de nouvelles charges consommatrices dans le réseau BT, comme c'est le cas des IRVE. La Figure 3-8 présente un graphique de dispersion montrant les prédictions en fonction des valeurs de DF.

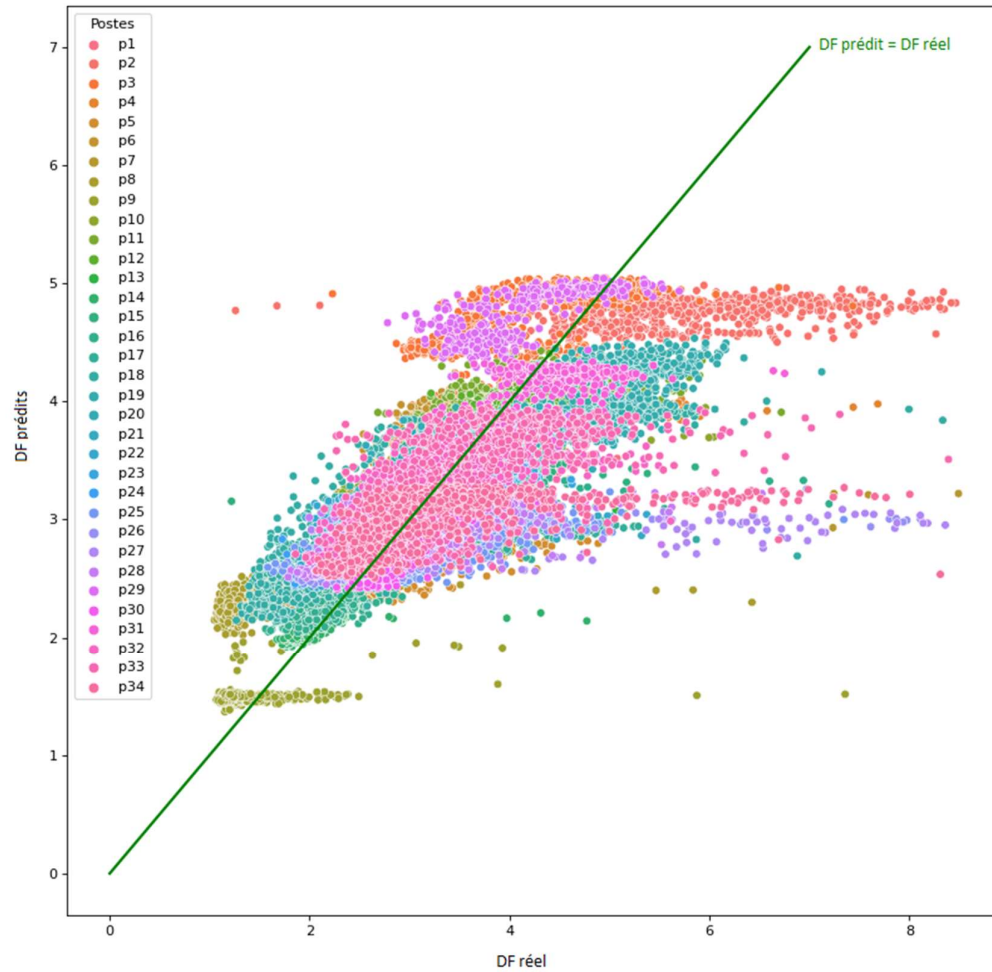


Figure 3-8 : Nuage de points – DF prédit x DF réel

On peut observer que le modèle a des difficultés à prédire des valeurs de DF plus élevées. Cela est logique car il y a peu de valeurs dans cette plage dans notre collection de données, ce qui signifie que le modèle a moins d'exemples d'entraînement. De plus, nous avons pu observer que les erreurs les plus importantes dans nos prédictions sont liées à quelques postes pour lesquels la présence de clients industriels est importante.

3.3.2 Analyse des indicateurs

Nous notons que certains de nos indicateurs semblent exprimer le même type d'information. Il a été possible de confirmer cela en utilisant la méthode de corrélation de Pearson par paires. Il s'agit d'une mesure statistique utilisée pour évaluer la relation linéaire entre deux variables quantitatives. Les indicateurs liés au taux de type de clients (catégorie 1 et 3) sont fortement corrélés de manière positive. Il en va de même pour les indicateurs liés au taux de puissance souscrite (catégorie 2 et 4). Ces corrélations sont montrées par la carte de chaleur de la Figure 3-9

Figure 3-9 : Carte de chaleur des indicateurs des catégories 1 à 4.

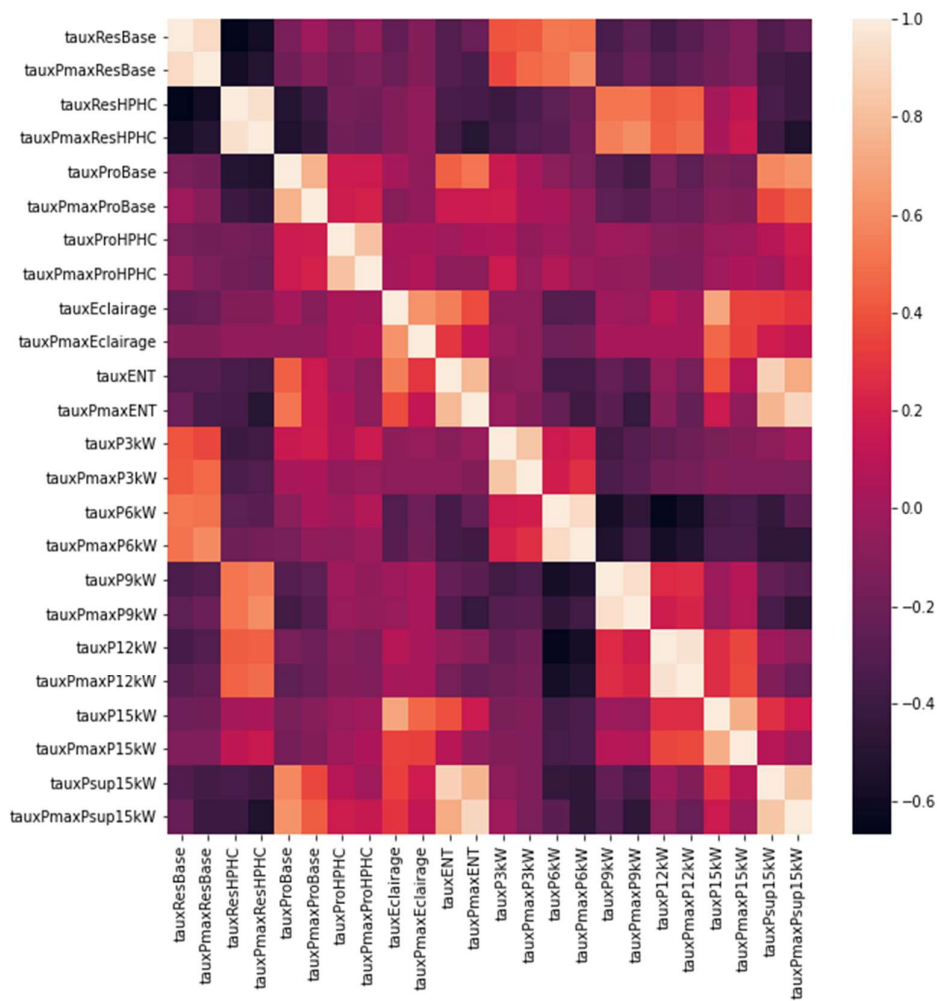


Figure 3-9 : Carte de chaleur des indicateurs des catégories 1 à 4

Par conséquent, ont été conservées uniquement les informations sur le taux de pointe dans les deux cas. La raison derrière cela est le fait que, pour les clients industriels, nous pouvons

avoir des postes HTA/BT avec un seul client industriel, mais ce client peut représenter une part dominante de la puissance soutirée sur le poste. Donc, les douze caractéristiques de la catégorie 1 et 2 (RES Base, RES HP/HC, PRO BASE, PRO HP/HC, Eclairage, ENT, P3kW, P6kW, P9kW, P12kW, P15kW, P>15kW) ont ainsi été écartées par l'utilisation des corrélations présentées par la Figure 3-9. La carte de chaleur pour l'ensemble des indicateurs est présentée dans annexe 6.1.2.

La prochaine étape consiste à examiner l'importance des indicateurs du modèle. La boîte à outils de la forêt aléatoire de la bibliothèque *scikit-learn* de Python offre la possibilité d'analyser l'importance des caractéristiques. Cette analyse est connue sous le nom d'Importance de Gini [147]. Cette métrique est fondée sur le concept de l'impureté de Gini, comme montre les équations 3-7 et 3-8 :

$$IC_{j,m}^{Gini} = IG_m - IG_l - IG_r \quad \text{Équation 3-7}$$

$$IG_m = \sum_{k \in K} \widehat{p}_k(m) * (1 - \widehat{p}_k(m)) \quad \text{Équation 3-8}$$

Où IC est l'importance de la caractéristique, IG est l'impureté de Gini, m est un nœud donné dans un arbre de décision unique qui effectue une division sur la caractéristique j, l et r sont les nœuds enfants du nœud m, $\widehat{p}_k(m)$ représente la fréquence de la classe k dans le nœud m. Lors de la construction d'un arbre de décision, *scikit-learn* utilise l'Importance de Gini pour évaluer l'importance de chaque caractéristique à chaque nœud de l'arbre. Plus précisément, il mesure comment chaque caractéristique contribue à réduire l'impureté dans l'ensemble de données. Plus la réduction de l'impureté est grande grâce à une caractéristique, plus cette caractéristique est considérée comme importante. Un exemple simple de l'impureté de Gini est illustré par la Figure 3-10. Cet exemple illustre l'importance des critères pour déterminer si une personne sera embauchée pour un poste ou non.

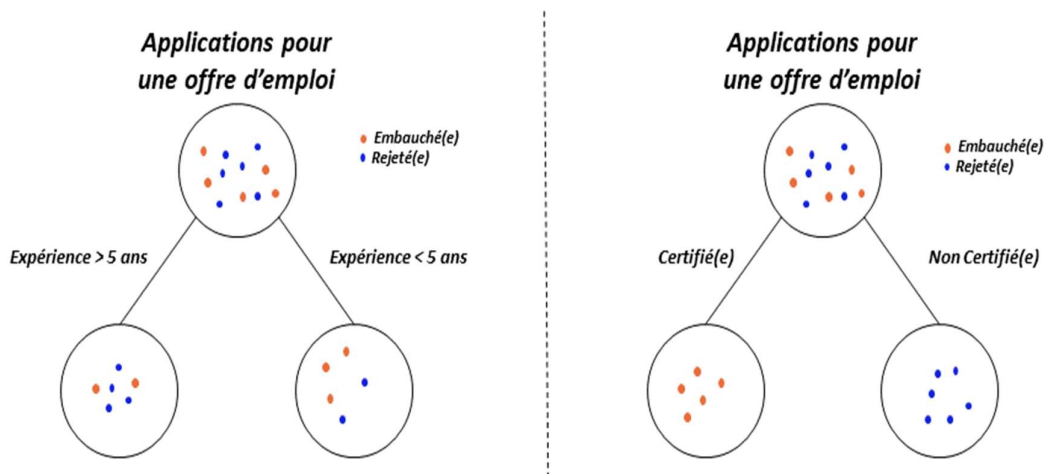


Figure 3-10 : Illustration du concept d'impureté de Gini

Dans cet exemple de la Figure 3-10, nous pouvons voir que la possession ou non d'une certification influe davantage sur la réduction de l'impureté de Gini que l'expérience. Ainsi, nous pouvons dire que la variable de certification est plus importante dans ce cas. Avec ce concept plus clair, nous pouvons calculer l'importance des indicateurs pour notre modèle de prédiction du DF. Le résultat de cette analyse montre que les catégories d'indicateurs 5, 6 et 9 (Nrj RES Base, Nrj RES HP/HC, Nrj PRO BASE, Nrj PRO HP/HC, Nrj Eclairage, Nrj ENT, Nrj P3kW, Nrj P6kW, Nrj P9kW, Nrj P12kW, Nrj P15kW, Nrj P>15kW, 1A, 1B, 1C, 2A, 2B, 2C, 2D, 3) peuvent être écartés car leur importance pour les résultats du modèle est faible. L'analyse des indicateurs nous laisse avec les 21 indicateurs appartenant aux catégories 3, 4, 7, 8 et 10, comme indique par le Tableau 3-2.

Catégorie	Indicateurs					
	Pmax RES Base	Pmax RES HP/HC	Pmax PRO Base	Pmax PRO HP/HC	Pmax Eclairage	Pmax ENT
3 - Taux Pointe par type de client						
4 - Taux Pointe par Puissance Souscrite	Pmax P3kW	Pmax P6kW	Pmax P9kW	Pmax P12kW	Pmax P15kW	Pmax P>15kW
7 - Climat	tempMoy	tempMin	tempMax	nebuMoy	-	-
8 - Type territoire	Urbain	Semi-Urbain	Rural	Dense	-	-
10 - Jour de la semaine	Weekend	-	-	-	-	-

Tableau 3-2 : Indicateurs retenus après analyse

3.3.3 Actualisation du modèle

Maintenant, nous sommes en mesure de recalculer les prédictions du modèle en utilisant les résultats des analyses des indicateurs. Le même ensemble de test de 34 postes HTA/BT a été utilisé pour voir l'impact des modifications apportées à nos données d'entrée. Ainsi, il est possible de comparer les résultats des modèles. Le modèle est ensuite ré-entraîné et les prédictions nous donnent le résultat présenté dans le Tableau 3-3 :

	Modèle initial	Modèle Actualisé
SMAPE (%)	16	12.9

Tableau 3-3 : Résultats des modèles

Nous pouvons effectuer une analyse Shapley additive (SHAP) pour interpréter la contribution de chaque indicateur à la sortie du modèle. Le SHAP repose sur le concept de la valeur de Shapley, emprunté à la théorie des jeux. Il attribue une contribution à chaque indicateur d'une entrée donnée, en mesurant son impact sur la prédiction finale du modèle. Pour ce faire, il considère toutes les permutations possibles des indicateurs et calcule la différence entre les prédictions avec et sans chaque indicateur. Ces différences sont ensuite moyennées pour obtenir une attribution globale [148]. Pour mieux comprendre les graphiques des figures 3-11 et 3-12, nous devons interpréter les valeurs de SHAP sur l'axe des ordonnées comme des écarts relatifs par rapport à une valeur moyenne de prédiction du modèle. Ainsi, des valeurs positives de SHAP indiquent une augmentation de la sortie du modèle par rapport à cette prédiction moyenne, ce qui dans notre cas signifie une augmentation de la valeur de DF. De même, des valeurs négatives de SHAP indiquent une diminution de la sortie du modèle par rapport à cette prédiction moyenne, ce qui dans notre cas signifie une diminution de la valeur de DF. La Figure 3-11 montre comment la température quotidienne moyenne contribue à la variation du DF pour tous les postes de notre ensemble de test. On peut observer que les températures plus élevées contribuent à augmenter le facteur de diversité. En revanche, lorsque la température diminue, elle contribue à la diminution de DF.

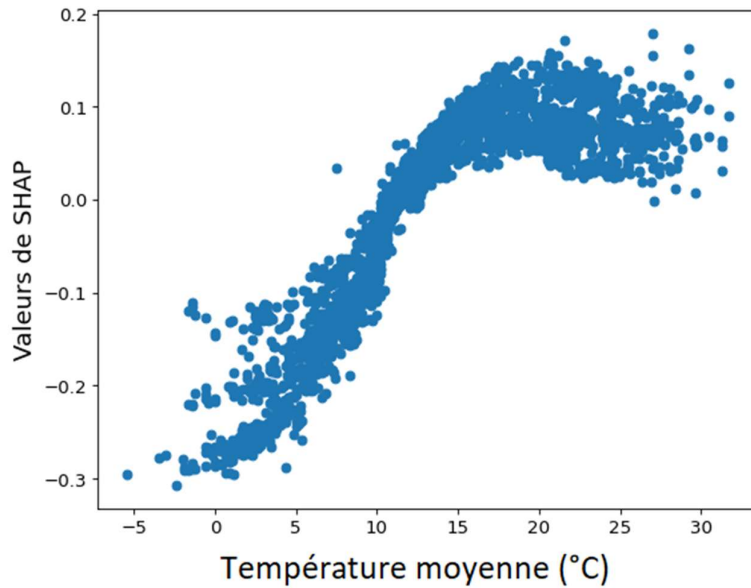


Figure 3-11 : Contributions de la température moyenne pour les variations du DF

Cela peut s'expliquer par le fait que la consommation en basse tension dépend de la température. En France, les clients résidentiels sont équipés de systèmes de chauffage et de chaudières qui ont tendance à être allumés aux mêmes horaires. De cette manière, la demande de pointe individuelle et la demande de pointe agrégée sont proches les unes des autres, réduisant ainsi le facteur de diversité. La Figure 3-12 présente le taux de pointe total quotidien par type de client résidentiel HP/HC. Les clients résidentiels HP/HC sont ceux qui ont un tarif énergétique fondé sur l'heure de la journée où l'énergie est consommée. Il y a les heures pleines où le tarif est plus cher et les heures creuses où le tarif est moins cher. Normalement, pour ces clients, la consommation est plus élevée pendant les heures creuses. On peut voir que si une grande partie de la consommation est due à ces clients, cela contribue à abaisser le DF, ce qui correspond au fait que leurs pointes de consommation sont proches dans le temps.

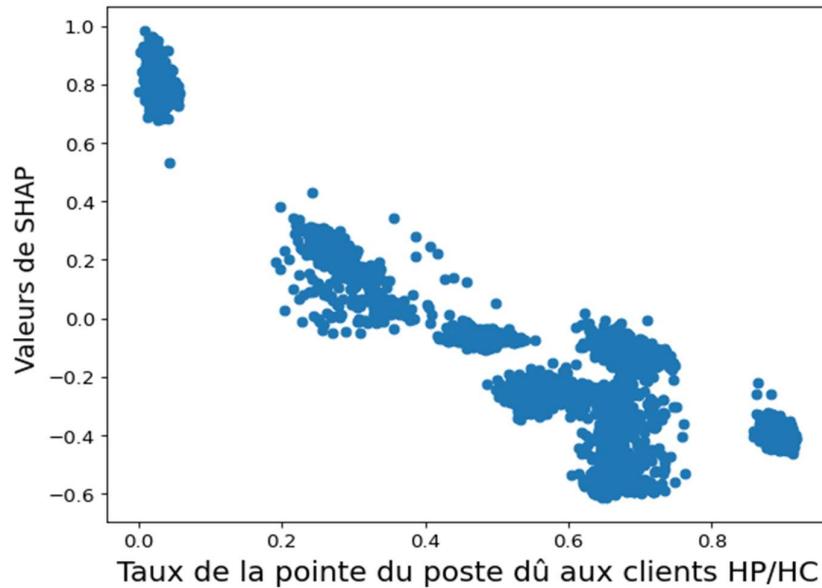


Figure 3-12 : Contribution des clients HP/HC pour les variations de DF

Avec cette analyse, nous pouvons comprendre que le comportement des indicateurs en sortie du modèle est cohérent avec les comportements classiques du réseau français. Cela garantit la qualité du modèle et de ses prévisions.

3.4 Comparaison avec des méthodes existantes

La prochaine étape pour évaluer la qualité de nos résultats de prédiction consiste à les comparer à d'autres méthodes. Cette section présente une comparaison entre les résultats obtenus par une méthodologie classique (appelée par la suite du chapitre comme méthode classique) de dimensionnement des pointes des postes HTA/BT inspiré de la norme NF C14-100 [150] et ceux de notre modèle. Pour ce faire, nous utilisons notre prédiction de DF pour calculer la pointe des poste HTA/BT, comme indiqué par l'équation 3-5. En revanche, la méthode classique se base sur les informations classiques dont le planificateur dispose sur son réseau, telles que le nombre de clients en aval du poste HTA/BT et leurs puissances souscrites respectives. L'estimation de la pointe du poste à partir de la méthode classique est déterminée par l'équation 3-9 :

$$PP = CP * \sum P_{sous} \quad \text{Équation 3-9}$$

Où, PP est la pointe du poste HTA/BT, CP est un coefficient de pondération fondé sur le nombre total de clients en aval du poste et P_{sous} est la valeur de puissance souscrite de ces clients. Les valeurs de CP sont calculées de manière à obtenir le SMAPE minimal entre PP et les valeurs de pointe réelles, mesurées par Pinky, pour l'intervalle de nombre de clients sélectionné. Ce problème d'optimisation est formulé par les équations 3-10 et 3-11 :

$$CP \rightarrow \text{minimisation}(\text{Erreur}_{CP}) \quad \text{Équation 3-10}$$

$$\text{Erreur}_{CP} = \text{SMAPE}(\text{Pointe Réelle} | PP) \quad \text{Équation 3-11}$$

Ce problème d'optimisation est résolu à l'aide de la bibliothèque `scipy` de Python en utilisant la fonction `optimize.minimize` avec la méthode '*Nelder-Mead*'. Cette méthode a été choisie car elle présente une bonne robustesse et est moins sensible aux conditions initiales. Pour se rapprocher de la littérature qui suggère que le nombre de clients est fondamental pour la définition du DF et par conséquent pour la pointe du poste, le calcul de CP est réalisé par tranches de nombre de clients. Cela va dans le sens de la méthodologie classique présentée dans la norme [150]. Les tranches de nombre de clients ont été choisies par paquet de 50 clients afin que chaque tranche contienne au moins 10 postes HTA/BT. Les valeurs de CP trouvées pour mon ensemble de données sont présentées dans le Tableau 3-4 :

Nombre de Clients	Valeur de CP
< 50	0.13
50 – 99	0.12
100 – 149	0.11
150 – 249	0.09
> 249	0.07

Tableau 3-4 : Valeurs de CP par tranche de nombre de clients

La Figure 3-13 présente les résultats de cette comparaison pour les 34 postes HTA/BT du jeu de test :

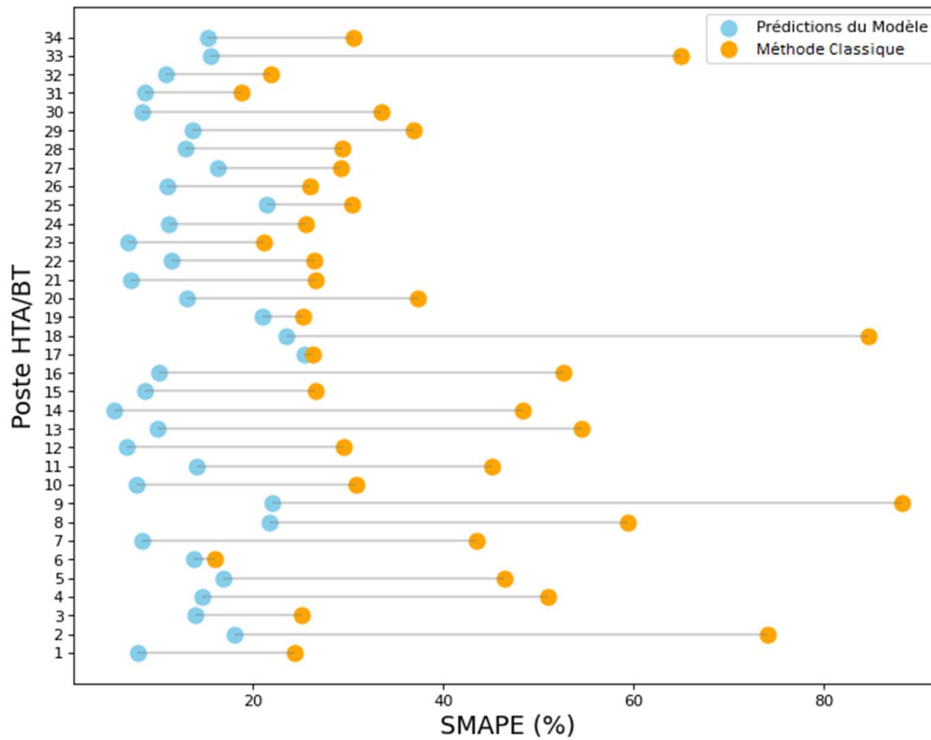


Figure 3-13: Comparaison des résultats pour les 34 postes HTA/BT

Le modèle de prédiction donne de meilleurs résultats dans les 34 postes HTA/BT. La différence moyenne entre les résultats du modèle de prédiction et les résultats de la méthode classique est de 20 % pour cet ensemble de jeu de test. Pour approfondir cette comparaison, nous avons décidé de l'étendre à 50 jeux de tests aléatoires de 34 postes parmi les 169 postes de notre jeu de données. Cela signifie que 50 fois, nous avons divisé notre jeu de données en jeu d'entraînement et jeu de test de manière aléatoire, nous avons entraîné un modèle avec les mêmes indicateurs énumérés précédemment, puis nous avons comparé les résultats obtenus avec la méthode classique. Pour chacun de ces 50 jeux de tests, une moyenne de l'erreur SMAPE a été calculée, et elle est présentée sous forme d'histogramme dans la Figure 3-14 :

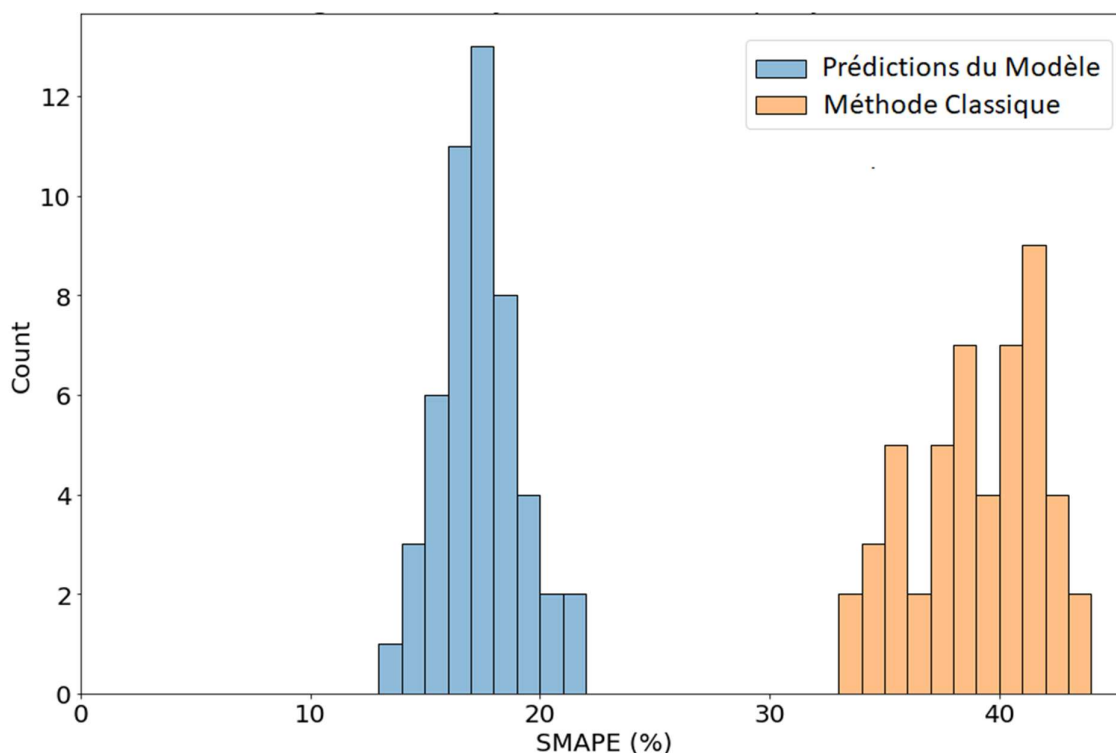


Figure 3-14 : Histogramme Moyenne de SMAPE par jeu de test

L'histogramme montre la différence entre les résultats des prédictions du modèle et les résultats de la méthode classique. En moyenne, les résultats du modèle de prédiction ont environ 50 % d'erreur en moins en comparaison de la moyenne des résultats estimés la méthode classique pour ces 50 ensembles de test. Cela nous montre la valeur ajoutée potentielle de notre modèle pour fournir une estimation plus précise de la puissance maximale consommée dans la maille d'un poste HTA/BT. Information qui est d'une importance extrême dans cet horizon 2035 où la vente de véhicules thermiques sera interdite en Europe, augmentant ainsi le nombre de VE et d'IRVE. C'est également important dans un contexte général d'électrification accrue.

3.5 Conclusions

Comme vu dans le chapitre 1, nous sommes insérés dans un contexte de nouveaux usages du réseau BT. Ces nouveaux usages, tels que les VE, insérés dans ce contexte d'électrification accrue, obligent le GRD à se préparer à accueillir une charge consommatrice plus importante dans le réseau BT. Cela, associé au déploiement massif de compteurs intelligents fournissant des données sur le réseau BT, nous permet d'imaginer des solutions pour mieux accueillir les charges du réseau. Comme vu dans le chapitre 2 à travers la cartographie, les principaux articles (les plus cités à partir de notre requête de recherche WoS) n'abordent pas cette thématique de la capacité d'accueil du réseau. Cela nous a motivés à approfondir ce sujet à travers la contribution (C2). Le *Diversity Factor* (coefficient de foisonnement) est une mesure de la diversité de la consommation du réseau qui est extrêmement importante pour le GRD afin d'estimer la puissance maximale consommée dans la maille du poste HTA/BT. Nous avons utilisé les données des compteurs intelligents Linky pour améliorer l'évaluation des variations du *Diversity Factor*. En utilisant des données de réseau réelles, il a été possible d'étudier et de comprendre comment les indicateurs locaux (comme les caractéristiques météorologiques et géographiques) et globaux (à partir d'indicateurs liés à la part de la consommation provenant des différents types clients) affectent le DF et d'utiliser cette connaissance pour développer un modèle d'apprentissage supervisé permettant de prédire cette valeur. Nos résultats et notre comparaison facilitent la visualisation de l'importance des données des compteurs intelligents pour la planification des réseaux. Nous avons pu obtenir un modèle avec une réduction d'erreur d'environ 50 % d'erreur sur notre ensemble de jeu de test, dans l'estimation de la charge de pointe dans la maille du poste HTA/BT en comparaison avec une méthodologie classique. Fournir une meilleure évaluation du *Diversity Factor* dans les réseaux basse tension peut permettre une meilleure estimation de la capacité d'accueil et, par conséquent, minimiser / retarder / éviter les investissements dans le réseau. Ainsi, la deuxième contribution scientifique (C2) relative à estimation et analyse du foisonnement des charges du réseau BT a été présentée de ce chapitre.

4 Modélisation des charges BT à partir de la désagrégation

Le modèle de désagrégation estime la courbe de charge des clients d'un réseau BT en désagrégeant la courbe de charge mesurée au niveau du Poste HTA/BT. Comme il n'y a rien de similaire dans la littérature, notre modèle est novateur et disruptif. Nous avons formulé deux modèles de désagrégation, l'un pouvant être mieux utilisé pour des applications en HTA et l'autre pouvant être plus approprié en BT. De cette manière, avec une estimation des comportements de consommation quotidiens des clients basse tension, nous pouvons envisager des solutions qui améliorent la capacité d'accueil du réseau.

4.1 Méthode de désagrégation : concept et état de l'art

Concernant les modèles des charges du réseau BT, ils estiment classiquement les courbes des consommations des clients BT avec une certaine résolution temporelle. En appliquant ces modèles à tous les clients en aval d'un poste HTA/BT, l'agrégation de toutes les courbes estimées ainsi que les pertes du réseau permettent d'estimer la courbe de charge du poste en question. Cela est illustré par la Figure 4-1 :

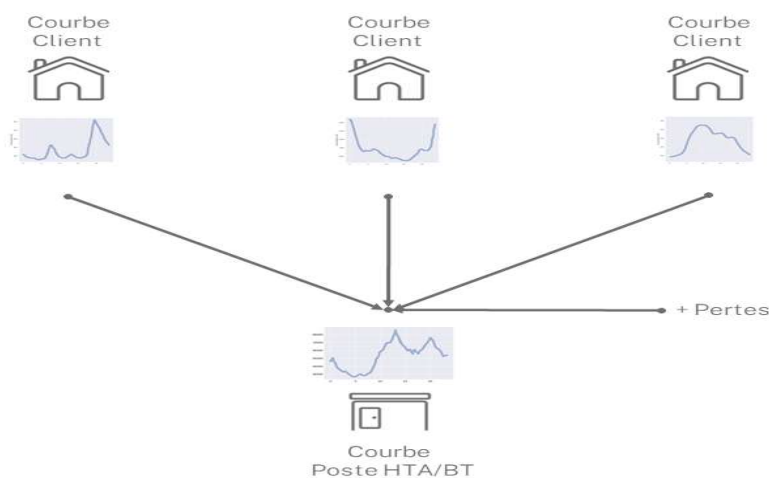


Figure 4-1 : Exemple d'obtention de la courbe de charge d'un poste avec le modèle BT actuel

Avec le développement massif du parc de compteurs communicants, nous pouvons imaginer que les données des profils de consommation des consommateurs seront disponibles. Cependant, l'exploitation de ces données nécessite une certaine attention et prudence. Dans l'Union européenne, les données de profil de consommation des consommateurs provenant des compteurs intelligents sont considérées comme des données personnelles par le Règlement général sur la protection des données. Par conséquent, le développement de nouveaux profils dans un scénario d'utilisation du réseau en évolution devient extrêmement important pour les gestionnaires de réseau de distribution

Grace à un système de comptage type Pinky, voir section 1.2, les données de la courbe de charge du poste sont disponibles pour les postes équipés de ces compteurs. Cette mesure peut ainsi être utilisée pour faciliter la modélisation des courbes de charge des clients BT. Étant donné que la courbe de charge Pinky est relevée quotidiennement, cela permet d'envisager une solution capable de répondre aux enjeux d'avoir un modèle qui estime la courbe réalisée quotidiennement, et non seulement dans une situation dimensionnante.

L'idée est de développer une méthode qui consiste à faire l'inverse de ce qui est actuellement fait, comme par exemple la RECOFLUX [151, 152]. Cela signifie partir de la courbe de charge du poste HTA/BT et la désagréger en courbes de charge individuelles des clients, comme illustré par la Figure 4-2. L'idée de ce modèle est d'être capable de réaliser des désagréations quotidiennes, ce qui nous permettrait d'accéder à la courbe à maille client chaque jour. La courbe client est la donnée d'entrée des outils de calcul du réseau, qui nous permettent de calculer les transits de puissance et de tension dans le réseau. Avec cette donnée de la courbe quotidienne générée par le client, nous pourrions envisager de nouvelles solutions de dimensionnement du réseau, en augmentant sa capacité d'accueil mais aussi en validant les impacts des activations des flexibilités.

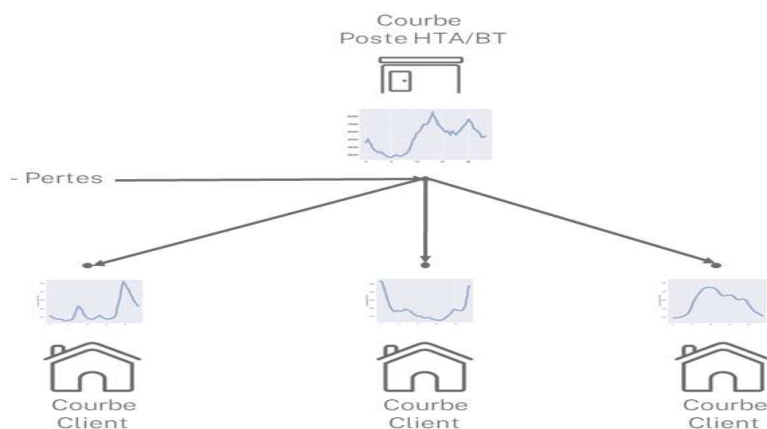


Figure 4-2 : Exemple d'obtention des courbes à la maille client à partir du modèle de désagrégation

En gardant cela à l'esprit, une recherche dans la littérature sur la modélisation de la charge est nécessaire pour déterminer si des solutions similaires ont déjà été créées. De telles solutions pourraient nous aider dans le développement du modèle, et ce modèle pourrait se concentrer sur la résolution de problématiques non explorées. Mais, comme décrit ci-dessous dans la section 4.1.1, rien de similaire à de notre proposition n'a été trouvée. Cela fait de notre modèle une contribution novatrice, fournissant un point de départ pour de nouvelles recherches.

4.1.1 Revue Bibliographique

Pour mieux comprendre où notre type de modèle s'insère, nous commençons cette recherche bibliographique en examinant les différents types de modèles de consommation électrique.

T. Barbier, dans sa thèse [153], a réalisé une classification des différents types de modèles de consommation électrique. Cette classification repose sur deux facteurs. Le premier facteur est une grille de lecture en 3 niveaux de consommation, comprenant le niveau individuel du client et de ses appareils, le niveau agrégé par type de client et par type d'appareil, ainsi que le niveau agrégé du client par zone spatiale et de ses appareils. Le deuxième facteur est une grille de variables influentes sur la consommation, comprenant le parc d'appareils, la technologie de ces appareils, le comportement d'utilisation, le niveau socio-économique, la météorologie, les variables relatives au lieu de consommation, l'abonnement du client et le marché de l'électricité. Ainsi, il est arrivé à classer 6 catégories de modèles différents :

- Modèles fondés sur les consommations agrégées d'une zone ;
- Modèles fondés sur les consommations par type de client ;
- Modèles fondés sur la consommation par usage ;
- Modèles fondés sur la consommation par type de client et par usage ;
- Modèles fondés sur la consommation par client ;
- Modèles fondés sur la consommation par appareils.

Modèles fondés sur les consommations agrégées d'une zone :

Comme l'explique l'auteur dans [153], cette catégorie englobe les modèles dont l'échelle est le niveau de consommation électrique agrégé par zone spatiale. Ce sont des modèles qui se concentrent essentiellement sur la composante temporelle de la consommation et se retrouvent classiquement lors des études de prévision ou « *forecasting* ».

Modèles fondés sur les consommations par type de client :

Cette catégorie englobe les modèles dont l'échelle est le niveau de consommation par type de client. Ce type de modélisation produit des courbes de charge différentes pour chaque type de client (résidentiel, industriel, etc.), et les clients du même type auront des courbes identiques, recalées en énergie. Cette catégorie peut être subdivisée en fonction des données d'entrée du modèle. Il existe des modèles qui utilisent les données de consommation individuelle des clients comme entrée, ce qui, selon l'auteur, est utilisé par de nombreux GRD pour planifier le réseau, comme c'est le cas du modèle Bagheera d'Enedis. Il existe également des modèles qui utilisent les données de consommation agrégées par zone en tant qu'entrée, comme c'est le cas du modèle développé par l'auteur [153].

Modèles fondés sur la consommation par usage :

Selon l'auteur [153], un exemple de ce type de modèle est celui qui décompose les usages thermosensibles des usages non thermosensibles à partir des courbes de charge de différentes zones. Il illustre l'exemple de la thermosensibilité en France, ce qui permet au GRT et aux GRD de comprendre l'influence de l'aléa météorologique sur le dimensionnement du réseau.

Modèles fondés sur la consommation par type de client et par usage :

Cette catégorie regroupe essentiellement des modèles pour le secteur résidentiel qui sont fondés sur la consommation par type d'utilisation d'équipements du secteur. Cette catégorie comprend des modèles tels que ceux de [156-158], qui consistent à désagréger les profils de consommation électrique avec des informations sur le taux de possession d'appareils et éventuellement avec des données socio-économiques. Ce sont des modèles de type « *bottom-up* » et « *top-down* ».

Modèles fondés sur la consommation par client :

Selon l'auteur [153], les modèles appartenant à cette catégorie utilisent des données provenant de compteurs intelligents pour effectuer des prédictions de consommation électrique à court terme pour des clients individuels.

Modèles fondés sur la consommation par appareils :

Cette catégorie vise à estimer la consommation à différents niveaux à partir de l'agrégation de la consommation des appareils. Nous pouvons donner comme exemple les modèles NILM. Les modèles NILM cherchent à décomposer la consommation individuelle des clients par usage en utilisant des mesures individuelles des clients. L'idée consiste à avoir une résolution temporelle des mesures suffisamment fine pour pouvoir capturer les signatures énergétiques des différents appareils.

A partir de cette analyse de la classification des différents types de modèles de consommation électrique fait en [153], nous pouvons conclure que les modèles de type « *bottom-up* » et « *top-down* » sont ceux qui se rapprochent le plus de l'idée initiale de modèle de désagrégation.

Cependant, l'échelle à laquelle nous souhaitons construire le modèle de désagrégation est celle du poste HTA/BT afin de reconstruire la courbe de charge de chaque client BT qui y est rattaché. Cela n'est pas clairement représenté dans cette classification. Les courbes de charge des clients sont les entrées des outils du GRD pour la réalisation des études de planification du réseau. Ainsi, les estimer de manière précise devient une tâche nécessaire dans ce contexte d'évolution des usages du réseau. L'utilisation de méthodologies de décomposition des usages telles que les « *bottom-up* » s'avère extrêmement dépendantes de mesures avec une résolution temporelle fine. Les GRD n'ont pas nécessairement ce degré de précision dans leurs mesures. Notre modèle de désagrégation échappe à cette situation, ce qui le rend plus accessible.

Pour approfondir notre compréhension de ce sujet, nous avons effectué une analyse bibliométrique le 4 août 2023, en nous appuyant sur la réputée base de données Web of Science (WoS). Cette recherche a permis d'identifier 47 publications « pertinentes ». Toutefois, il convient de nuancer en soulignant que bien que ces travaux abordent des questions relatives à la désagrégation de courbes de charge électrique dans les habitats, ce qui est désigné dans la littérature comme des modèles « *bottom-up* ». Des questions liées à l'identification des signatures d'appareils, également connues sous le nom de NILM (« Non-Intrusive Load Monitoring »), sont également abordées. Aucun ne semble explorer directement ou discuter notre proposition spécifique de modèle de désagrégation.

Bagheera : le modèle de charge BT d'Enedis pour la planification des réseaux

En plus de la littérature scientifique, on peut regarder le modèle de charge BT d'Enedis pour la planification des réseaux. BAGHEERA [153, 163-166] est l'acronyme pour modèle BASse tension de Génération des cHarges et de calcul de l'Etat Electrique des Réseaux Arborescents. Il permet d'estimer la demande d'électricité dans une situation de référence puis de calculer l'état électrique du réseau. Dans ce manuscrit nous nous intéressons uniquement à la fonctionnalité « estimation de la demande ». Le modèle a été développé au début des années 90 pour répondre aux limitations du précédent modèle BATECA. Contrairement aux réseaux HTA pour lesquels on peut reconstituer à l'aide de quelques informations disponibles la puissance appelée en tête de départ HTA, pour les réseaux BT et avant Linky, on ne disposait d'aucune mesure permettant de le faire. C'est pourquoi BAGHEERA a comme base l'analyse statistique de campagnes de mesure de puissance (faites au début des années 90), au pas de temps 10 minutes, de 2000 courbes de charge représentatives des clients BT. On retiendra que par construction, le modèle cherche à capter la sensibilité de la demande d'électricité en période d'hiver. BAGHEERA compte trois types de données d'entrée, à savoir :

- Données Clients : Ce sont les données contractuelles (puissance souscrite, tarif, option tarifaire, activité) et les données de relevé (consommation et dates de relevé) ;
- Données Réseau : Ce sont les données qui décrivent le réseau (topologie, caractéristiques, raccordements, entre autres) ;
- Données Météo : Ce sont les températures moyennes journalières relevées par les différentes stations météorologiques.

A partir des données clients, différentes classes sont formées pour caractériser en comportements homogènes l'ensemble des clients BT. La classe de chaque client dépend du type de tarif, de l'option tarifaire et de sa catégorie d'activité. BAGHEERA fait le choix de représenter la sensibilité de chaque client à la température, pour cela il détermine pour chaque client :

- La sensibilité à la température (s) : C'est une variable qui représente la consommation quotidienne supplémentaire nécessaire quand la température baisse de 1°C , exprimé en kWh/dégré/jour ;
- La consommation journalière indépendante de la température (E_0) : en kWh.

Pour chaque client, le modèle estime une courbe de charge horaire en 48 points pour deux journées dimensionnantes (les jours plus froids de l'hiver) totalisant 24 puissances pour un jour ouvré et 24 puissances pour un jour weekend. La forme de la courbe dépend du tarif, de l'option tarifaire et de la catégorie d'activité du client. Le niveau de la courbe dépend de l'historique des consommations du client et de la température. Pour les décisions de développement de réseau, l'étude doit avoir lieu à la température minimale de base (TMB) et prendre en compte l'aléa sur la charge. La TMB peut être définie de la façon suivante : pour chaque station météorologique prise en référence, la Température Minimale de Base est définie comme la 30ème plus petite valeur moyenne journalière calculée sur une période de 30 ans. C'est donc la température froide qui n'est atteinte ou dépassée que 30 fois sur 10950 soit environ 1,6%. A chaque point horaire, le modèle estime une puissance appelée, variable aléatoire qu'il caractérise par deux éléments de sa distribution de probabilité :

- Valeur moyenne : $P_{\text{moy}(h)}$;
- Dispersion autour de sa valeur moyenne (mesurée par sa variance). La variance est un indicateur de la précision de l'estimation.

Après avoir connaissance de ces informations, le cas général d'estimation des puissances, où s et E_0 ont pu être estimés, est décrit par l'équation 4-1 :

$$P_{moy(h)} = a_h E_0 + b_h s (t_{NC} - t)_+$$

$$v_h = \sigma_h^2 E_n$$

Équation 4-1

Où :

- $P_{moy(h)}$ est la valeur moyenne de puissance pour l'heure h ;
- v_h est la variance représentative de la précision d'estimation pour l'heure h ;
- s et E_0 sont les caractéristiques du client ;
- E_n représente la consommation annuelle du client pour une année climatique normale ;
- t est la température de calcul ;
- t_{NC} est la température de non chauffage ;
- $(t_{NC} - t)_+$ indique que ce terme est nul quand $(t_{NC} - t)$ est négatif ;
- a_h, b_h et σ_h sont des paramètres issus de l'analyse des campagnes de mesure.

Lorsque la température t du calcul est égale à TMB, c'est là que le modèle Bagheera va calculer la courbe de puissance maximale consommée par le client. Cette valeur de charge extrême est utilisée en entrée des outils de calcul électrique du réseau pour la réalisation des études de planification. Pour atteindre l'objectif de satisfaire la demande d'électricité en respectant le compromis entre la qualité et le coût de fourniture, le modèle prend en compte un certain risque r % que la puissance au jour plus froid de l'année soit dépassée. Les valeurs de r proposées sont 2 % ou 5 % ou 10 %. Lors des études de développement de réseau, le risque est normalement pris égal à 10%.

Cette revue bibliographique nous confirme que nous cherchons à réaliser une recherche novatrice qui contribuera à élargir les horizons du domaine. En gardant cela à l'esprit, explorons l'idée de notre modèle en général.

4.1.2 Méthode de désagrégation en 2 étapes

Le principe général du modèle est de désagréger la courbe de charge du poste en des courbes de charge de chacun des clients en aval du poste. Pour ce faire, nous allons réaliser ce processus en 2 étapes :

- Réalisation d'une segmentation des clients : Les N clients d'un poste ont des caractéristiques similaires parmi eux, que ce soit parce qu'ils appartiennent aux mêmes catégories (résidentiel, professionnel, etc.) ou parce qu'ils ont des habitudes de consommation similaires. Dans cette perspective, la segmentation (ou ainsi le nombre de clusters) de ces clients est réalisée pour trouver les K groupes de clients similaires parmi les N clients du poste. Où $K < N$;
- Désagrégation de la Courbe du Poste : La courbe de charge du poste est ensuite désagrégée en K courbes, qui représentent les K groupes parmi de clients similaires du poste. Ces K courbes sont ensuite recalés en énergie pour attribuer à chaque client la courbe du groupe auquel il appartient.

4.1.3 Données

Le périmètre de données choisi pour le développement de ce modèle comprenait 48 postes HTA/BT, où les données de la courbe de charge Pinky ont été collectées pour la période de mars 2022 à mars 2023, ainsi que les données de puissance maximale et d'énergie des clients Linky (5318 au total) en aval de ces postes pour chaque jour de la période. Il est nécessaire de préciser que les données Linkys des clients utilisées dans cette étude sont associées à un panel de clients qui s'appelle « panel réseaux ». Ce sont des clients ayant donné leur consentement pour l'utilisation de ces données dans le cadre des études sur le réseau, auxquelles cette thèse est intégrée, conformément au règlement général sur la protection des données (RGPD).

Ces 48 postes ont été choisis car ils ne contiennent aucun client de la catégorie entreprise. De plus, tous les clients de type « entreprises » sont abonnés en mode courbe de charge depuis 2021, donc pour un poste qui a une entreprise, l'information de sa courbe de charge est déjà connue. L'objectif d'utiliser uniquement ces données (énergie et puissance maximale mesurées par Linky et la courbe de charge du poste mesurée par Pinky) est de construire le modèle avec moins de données possibles. Cela laisse le modèle plus facilement reproductible car ces données ne sont pas de caractéristiques exclusives du réseau français. Ainsi, notre méthode de désagrégation peut être explorée à partir de données provenant d'autres compteurs intelligents, rendant ainsi notre modèle adaptable à des utilisations en dehors du contexte français. La Figure 4-3 illustre la diversité des postes de notre jeu de données. Avec

des postes dans différentes régions du territoire, nous couvrons différents types de comportements de consommation qui se refléteront dans les données des compteurs Pinky :

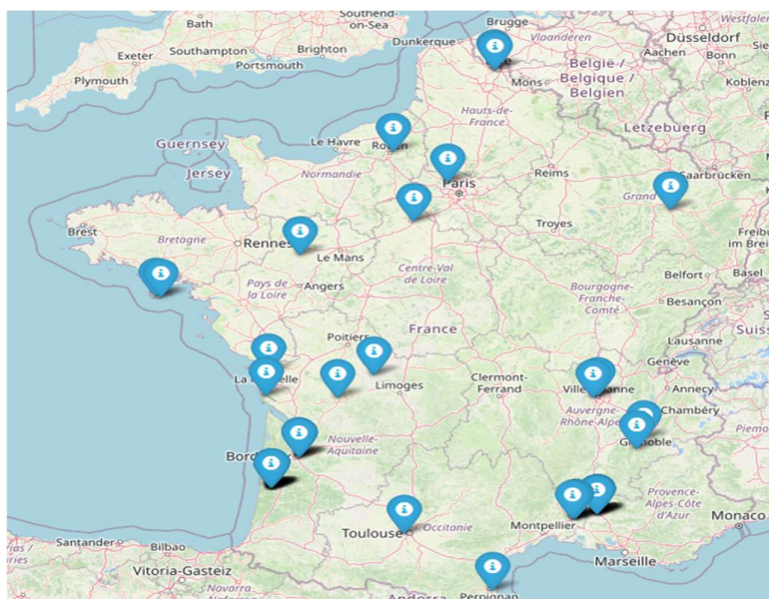


Figure 4-3 : Carte représentant les localisations des postes HTA/BT choisis pour l'étude

Avec toutes ces informations en tête, nous pouvons approfondir les détails de chacune des étapes du modèle. Afin de rendre la lecture des prochains sous-chapitres plus compréhensible et de leur donner une certaine continuité, le même poste HTA/BT nommé *Poste Exemple* sera utilisé comme exemple pour illustrer les résultats pour la saison d'hiver (01/12/2022 à 31/03/2023). Ensuite, une analyse sera réalisée pour montrer et discuter des résultats de l'application du modèle à d'autres postes.

4.2 Etape 1 : Segmentation des clients

L'objectif de cette étape est de trouver les K groupes de clients similaires d'un poste ayant N clients. L'idée derrière cela est de simplifier l'étape suivante de désagrégation, puisque nous allons désagréger la courbe de charge du poste en moins de courbes ($K < N$). De plus, travailler sur une maille individuelle du client est extrêmement compliqué car la variabilité de la consommation est énorme. Par conséquent, travailler sur une maille légèrement plus élevée (K groupes de clients similaires) est la meilleure stratégie. Plusieurs échelles temporelles pourraient être utilisées pour segmenter les clients d'un poste, par exemple : quotidienne, par jours de la semaine, par mois, par saison, par année, entre autres. Après plusieurs essais, une

segmentation par saison a été choisie car elle apporte plus d'avantages. Elle nous fournit un volume de données suffisant pour effectuer la segmentation des clients. Elle nous permet également de visualiser l'évolution du comportement de l'ensemble des clients au fil du temps, car les clients qui ont un comportement similaire au sein d'une saison donnée ne le présentent pas nécessairement au sein des autres saisons. Les données choisies pour la construction de cet ensemble de données sont les données de puissance maximale. La raison de n'utiliser que cette donnée est de rendre la construction aussi simple que possible en exploitant l'une des informations collectées par les compteurs Linky. De plus, plus les informations utilisées dans la construction du modèle sont simples, plus il sera facile de le reproduire dans d'autres recherches. Pour tous les clients et pour tous les jours de la saison, nous avons les informations de la Puissance Maximale mesurée par Linky :

- La valeur de puissance maximale en W ;
- L'instant de la journée (HH:MM) où la puissance maximale s'est produite converti en une valeur entière ($0 \leq T < 48$).

Pour le Poste Exemple, le nombre de variables du jeu de données est de 240, car la saison d'hiver en question a 120 jours (01/12/2022 à 31/03/2023). Avec notre ensemble de données construit, la prochaine étape consiste à traiter les cas particuliers. Dans ce cas spécifique, les cas particuliers sont les clients dont la valeur de puissance maximale est nulle pendant une proportion considérable des jours de la saison de l'année pour laquelle la segmentation sera réalisée. Une puissance maximale égale à zéro indique que le client ne consommait pas ce jour-là. La raison la plus courante sont les clients résidence secondaire, où le client partage son temps de consommation entre plusieurs résidences tout au long de l'année. Le critère choisi pour déterminer qu'un client est un « client résidence secondaire » est s'il a passé plus de la moitié des jours de la saison avec puissance maximale égale à zéro. Pour le *Poste Exemple*, parmi les 217 clients, 19 ont été identifiés comme des clients résidence secondaire. Ce groupe de clients est mis de côté et ne sera pas utilisé pour la suite de la segmentation. Une fois ces cas particuliers enlevés, la segmentation des clients est décrite dans les parties 4.2.1 et 4.2.2.

4.2.1 Analyse en Composantes Principales

Pour prétraiter les données, la technique de réduction de dimensionnalité appelée analyse en composantes principales (ACP) a été utilisée. Il s'agit d'une technique statistique utilisée pour

réduire la dimensionnalité des données tout en préservant, autant que possible, les informations essentielles contenues dans ces données. En particulier, ce type de méthode est intéressant pour la réduction de la taille des données traitées, pour limiter la redondance des données, et pour la visualisation plus facile des données, ce qui permet une meilleure compréhension [159]. L'ACP a été choisie parmi d'autres techniques de réduction de dimensions telles que t-SNE et UMAP, car elle présente des performances computationnelles élevées en plus d'être une méthode facile à comprendre. L'ACP fonctionne en transformant un ensemble de données en un nouvel ensemble de variables, appelées composantes principales, qui sont des combinaisons linéaires des variables d'origine. La matrice de covariance des données est calculée. A partir de cette matrice, les vecteurs propres et les valeurs propres sont calculés. Les vecteurs propres sont classés en fonction de leurs valeurs propres dans l'ordre décroissant. Les vecteurs propres ayant les plus grandes valeurs propres représentent les directions de la plus grande variation dans les données et sont appelés composantes principales. Comme les combinaisons linéaires sont orthogonales les unes par rapport aux autres, cela signifie que les composantes principales ne sont pas corrélées entre elles. Cela contribue à réduire la redondance des données. L'objectif est de réduire le nombre de ces nouvelles variables tout en conservant la variance maximale des données, ce qui permet de simplifier l'analyse tout en maintenant l'information significative. Pour le cas du Poste Exemple, nous avons réduit les variables dans le but de conserver au moins 80% de la variance de l'ensemble de données. Cela a été obtenu avec 3 composantes principales, maintenant ainsi 81,3 % de la variance.

4.2.2 Méthodes de Segmentation et choix de K nombre de clients similaires

Dans la littérature, il existe différentes méthodes pour effectuer une segmentation. Il existe des méthodes de partition qui utilisent la distance entre les individus comme base pour la segmentation. Il existe des méthodes hiérarchiques qui sont généralement représentées par une structure arborescente appelée dendrogramme. Il existe des méthodes fondées sur la densité qui modélisent les clusters comme des régions denses dans l'espace de données. Pour la réalisation de cette étape du modèle de désagrégation, nous avons choisi de travailler avec les méthodes de partition, Kmeans, en raison de leur simplicité dans la compréhension des résultats proposés. Une fois notre méthode de segmentation définie, il faut déterminer la méthode pour sélectionner la quantité K de groupes dans lesquels notre ensemble de données sera divisé. Habituellement, on sélectionne le nombre K de clusters en fonction de la quantité K qui produit les clusters de meilleure qualité selon un indicateur de qualité, pour une plage prédéterminée de valeurs de K. La littérature présente différents indicateurs qui mesurent la qualité des clusters produits. Parmi eux, on peut citer l'indice Davies-Bouldin, le critère

Calinski-Harabasz, la somme des carrés des erreurs (SSE), l'indice Silhouette, entre autres [167]. L'indicateur choisi était la somme des carrés des erreurs pour réaliser l'analyse du coude afin de définir le nombre K de clusters optimal qui conserve le meilleur compromis nombre/représentativité.

Kmeans clustering

Le clustering Kmeans est l'un des algorithmes d'apprentissage non supervisés les plus connus et utilisés. En général, les algorithmes non supervisés font des inférences à partir d'ensembles de données en utilisant uniquement des vecteurs d'entrée sans faire référence à des résultats connus ou étiquetés [168]. L'objectif du clustering Kmeans est de regrouper des points de données similaires et de découvrir des groupes sous-jacents. La Figure 4-4 montre un exemple du fonctionnement du Kmeans. Pour traiter les données d'apprentissage (Figure 4-4 – A), l'algorithme Kmeans commence par créer un premier groupe de centroides sélectionnés qui servent de points de départ à chaque cluster (Figure 4-4 – B), puis il fait l'assignation des points au centroïde le plus proche (Figure 4-4 – C) et après il déplace les centroides à la moyenne du cluster (Figure 4-4 – D). Des calculs itératifs sont effectués pour optimiser les positions des centroides (Figure 4-4 – E). Le processus s'arrête lorsque ces centroïdes se sont stabilisés ou lorsque que le nombre d'itérations maximal prédéfini a été atteint (Figure 4-4 – F). Le principal avantage de Kmeans réside dans sa simplicité de compréhension et de fonctionnement. De plus, il est efficace sur le plan computationnel car le calcul de la distance euclidienne est rapide à effectuer. Cependant, parmi les inconvénients, on peut mentionner sa sensibilité à l'initialisation des centroides, ce qui peut conduire à des résultats différents. Il suppose également que les clusters sont bien séparés et sphériques. Dans le cas où les données ne suivent pas cette forme, les résultats peuvent ne pas être précis est des méthodes comme DBSCAN sont plus adaptées.

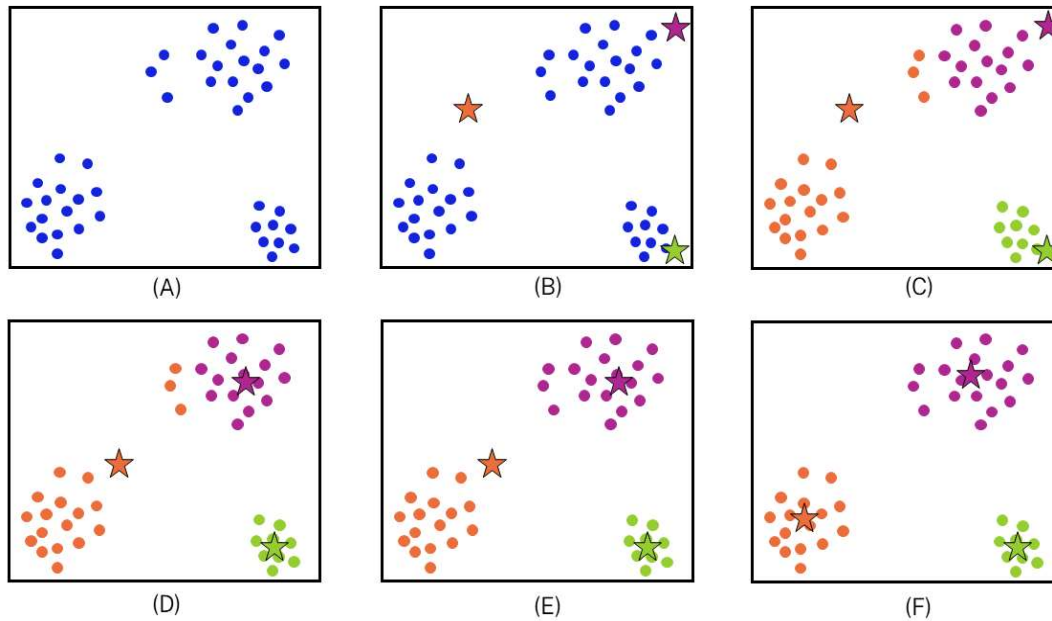


Figure 4-4 : Exemple du fonctionnement du clustering Kmeans

Méthode du Coude

La méthode du coude vise à identifier le nombre de clusters K qui offre un bon équilibre entre la réduction de la somme des carrés des erreurs (SSE) et la complexité du modèle. SSE mesure la somme des distances au carré entre chaque point de données et son centre de cluster assigné. En d'autres termes, la méthode du coude cherche le point où l'ajout de clusters supplémentaires n'améliore plus significativement la qualité du regroupement des données. C'est une méthode facile à comprendre et à appliquer, ce qui en fait un outil pratique pour choisir K . Cependant, l'inconvénient est la subjectivité de la méthode une fois que le coude peut être difficile à identifier. Pour cette raison, on a choisi d'utiliser la librairie python *Kneed* [169] pour faire la détection des coudes automatiquement et éliminer la partie de subjectivité. Cette librairie identifie le coude grâce au point de courbure maximale de la courbe. La Figure 4-5 montre l'application de la méthode du coude pour le jeu de données du *Poste Exemple* réduits avec l'ACP.

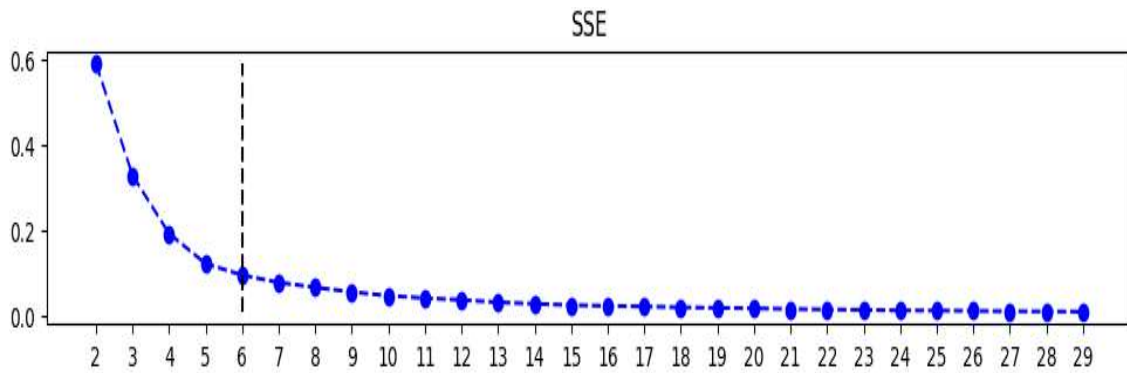


Figure 4-5 : L'application de la méthode du Coude pour le Poste Exemple

Pour le cas de notre *Poste Exemple*, pendant la saison d'hiver, sans les clients résidentiels secondaires, cette étape de segmentation (Kmeans et Méthode du Coude) est réalisée. Cette étape a identifié un nombre K de clusters égal à 6. La répartition du nombre de clients par cluster est présentée dans le Tableau 4-1.

Cluster	0	1	2	3	4	5
Nb Clients	72	74	5	3	24	20

Tableau 4-1 : Résultats de la segmentation pour le Poste Exemple avec K = 6

Nous pouvons constater que certains clusters sont plus peuplés que d'autres, ce qui ne signifie pas nécessairement quelque chose d'anormal. Comme expliqué précédemment, la construction de ce modèle vise à permettre de trouver des solutions simples qui répondent bien à l'objectif final de la modélisation, cela a été obtenu. Avec l'étape de segmentation bien structurée, nous pouvons passer à la prochaine étape, qui est l'étape de désagrégation. Nous utiliserons les résultats du *Poste Exemple* présentés ici pour la prochaine étape.

4.3 Etape 2 : Désagrégation

L'objectif de cette étape est de formuler une méthode pour désagréger la courbe de charge d'un poste HTA/BT en K courbes qui représentent les K groupes parmi les N clients identifiés lors de l'étape de segmentation. Pour commencer à réfléchir à la manière d'atteindre cette solution, commençons par formuler mathématiquement le problème.

$$\text{Courbe Charge}_{\text{poste HTA/BT}} = \sum_{k=1}^K \text{courbe cluster } (k) \quad \text{Équation 4-2}$$

Comme le montre l'équation 4-2, la désagrégation de la courbe de charge du poste peut être vue comme la somme des K courbes étant égale ou aussi proche que possible de la courbe du poste. En envisageant le problème sous cet angle, une fois que les équations décrivant les K courbes sont définies, ce problème peut être résolu par l'application d'une méthode d'ajustement de courbe (*curve fitting en anglais*). C'est une technique couramment utilisée en analyse de données pour estimer les paramètres d'une fonction mathématique qui décrit au mieux un ensemble de données [160]. Son principe de base consiste à trouver les paramètres optimaux d'une fonction mathématique, $f(x, \theta)$, qui minimise le carré de l'écart entre les données observées (x_i, y_i) où $1 \leq i \leq M$ et les valeurs prédites par la fonction, comme montre l'équation 4-3 :

$$\min \left(\sum_{i=1}^M (y_i - f(x_i, \theta_{\text{optimal}}))^2 \right) \quad \text{Équation 4-3}$$

L'objectif est de trouver θ_{optimal} qui minimisent cette somme des carrés, appelée fonction coût. En d'autres termes, l'ajustement de courbe consiste à résoudre un problème d'optimisation où le besoin est identifier les valeurs des paramètres θ qui rendent la fonction de coût aussi petite que possible. Pour résoudre ce problème d'ajustement de courbe, la bibliothèque Python *Imfit* [161] a été choisie. Comme expliqué par les développeurs, *Imfit* offre une interface conviviale pour définir des modèles, spécifier des paramètres, ajuster les données et récupérer les résultats. Elle permet aussi le choix sur une variété de méthodes d'optimisation. Dans le cas de ce problème de désagrégation, la méthode choisie est le L-BFGS-B (Limited-memory Broyden-Fletcher-Goldfarb-Shanno with Bound constraints) [162]. L-BFGS-B est un algorithme d'optimisation qui combine les avantages des méthodes quasi-Newton avec une gestion efficace de la mémoire et la capacité de traiter des contraintes de bornes sur les variables.

Avec la définition de la manière dont le problème de désagrégation sera abordé, il est maintenant nécessaire de définir quelle fonction mathématique décrit le mieux nos données. Les données sur lesquelles l'ajustement de courbe sera réalisé sont les courbes de charge quotidiennes des postes HTA/BT mesurées par les compteurs Pinky.

4.3.1 Modèle purement mathématique

L'option retenue qui semble répondre à la question de quelles équations mathématiques décrivent le mieux la somme de K courbes de l'équation 4-2 est une somme de N gaussiennes, comme le montre l'équation 4-4 :

$$f(x, a, b, c) = \sum_{k=1}^N (a_k * \exp \frac{-(x-b_k)^2}{2*c_k^2}) \quad \text{Équation 4-4}$$

Où : θ de la fonction $f(x, \theta)$ correspond aux paramètres a_k, b_k, c_k . a_k détermine la hauteur de la gaussienne, b_k détermine la position du centre de la gaussienne et c_k détermine la largeur de la gaussienne.

La raison qui justifie l'hypothèse des gaussiennes comme un bon point de départ est la suivante. Si nous examinons la forme des profils de charge des clients [151, 152], les larges pics de consommation à différents moments de la journée rappellent un comportement gaussien. Ces profils sont la représentation du comportement d'un ensemble de clients faisant partie de la même classe. La Figure 4-6 montre la forme des profils résidentiels tarif de base et des profils professionnels tarif de base :

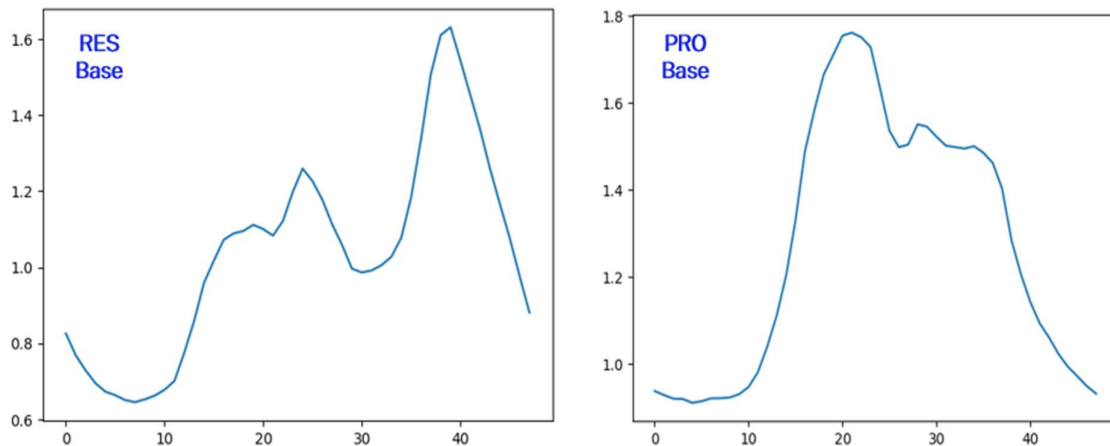


Figure 4-6 : Format de la courbe des profils résidentiels tarif de base et des profils professionnels tarif de base

De plus, une gaussienne est un candidat naturel car c'est une fonction classique en statistique liée au théorème central limite. Ce théorème joue un rôle fondamental dans la théorie statistique en montrant que la moyenne de nombreuses variables aléatoires indépendantes et identiquement distribuées suit approximativement une distribution gaussienne, quelle que soit la distribution initiale de ces variables. Ce qui démontre son omniprésence dans la modélisation de phénomènes aléatoires, comme la consommation électrique des clients BT. Si nous observons le comportement de la courbe de charge d'un poste HTA/BT, nous pouvons remarquer que tout au long de la journée, il est possible d'observer plusieurs pointes de consommation qui rappellent une forme gaussienne, comme le montre la Figure 4-7 :

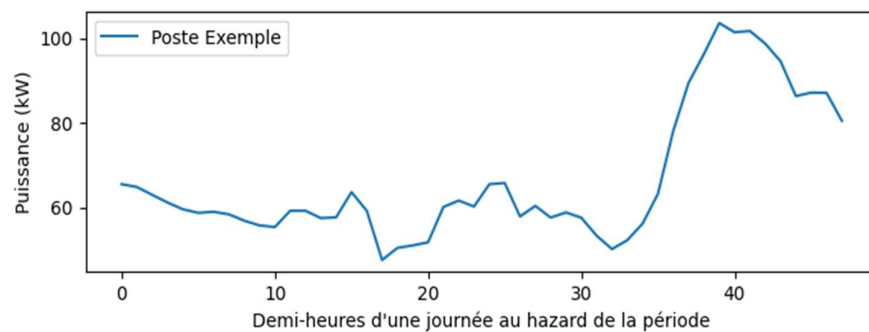


Figure 4-7 : Mesure Pinky du Poste Exemple pour une journée au hasard

On peut imaginer que la pointe de la 38ème demi-heure pourrait être représenté par une gaussienne, tout comme les pointes de la 15ème demi-heure et de la 25ème demi-heure. A ce stade, des questions se posent comme : combien de gaussiennes utiliser par cluster ? Le nombre de gaussiennes doit-il être le même pour tous les clusters ? Peut-on déterminer à l'avance la position b des gaussiennes ? La suite de ce sous-chapitre apportent des réponses à ces questions. Le Poste Exemple continue d'être utilisé comme base pour tester les résultats générés par les différentes compositions de cette fonction $f(x,a,b,c)$ somme de gaussiennes.

4.3.1.1 Modèle 1 « gaussienne simple » : une gaussienne par cluster

La version la plus simplifiée de notre fonction $f(x,a,b,c)$ est avec une gaussienne par cluster, où les paramètres (a, b, c) n'ont pas de contraintes. Cela est montré par l'équation 4-5 :

$$f(x, a, b, c) = \sum_{k=1}^K (a_k * \exp(\frac{-(x-b_k)^2}{2*c_k^2})) \quad \text{Équation 4-5}$$

L'ajustement de courbe va trouver les paramètres optimaux pour minimiser l'erreur entre notre fonction décrite par l'équation 4-5 et, dans notre cas, une courbe Pinky. Dans ce contexte, notre fonction décrit que "la courbe Pinky" est égale à la somme des comportements par cluster, et ces comportements sont traduits à partir d'une gaussienne.

Nous allons appliquer cette hypothèse pour un jour aléatoire de la saison d'hiver du Poste Exemple. Le nombre K de clusters pour le Poste Exemple pendant la saison d'hiver est de 6 clusters, comme expliqué précédemment. Les clients en résidence secondaire ne seront pas pris en compte pour une question de simplification. Les résultats de l'ajustement de courbe sont présentés dans la Figure 4-8 et la Figure 4-9. La Figure 4-8 nous montre la gaussienne de chaque cluster, calculée à l'aide des paramètres optimaux trouvés par l'ajustement de courbe. Nous pouvons observer que seuls 2 clusters ont des valeurs non nulles. La somme des gaussiennes de chaque cluster nous présente la courbe ajustée résultante de l'équation 4-5, telle qu'illustrée dans la Figure 4-9.

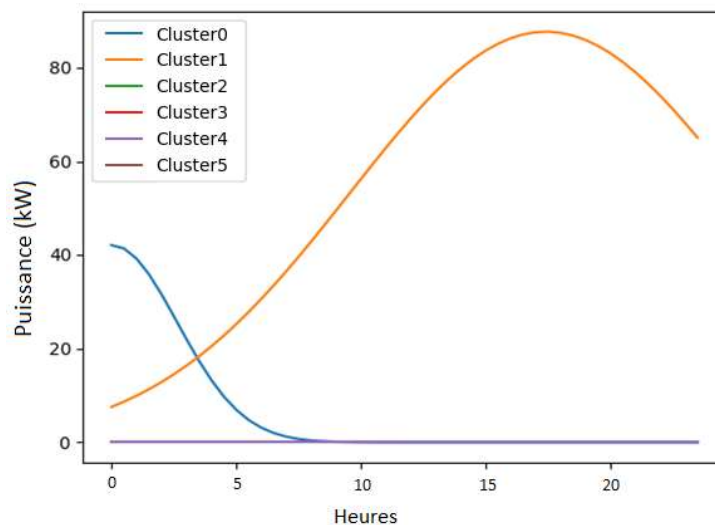


Figure 4-8 : Résultat de l'ajustement de courbe avec l'équation 4-5 – courbes par cluster

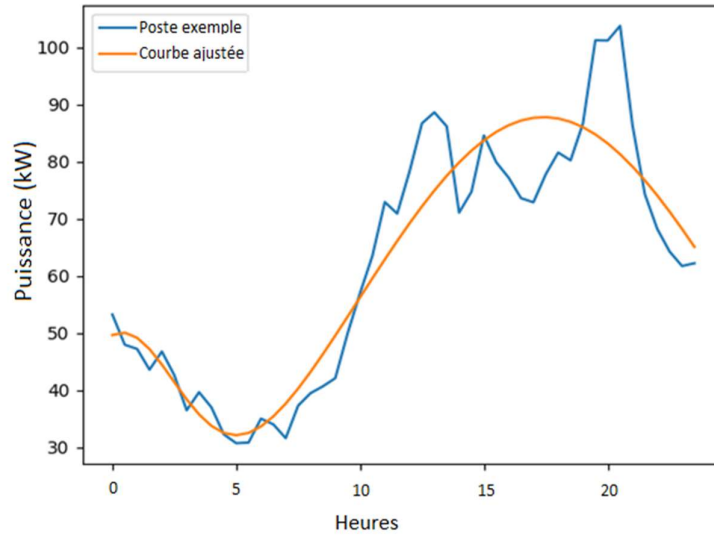


Figure 4-9 : Comparaison entre le résultat de l'ajustement de courbe avec l'équation 4-5 et la courbe Pinky du poste Exemple dans la journée d'hiver

L'erreur SMAPE entre la courbe du Poste Exemple et la courbe ajustée pour le cas de l'équation 4-5 est de 8,63 %. Ce que nous pouvons interpréter à partir de la figure, c'est que la courbe ajustée parvient à retrouver la tendance de la courbe du Poste Exemple. La région de la vallée et la région de la pointe sont aux moments appropriés de la journée, mais les pointes « secondaires », comme de la 24^{ème} demi-heure et de la 42^{ème} demi-heure ne sont pas bien représentés.

4.3.1.2 Modèle 2 « somme de gaussiennes » : deux gaussiennes par cluster

Pour capturer de manière plus efficace les différents pics de consommation de la journée, nous augmentons le nombre de gaussiennes par cluster pour avoir deux gaussiennes par cluster, comme le montre l'équation 4-6 :

$$f(x, a, b, c) = \sum_{k=1}^K \left(a1_k * \exp \frac{-(x-b1_k)^2}{2*c1_k^2} + a2_k * \exp \frac{-(x-b2_k)^2}{2*c2_k^2} \right) \quad \text{Équation 4-6}$$

$$0 < a < \max(\text{Courbe Pinky})$$

Le paramètre a a ses plages définies dans le but de ne pas permettre l'apparition de gaussiennes négatives, étant donné que nous étudions l'application du modèle pour les clients

d'un réseau qui n'a pas de producteurs d'énergie raccordés. Par conséquent, avoir des gaussiennes négatives n'aurait pas de sens pratique dans ce cas. L'ajustement de courbe a été appliqué au *Poste Exemple* le même jour que précédemment. La Figure 4-10 nous montre la « somme de gaussiennes » résultant de chaque cluster, calculée à l'aide des paramètres optimaux trouvés par l'ajustement de courbe. La comparaison entre le courbe ajustée et la courbe du Poste exemple est présenté sur la Figure 4-11 :

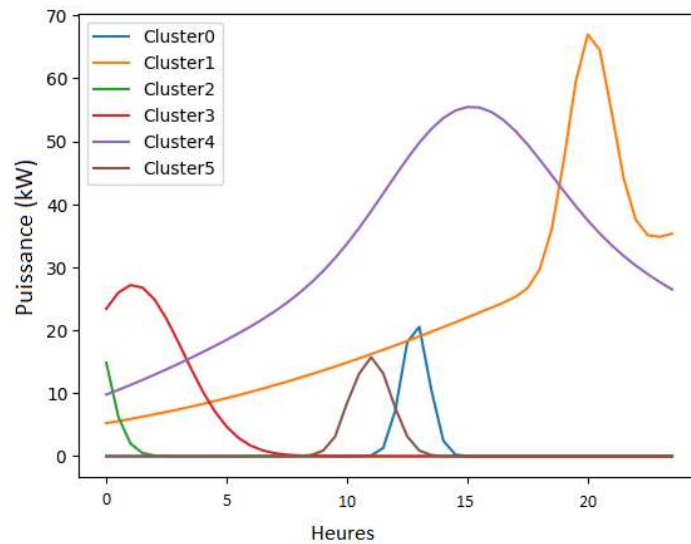


Figure 4-10 : Résultat de l'ajustement de courbe avec l'équation 4-6 – courbes par cluster

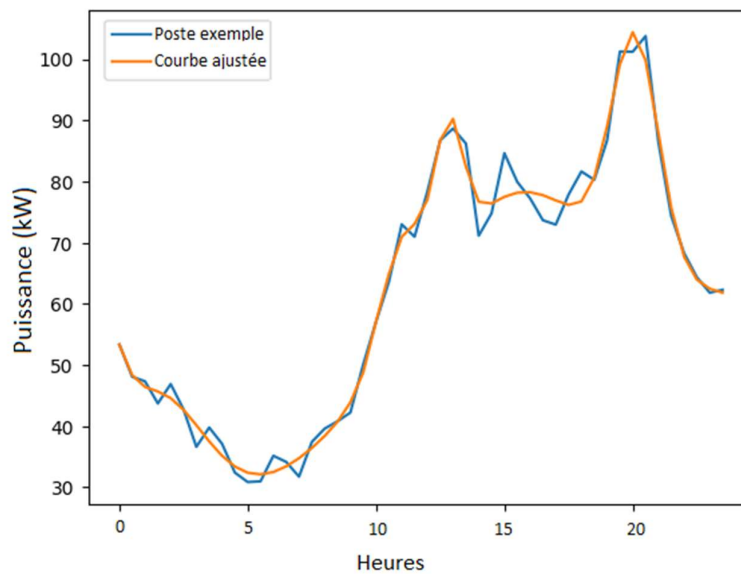


Figure 4-11 : Comparaison entre le résultat de l'ajustement de courbe avec l'équation 4-6 et la courbe Pinky du poste Exemple dans la journée d'hiver

L'erreur SMAPE entre la courbe du Poste Exemple et la courbe ajustée pour le cas de l'équation 4-6 est de 3,15 %. Cela représente une réduction du pourcentage d'erreur par rapport à l'hypothèse précédente de 63%. Cela indique que l'option de la fonction $f(x, a, b, c)$ décrite par l'équation 6 étant une somme de gaussiennes est une solution viable pour le problème. La suite de l'étude de la définition de la fonction cherchera à rendre les courbes ajustées plus proches de ce qui se passe dans le réseau, car nous voulons avoir un modèle fonctionnel.

4.3.2 Ajout de propriétés électriques au modèle purement mathématique

Pour l'instant, aucune information sur les données de puissance maximale et d'énergie associées aux individus faisant partie des clusters n'a été utilisée dans la formulation de notre ajustement de courbe. Pour rapprocher le modèle d'un comportement réel des clients du réseau, ces informations doivent être exploitées, et cela sera fait au cours de cette section.

4.3.2.1 *Modèle 3 « Somme de Gaussiennes-Pmax » : Somme de Gaussiennes corrélées avec la donnée de Puissance Maximale*

Initialement, la puissance maximale a attiré notre attention car cette donnée, qui vient du « panel réseaux », est accompagnée de l'heure et de la minute à laquelle elle s'est produite. Cette information temporelle semblait d'une certaine manière utile pour définir le paramètre b , centre des gaussiennes. Nous avons donc choisi d'approfondir cette donnée. En analysant, pour un même client, tous les horaires auxquels sa puissance maximale a été enregistrée au cours de la saison, il est possible de constater que ceux-ci ne sont pas toujours identiques. Cela est illustré par la Figure 4-12, où chaque point bleu représente une occurrence de puissance maximale pour 5 exemples de clients et pour la saison d'hiver.

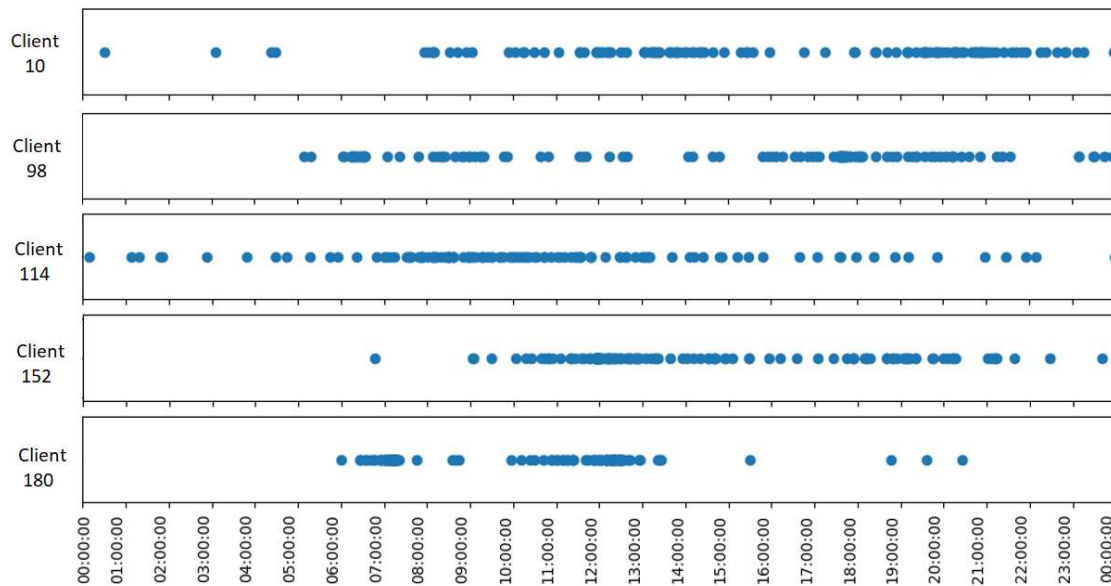


Figure 4-12 : Occurrences de puissance maximale de 5 clients au cours de la saison

Il est possible d'observer que chaque client présente des tendances de consommation différentes et qu'elles varient considérablement tout au long de la saison. Cela nous amène à penser que l'observation d'un groupe de clients agrégés pourrait nous montrer des motifs plus faciles à identifier. Il a donc été choisi d'examiner le groupe de clients appartenant au même cluster. La Figure 4-13 montre l'occurrence de la puissance maximale pendant toute la saison d'hiver pour tous les clients des clusters 0 et 5, à partir d'un histogramme où chaque barre correspond à un intervalle de 10 minutes. L'histogramme de chaque cluster du Poste Exemple se retrouve dans l'annexe 6.2.1.

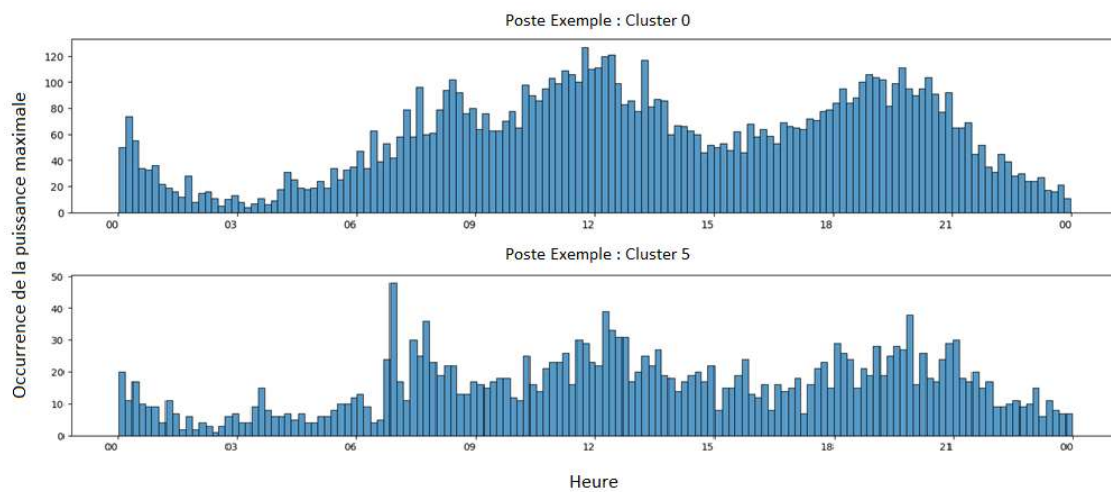


Figure 4-13 : Occurrences de la puissance maximale par les clients des cluster 0 et 5

En observant les occurrences par groupe, les tendances deviennent plus évidentes. Pour le cluster 0, on peut noter quatre concentrations d'occurrences de puissance maximale : autour de midi, 20h, 8h et minuit. Pour le cluster 5, on remarque une grande concentration à 7 heures du matin et des concentrations moins marquées vers midi et vers 20 heures. Maintenant, il est nécessaire de traduire ces informations en réglant les paramètres de notre fonction $f(x,a,b,c)$.

Les concentrations de puissance maximales à une période donnée indiquent si c'est une période importante pour le cluster, méritant une bonne résolution pour ajuster plus finement les pointes. Donc, pour chaque cluster, nous pouvons dire que les concentrations les plus marquées des histogrammes indiqueront le nombre maximal de gaussiennes, ainsi que la valeur initiale du paramètre "b" de chacune de ces gaussiennes. Pour identifier ces concentrations, la fonction *find_peaks* de la bibliothèque Python *SciPy* a été utilisée [170]. Elle permet d'identifier et de localiser les maxima locaux dans un ensemble de données unidimensionnelles, par simple comparaison des valeurs voisines. Un sous-ensemble de ces maxima locaux peut être sélectionné en spécifiant des conditions pour les propriétés d'un maxima. Les propriétés utilisées sont la « distance » et la « proéminence ». La « distance » correspond à la distance minimale en nombre de points entre les maxima. La « proéminence » est une mesure qui quantifie à quel point un maxima est « élevé » et « isolé » par rapport aux maxima voisins. Elle correspond à la hauteur du maxima au-dessus du point le plus bas entre les maxima voisins (creux). Si un maxima n'a pas une valeur de « proéminence » supérieure à la valeur fixée, il ne sera pas considéré comme un maxima. Ensuite, les données unidimensionnelles utilisées sont le carré des occurrences de la puissance maximale afin de différencier davantage les pointes de l'histogramme. Le paramètre de distance a été réglé à 6, car chaque point représente un intervalle de 10 minutes, et 6 points représentent 1 heure. Le paramètre de proéminence a été réglé à p , où p est égal au nombre de clients du cluster au carré (il est au carré car nos données unidimensionnelles sont au carré) multiplié par le nombre de jours de la saison divisé par 288 (nombre de intervalles de 10 minutes dans la journée multipliée par deux). Comme le nombre de clients n'est pas le même pour tous les clusters, l'échelle des valeurs d'occurrence ne sera pas la même. Par conséquent, il n'est pas intéressant d'avoir une seule valeur de proéminence pour tous les clusters. Pour les clusters 0 et 5, les résultats de ce calcul sont présentés dans la Figure 4-14, où chaque point orange représente un maxima. Les points de maxima de chaque cluster du Poste Exemple se retrouve dans l'annexe 6.2.1.

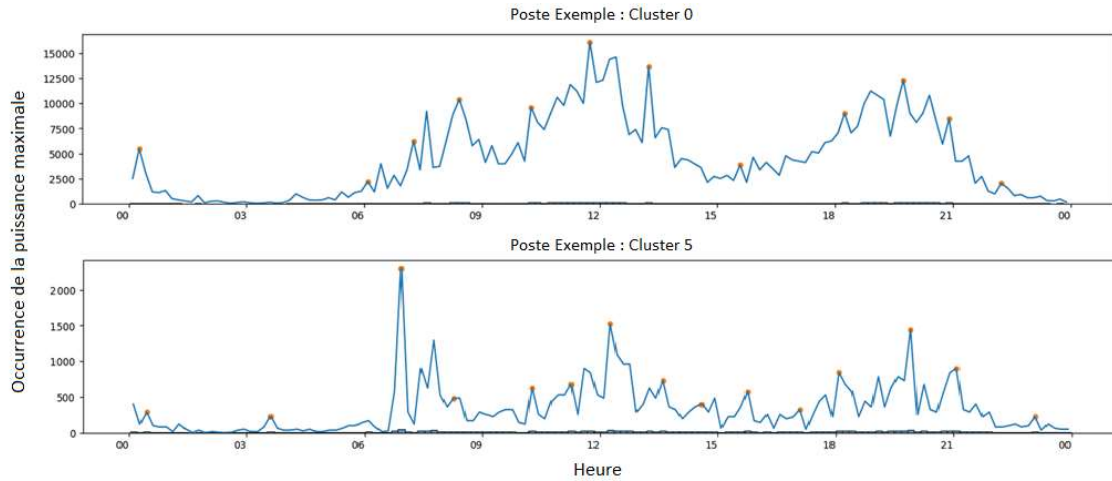


Figure 4-14 : Points de maxima de l'histogramme des occurrences de puissance maximale

Pour les clusters 0 et 5, nous avons trouvé respectivement 12 points et 15 points de maximum. Ces points représentent les heures de la journée où il y a le plus de chances d'occurrence de la puissance maximale pour chaque cluster tout au long de la saison. Les points de maxima seront appelés par la lettre t et les heures d'occurrences pendant lesquels ils apparaissent seront appelés h_t . Cela nous oriente vers une autre façon d'écrire notre fonction somme de gaussiennes $f(x,a,b,c)$ et du réglage des bords des paramètres. Pour chaque cluster k , nous aurons T nombres de gaussiennes trouvées à partir de l'histogramme des occurrences de puissance maximale, comme indiqué par l'équation 4-7 :

$$f(x, a, b, c) = \sum_{k=1}^K \sum_{t=1}^T \left(a_{t,k} * \exp \frac{-(x-b_{t,k})^2}{2*c_{t,k}^2} \right)$$

Équation 4-7

$$\begin{aligned} 0 &\leq a_{t,k} < \max(\text{Courbe Pinky}) \\ h_t - 1 &\leq b_{t,k} \leq h_t + 1 \\ 2 &< c_{t,k} < 12 \end{aligned}$$

Avoir le paramètre $a_{t,k}$ pouvant avoir une valeur de zéro permet à l'optimisation de se débarrasser de la gaussienne (t, k) si elle n'est pas intéressante pour l'ajustement de la courbe, en fixant sa hauteur à zéro. Cela représente le comportement variable traduisant le fait que les clients de ce groupe n'auront pas leur puissance maximale à cet horaire tous les jours. Le paramètre $b_{t,k}$ pouvant varier d'environ une demi-heure autour de l'heure de h_t permet de déplacer le centre de la gaussienne, ce qui est cohérent avec les occurrences de puissance maximale autour de chaque maxima. Le paramètre $c_{t,k}$ représente la largeur de la gaussienne,

avec le bord inférieur à deux demi-heures, cela nous permet d'éviter la présence de pics qui apparaissent et disparaissent très rapidement, car nous partons de l'hypothèse que la consommation du groupe de clients présente une certaine inertie dans ses variations. Le paramètre $c_{t,k}$ a le bord supérieur limité à 12 demi-heures pour éviter ainsi de perdre l'information d'une pointe possible. Ainsi, notre fonction d'ajustement de courbe est définie à partir des données des clients appartenant à chaque cluster. La nouvelle hypothèse de la fonction décrite par l'équation 4-7 d'ajustement de courbes est testée pour le Poste Exemple pour tous les jours de la saison d'hiver. Les résultats sont résumés dans le Tableau 4-2 :

	Modèle 1 « gaussienne simple »	Modèle 2 « somme de gaussiennes »	Modèle 3 « somme de gaussiennes-Pmax »
SMAPE minimal (%)	4,09	1,16	1,60
SMAPE moyen (%)	6,36	2,81	2,64
SMAPE maximal (%)	10,81	6,11	3,99

Tableau 4-2 : Erreurs SMAPE d'ajustement de courbes pour la saison d'hiver pour les trois modèles proposés

Nous pouvons constater que nous avons maintenu l'erreur à 2 % en moyenne. De plus, le modèle 3 a réussi à réduire l'erreur maximale d'un facteur de 1,5. Cela nous indique que nous n'avons pas dégradé nos résultats par rapport à l'hypothèse précédente avec une construction de la fonction $f(x,a,b,c)$ qui tient compte des comportements des clients des différents clusters. L'intérêt de prendre en compte le comportement des clients des différents clusters est de garantir une cohérence avec la réalité, rapprochant ainsi de plus en plus le modèle de ce qui se passe dans le réseau. La Figure 4-15 montre la comparaison entre l'ajustement de courbe avec l'équation 4-7 et la courbe Pinky du Poste Exemple dans la journée avec le plus faible erreur SMAPE de 1,6 %. Il est possible d'observer que la courbe ajustée a réussi à capturer les différents pics de consommation représentés par la courbe Pinky. Cela met en évidence la cohérence entre la distribution des occurrences de puissance maximale par cluster et les mesures de consommation de l'ensemble des clients observées par Pinky dans la maille du Poste HTA/BT. La Figure 4-16 montre les courbes résultantes par cluster de cette journée en question. On les appellera des profils pour simplifier la lecture et pour faire référence aux profils actuels utilisés par les responsables d'équilibre. Les courbes des clusters 1 et 4 de la Figure 4-16 nous montrent une grande similitude avec les courbes qui représentent les formes des profils présentées dans la Figure 4-6 (profils RES base et PRO base construits historiquement par Enedis). Comme à aucun moment des données représentant ces profils

classiques n'ont été utilisées dans la construction du modèle, avoir des résultats qui leur ressemblent est un autre indicateur de la qualité de notre modélisation, car elle nous montre une cohérence avec les comportements de l'ensemble de clients mesurés par les profils classiques. Pour continuer à rapprocher notre modèle des conditions réelles du réseau, nous devons maintenant nous pencher sur un autre facteur très important : l'énergie de nos profils pour chaque cluster. Afin de pouvoir réattribuer ces profils aux clients qui en font partie, nous devons nous assurer que l'énergie pour chaque profil soit cohérente avec l'énergie mesurée par le compteur Linky. C'est le prochain sujet qui sera discuté dans la section 4.3.2.2.

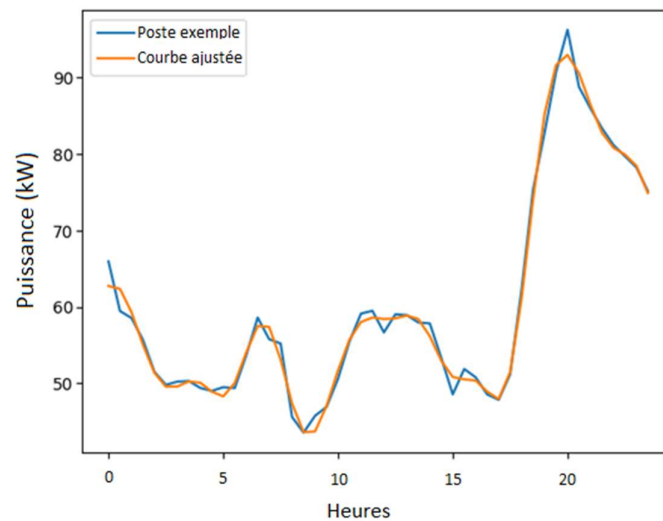


Figure 4-15 : Comparaison entre le résultat de l'ajustement de courbe avec l'équation 4-7 et la courbe Pinky du poste Exemple dans la journée avec le plus faible erreur SMAPE de 1,6 %

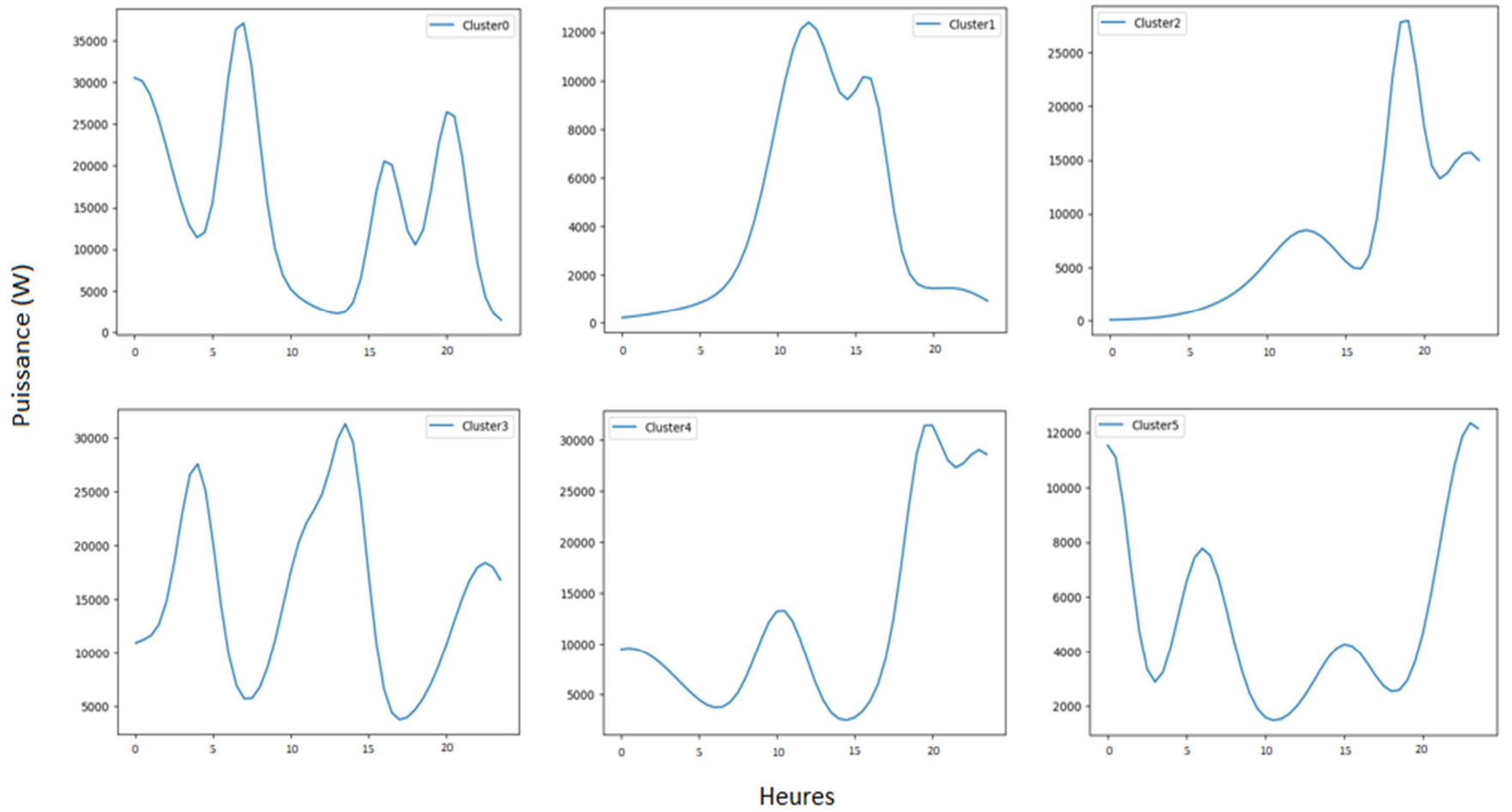


Figure 4-16 : Courbes résultants de la désagrégation par cluster au jour où le SMAPE est minimal

4.3.2.2 Modèle 4 « Somme de Gaussiennes-Pmax-énergies » : : Ajout d'une contrainte liée aux énergies des clients

Le modèle 4 est une version mise à jour du modèle 3 avec l'ajout de 2 nouveaux facteurs pour obtenir une cohérence énergétique, à savoir : une constante dans notre fonction $f(x,a,b,c)$ et une pénalité dans notre optimisation. Plus de détails sur ces deux facteurs seront expliqués tout au long de cette section. Pour envisager cette cohérence énergétique dans notre ajustement de courbes, nous devons prendre en compte les pertes dans le réseau, ainsi que l'énergie des clients résidentiels secondaires. Les clients résidentiels secondaires sont regroupés dans leur propre cluster et seront pris en compte lors de l'ajustement des courbes. Donc, dans ce cas, le Poste Exemple aura un cluster de plus que dans les modèles précédents pour accueillir ces clients. Comme l'objectif final de ce modèle est de désagréger la courbe du poste HTA/BT en courbes des clients, les courbes des clients doivent être cohérentes avec l'énergie qu'ils consomment. Nous pouvons exprimer cela mathématiquement comme suit :

$$E_{profils_k} = E_{cluster_k} \quad \text{Équation 4-8}$$

Où $E_{profils}$ est l'énergie de la courbe profil provenant de la désagrégation, $E_{cluster}$ est la somme de l'énergie mesurée par le compteur Linky des clients qui font partie du cluster. Une façon de prendre cela en compte dans notre ajustement de courbe est d'ajouter une pénalité, nommée *Pénalité*. Elle peut être formulée par l'erreur quadratique exprimées dans l'équation 4-9 :

$$Pénalité = \sum_{k=1}^k (E_{profils_k} - E_{cluster_k})^2 \quad \text{Équation 4-9}$$

La pénalité est donc la somme des erreurs carrés entre l'énergie du profil et l'énergie des clients pour chaque cluster. Le choix d'une pénalité quadratique est dû au fait qu'elle accorde un poids plus important aux erreurs importantes, ce qui simplifie la recherche des minima lors de l'optimisation des paramètres pour l'ajustement des courbes. Comme la pénalité que nous allons introduire est quadratique, nous pouvons réutiliser notre méthode d'optimisation L-BFGS-B, car les méthodes quasi-Newton sont efficaces pour les fonctions faiblement non linéaires, continues et différentiables, comme est le cas d'une fonction quadratique. Ainsi, nous pouvons mettre à jour l'équation 4-3 du principe de base de l'ajustement de courbes pour prendre en compte notre pénalité :

$$\min \left(\sum_{i=1}^M (y_i - f(x_i, \theta_{optimal}))^2 + \lambda * Pénalité \right) \quad \text{Équation 4-10}$$

L'équation 4-10 nous indique que les paramètres optimaux de notre fonction $f(x, \theta)$ doivent minimiser la somme de deux termes : le carré de la différence entre la courbe à ajuster et la courbe produite par la fonction $f(x, \theta)$, plus la pénalité multipliée par une constante lambda (λ). En d'autres termes, en plus de chercher à minimiser les erreurs de l'ajustement de courbes, il est nécessaire de minimiser également les erreurs des courbes résultantes en termes d'énergie. La constante lambda a pour objectif d'équilibrer l'importance du terme de pénalité lors de l'optimisation. Plus la valeur de lambda est élevée, plus le poids de la pénalité lors de l'optimisation sera important, et plus sa valeur est faible, moins le poids de la pénalité sera important lors de l'optimisation. Notre objectif était d'équilibrer l'importance des deux termes, et nous avons réussi à trouver ce compromis avec une valeur de pondération de lambda de 0,0001 (les résultats pour différentes valeurs de lambda sont dans l'annexe 6.2.2). Afin de trouver cet équilibre dans cet ajustement de courbes avec une pénalité, nous avons ajouté une constante dans notre fonction $f(x, a, b, c)$, comme le montre l'équation 4-11 :

$$f(x, a, b, c) = \sum_{k=1}^K \sum_{t=1}^T \left(a_{t,k} * \exp \frac{-(x-b_{t,k})^2}{2*c_{t,k}^2} \right) + l_k * Ecluster_k$$

$$\begin{aligned} 0 &\leq a_{t,k} < \max(\text{Courbe Pinky}) \\ h_t - 1 &\leq b_{t,k} \leq h_t + 1 \\ 2 &< c_{t,k} < 12 \\ 0 &< l_k < 1 \end{aligned} \quad \text{Équation 4-11}$$

Le terme constant ajouté à la fonction permet d'ajuster le niveau de la courbe pendant l'optimisation. La construction de la fonction et des valeurs de bordure pour la partie gaussienne se fait de la même manière qu'expliqué dans l'équation 4-7 à travers la distribution des occurrences de puissance maximale pour les clients de chaque cluster. Ce qui a été ajouté, c'est la constante, $l_k * Ecluster_k$, qui règle le niveau du profil de chaque cluster en fonction de leurs énergies respectives. L'idée d'ajouter cette constante dans notre fonction est de représenter la consommation moyenne des clients. Étant donné que l_k est également un paramètre à optimiser, cela permettra à l'ajustement des courbes de mieux conserver l'énergie, réduisant ainsi la différence entre $Ecluster_k$ et $Eprofils_k$. La nouvelle fonction

définie par l'équation 4-11 avec la pénalité dans l'ajustement des courbes a été testée pour le Poste Exemple pour tous les jours de la saison d'hiver. Les résultats de l'erreur SMAPE entre les courbes ajustées et les courbes Pinky pour la saison d'hiver sont résumés dans le Tableau 4-3. Le tableau 4-4 nous présente les résultats sur les énergies qui sont calculés comme le $SMAPE(E_{cluster_k}, E_{profils_k})$ de chaque jour de la saison d'hiver.

	Modèle 1 « gaussienne simple »	Modèle 2 « somme de gaussiennes »	Modèle 3 « somme de gaussiennes- Pmax »	Modèle 4 « somme de gaussiennes- Pmax-énergies »
SMAPE minimal (%)	4.09	1.16	1.60	2.06
SMAPE moyen (%)	6.36	2.81	2.64	4.86
SMAPE maximal (%)	10.81	6.11	3.99	11.30

Tableau 4-3 : Nouvelles erreurs SMAPE de l'ajustement de courbes pour la saison d'hiver

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster Rés. Secondaire
SMAPE minimal (%)	0,01	0,01	0,28	0,10	0,02	0,01	0,33
SMAPE moyen (%)	0,29	0,21	26,27	44,38	0,50	0,47	95,89
SMAPE maximal (%)	1,03	0,62	127,79	200,00	1,92	1,75	200,00

Tableau 4-4 : Les erreurs SMAPE des énergies du modèle 4

Les résultats obtenus par le modèle 4 présentent une dégradation par rapport aux résultats du modèle 3 issus de l'hypothèse qui ne tient pas compte de la pénalité liée à la différence d'énergie comme montre le Tableau 4-3. Cela peut être considéré comme naturel, car plus nous approchons le problème des conditions réelles du réseau, plus il devient complexe. Par conséquent, cela conduit à une optimisation plus contrainte, ce qui explique la légère dégradation des résultats. Cependant, avec cette formulation du modèle de désagrégation

définie par l'équation 4-11 avec la pénalité dans l'ajustement des courbes, nous avons pu trouver un compromis entre la qualité des résultats obtenus et l'approximation des conditions réelles du réseau. Nous pouvons observer dans le Tableau 4-4 que pour les clusters avec au moins 20 clients (Cluster0, Cluster1, Cluster4, Cluster5), nous avons une erreur d'énergie moyenne inférieure à 1%. En revanche, pour les clusters avec une consommation énergétique plus faible (Cluster1, Cluster2, Cluster Rés. Secondaire) soit en raison d'un faible nombre de clients, soit parce qu'ils sont constitués de résidences secondaires, c'est là que l'on trouve les erreurs les plus importantes. Faire ce compromis est essentiel pour garantir que le modèle ait une cohérence énergétique avec la consommation énergétique des clients mesurée par le compteur Linky. Même si cette cohérence présente une erreur élevée pour certains cas spécifiques, son ajout assure l'intérêt d'utiliser le modèle 4 pour répondre aux situations réelles du réseau à la maille BT contrairement au modèle 3, qui peut avoir un intérêt à la maille HTA en raison de ses erreurs plus faibles par rapport aux mesures Pinky. Pour la suite du chapitre, nous allons analyser les résultats produits par cette version du modèle pour les autres Postes HTA/BT de notre échantillon de données.

4.4 Analyse des résultats de l'application du modèle sur la totalité des postes à disposition

Jusqu'à présent, le modèle a été présenté sur un seul poste HTA/BT, le *Poste Exemple*. Une analyse de l'application du modèle 4 (somme de gaussiennes-Pmax-énergies) aux 48 postes HTA/BT de notre échantillon de données a été réalisée. De la même manière que pour le Poste Exemple, le modèle est appliqué à tous les jours de la saison d'hiver. Cette analyse vise à évaluer les erreurs du modèle pour d'autres postes, ce qui permet d'avoir une vue plus large de la qualité des résultats. Elle permet également d'identifier si le modèle fonctionne mieux pour certains postes que pour d'autres, et pourquoi. A partir de cette évaluation, nous pouvons identifier des pistes d'amélioration potentielles à explorer pour le modèle. De la même manière que l'erreur a été calculée pour le Poste Exemple, elle sera calculée pour les autres postes. La moyenne des erreurs SMAPE est résumée dans le Tableau 4-5:

	Moyenne des SMAPE (%)
Minimal	6.47
Moyen	12.06
Maximal	21.44

Tableau 4-5 : Moyenne des SMAPE pour les 48 postes HTA/BT

Comparé au Poste Exemple, on peut observer que l'erreur a considérablement augmenté. Si l'on regarde la valeur moyenne de la moyenne des erreurs, on peut voir que l'erreur a augmenté de 148 %. Pour mieux restreindre l'analyse, la Figure 4-17 regroupe les postes en tranches d'erreur en fonction de la moyenne des erreurs.

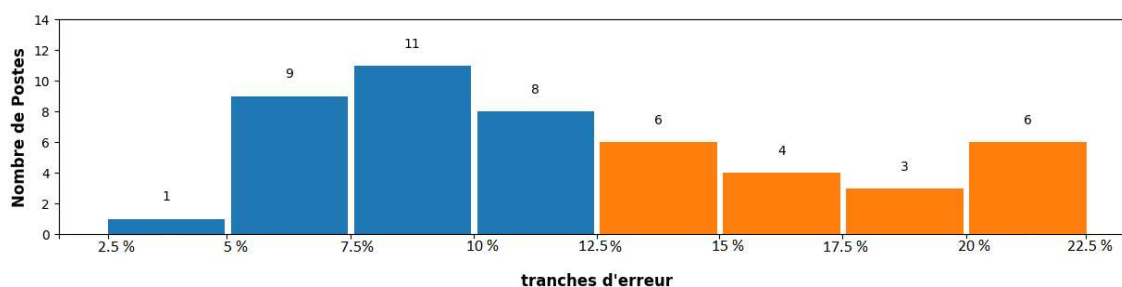


Figure 4-17 : Nombre de postes par tranche d'erreur

Ce regroupement des postes nous permet de mieux comprendre le comportement des erreurs du modèle. Le poste avec l'erreur la plus basse est le Poste Exemple. Le *Poste Exemple* pour lequel l'erreur était la plus faible a été choisi complètement au hasard, il n'a pas été sélectionné pour cette raison. De plus, notre modèle ne comporte pas de nombreux paramètres qui pourraient avoir conduit à un sur-apprentissage. La valeur de 12,5 % d'erreur a été choisie comme seuil pour diviser les postes en deux groupes. Ces 29 postes seront appelés "postes à faibles erreurs", tandis que les 19 autres seront appelés "postes à fortes erreurs". Cette division sera utilisée pour essayer de comprendre s'il y a des différences entre ces deux groupes qui expliquent la différence des erreurs. Les quatre caractéristiques seront analysées :

- Caractéristique 1 : Nombre de clients ;
- Caractéristique 2 : Nombre de résidences secondaires ;
- Caractéristique 3 : Energie moyenne du poste ;
- Caractéristique 4 : Valeurs maximales de la dérivée seconde de la courbe de charge du poste ;

La première caractéristique des postes comparée était le nombre de clients en aval des postes. Cela nous permet de comprendre si le modèle a besoin d'un nombre spécifique de clients pour obtenir de bons résultats ou non. Comme le nombre de gaussiennes par cluster est calculé à partir de l'historique des occurrences de la puissance maximale, nous pouvons supposer que plus le nombre de clients en aval du poste est faible, plus il est difficile de capturer les tendances de chaque cluster. Par conséquent, cela pourrait rendre le nombre de gaussiennes

et leurs positions imprécises. La deuxième caractéristique analysée est le nombre de clients résidentiels secondaires des deux groupes. Comme ce sont des types de clients atypiques, ils peuvent avoir une consommation d'énergie très variable, allant de zéro à une consommation normale. Il est donc possible que de les avoir en grande quantité puisse causer des difficultés dans l'optimisation. En effet, la pénalité peut devenir plus complexe en raison de leur comportement inhabituel. Nous avons poursuivi notre analyse en examinant maintenant la caractéristique de la consommation des postes. Nous avons cherché à savoir si l'énergie consommée par le poste tout au long de la saison était liée aux erreurs. L'objectif est de comprendre si un certain niveau de consommation, qu'il soit très bas ou très élevé, peut entraîner des difficultés dans la partie de la pénalisation du modèle. De plus, nous avons décidé d'examiner la deuxième dérivée de ces courbes. Elles peuvent nous donner une indication si notre hypothèse pour régler les limites des paramètres $c_{t,k}$ de notre gaussienne est vraie. Si la courbe de charge de ces postes présente des changements de comportement très brusques et variés, cela nous indique que l'hypothèse doit être ajustée. Les figures 4-18 à 4-20 correspondent respectivement aux caractéristiques 1 à 3 :

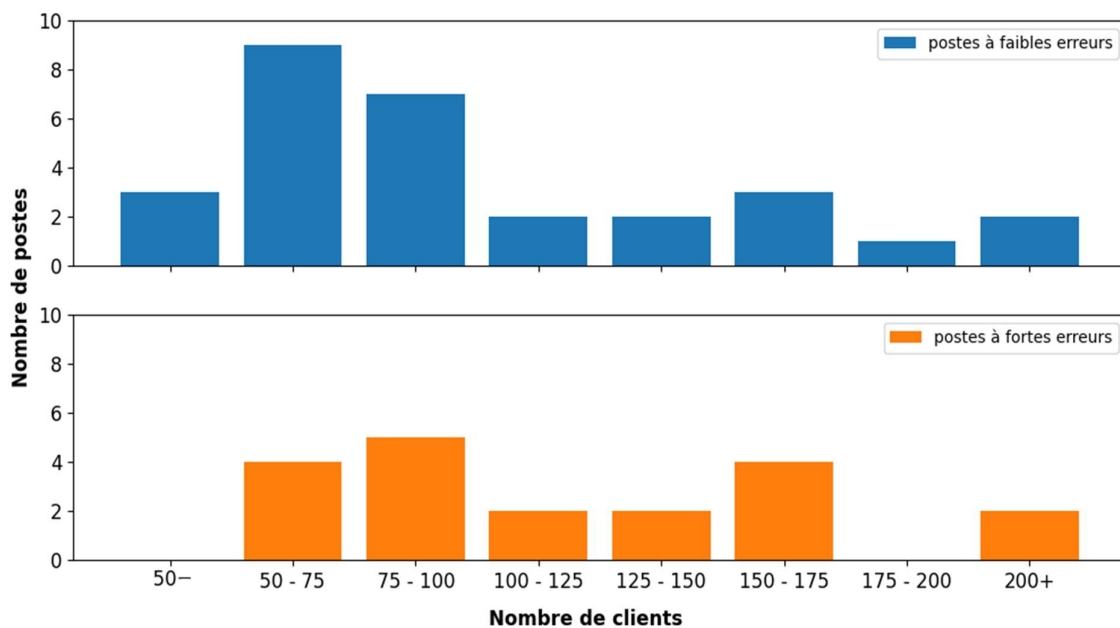


Figure 4-18 : Histogramme du nombre de clients pour les deux groupes des postes

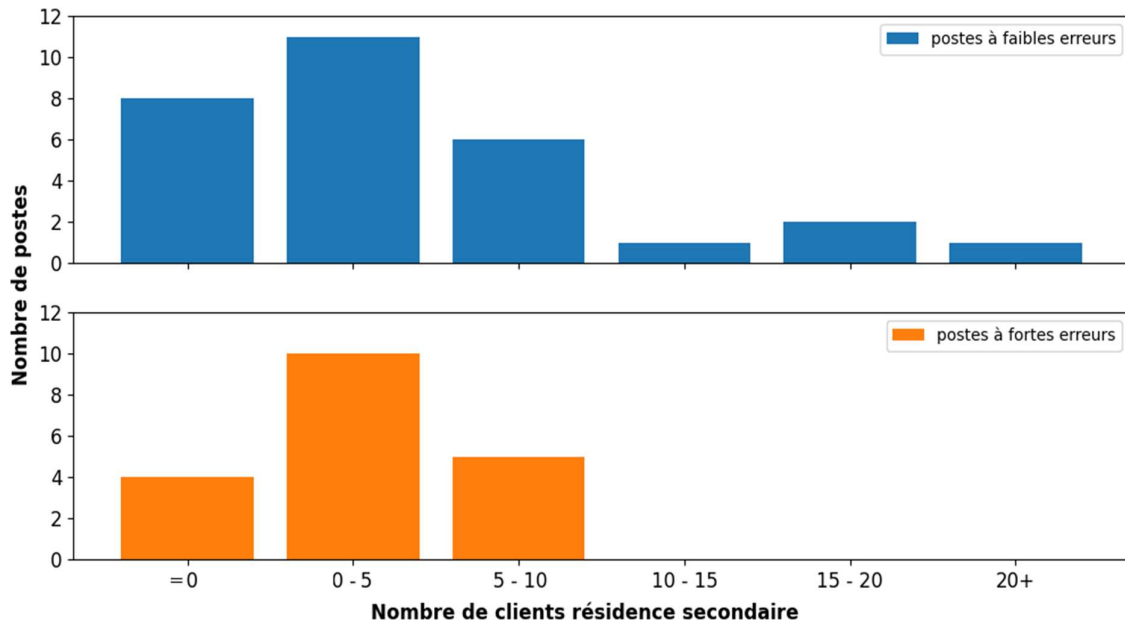


Figure 4-19 : Histogramme du nombre de clients résidence secondaire pour les deux groupes des postes

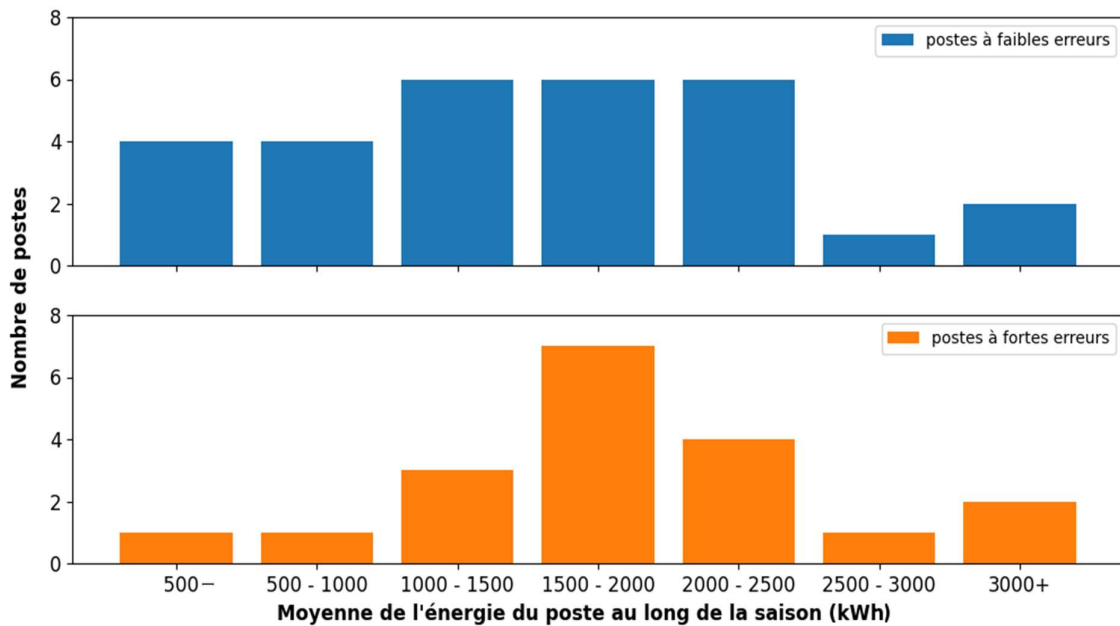


Figure 4-20 : Histogramme de l'énergie moyenne pour les deux groupes des postes

La Figure 4-18 nous montre que les postes à faibles erreurs présentent différents nombres de clients, allant de postes avec moins de 50 clients à des postes avec plus de 250. Cela confirme que le modèle de désagrégation est capable d'obtenir de bons résultats indépendamment du nombre de clients et que le nombre de clients du poste n'a pas de relation directe avec les

erreurs du modèle. La Figure 4-19 nous montre que le groupe de postes à faibles erreurs comprend des postes avec divers nombres de clients résidentiels secondaires. Cela indique que le modèle est adapté pour traiter ce type de client et que le nombre de clients résidence secondaire du poste n'a pas de relation directe avec les erreurs du modèle. La Figure 4-20 nous montre que les postes à faibles erreurs ainsi que les postes à fortes erreurs présentent différents niveaux de consommation allant de postes avec une moyenne de 500 kWh à des postes avec plus de 3000 kWh. Ce dernier point ne nous permet pas de conclure quant à la corrélation entre le niveau d'énergie et les erreurs. La Figure 4-21 montre la différence concernant la caractéristique 4 pour les deux groupes de postes :

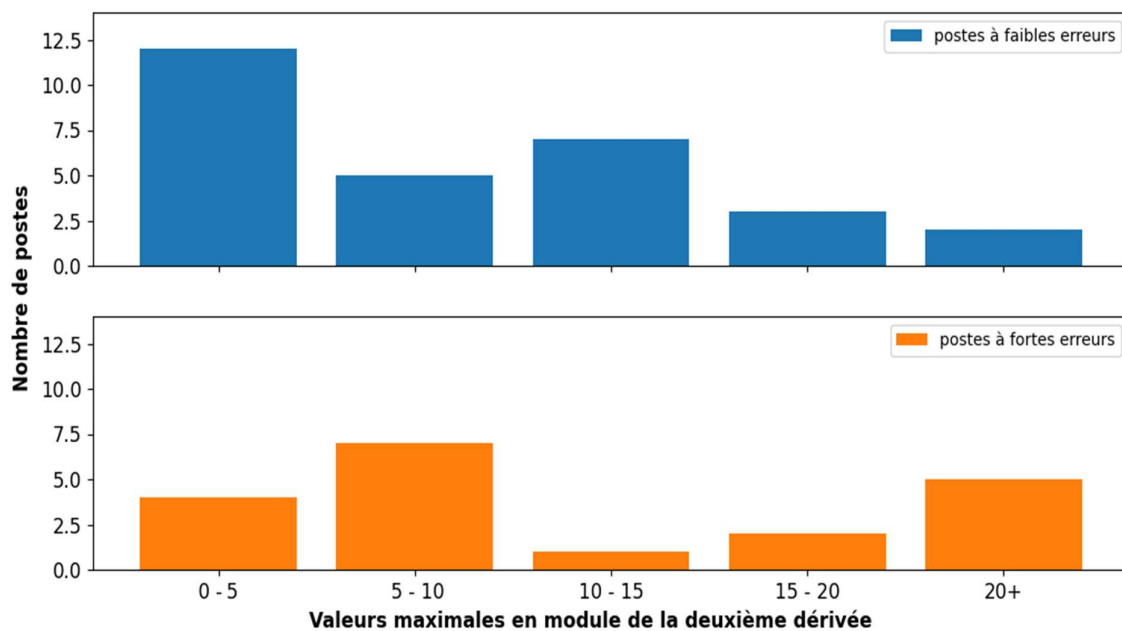


Figure 4-21 : Histogramme des valeurs maximales en module de la deuxième dérivée pour les deux groupes des postes

Il est possible d'observer que certains postes à fortes erreurs présentent des valeurs de la deuxième dérivée plus élevées que les postes à faibles erreurs. Cela nous indique que ce ne sont pas tous les postes qui présentent une grande inertie de consommation. La Figure 4-22 montre la courbe de charge d'une poste à fortes erreurs pour une journée aléatoire de la période :

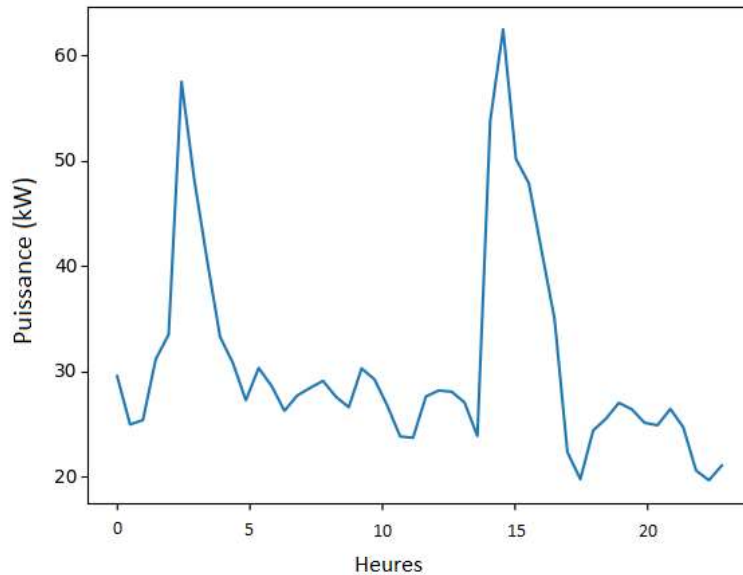


Figure 4-22 : Courbe de charge d'une journée au hasard pour une poste à forte erreur

Il est possible d'observer que la courbe de charge présente des pointes plus prononcées avec de grandes valeurs de la deuxième dérivée. Par conséquent, nous pouvons comprendre que ce ne sont pas tous les postes pour lesquels notre pour régler les limites des paramètre $c_{t,k}$ de notre gaussienne est valide. Un ajustement des limites du paramètre $c_{t,k}$ de notre modèle peut être une piste pour nous aider à réduire l'erreur. De plus, un autre facteur qui peut entraîner des erreurs est un mauvais rattachement d'un client dans le système d'informations géographiques. Comme le modèle pénalise les écarts d'énergie, avoir un client mal rattaché au bon poste peut être une cause d'erreurs. Cependant, cette piste est plus difficile à explorer car il n'est pas simple d'identifier si le SIG présente des erreurs de rattachement et quel est le client responsable des erreurs. Nous pouvons conclure, à partir de cette analyse, que nous devons encore travailler davantage pour comprendre et améliorer la performance du modèle 4 face aux erreurs de l'ajustement de courbes. Le tableau 4-6 illustre les forces et faiblesses des modèles 3 et 4 ainsi que la maille du réseau où ces modèles peuvent être mieux appliqués.

	Modèle 3 « somme de gaussiennes- Pmax »	Modèle 4 « somme de gaussiennes- Pmax-énergies »
(+)	Estimation de la courbe du Poste HTA/BT.	Cohérence énergétique avec les mesures Linky des clients
(-)	Cohérence énergétique avec les mesures Linky des clients	Estimation de la courbe du Poste HTA/BT.
Maille du réseau	HTA	BT

Tableau 4-6 : Forces et faiblesses des modèles de désagregation

Le modèle 3, en raison de son estimation plus précise de la courbe sur la maille du Poste HTA/BT, peut mieux servir aux applications en HTA. Cela est dû au fait qu'en HTA, il n'est pas nécessaire d'avoir une estimation de ce qui se passe en aval du Poste HTA/BT. Quant au modèle 4, en raison de sa cohérence énergétique avec les mesures Linky, il est plus opportun pour les applications en BT.

Une comparaison entre les résultats produits par le modèle de désagrégation et par le modèle classique Bagheera a été réalisée pour comprendre la qualité du modèle de désagrégation dans un contexte d'utilisation réelle. Cependant, les résultats de cette comparaison ne seront pas présentés ici pour des raisons de confidentialité.

4.5 Conclusions

Tout au long du chapitre 1, nous pouvons comprendre l'évolution des usages du réseau. Dans ce contexte d'électrification accrue, plus d'électricité sera consommée par les clients du réseau et le rôle du réseau de distribution deviendra de plus en plus important dans la transition énergétique. Ainsi, avoir des modèles de consommation des clients qui s'adaptent à ces évolutions est essentiel pour accueillir ces nouvelles charges sur le réseau. Grâce à notre contribution (C1) présentée dans le chapitre 2, nous pouvons comprendre comment les données des compteurs ont été exploitées à travers notre cartographie. Elle nous a montré que le sujet de la modélisation de la consommation à partir des données des compteurs intelligents pour améliorer la capacité d'accueil du réseau n'est pas très exploré. C'est pourquoi nous plongeons dans ce sujet tout au long de ce chapitre 4 pour apporter la contribution (C3). Le but de ce chapitre était de développer un modèle de courbes de charge client de manière innovante et disruptive. Pour ce faire, nous avons décidé d'inverser le processus naturel, qui consiste à partir du développement de courbes clients agrégés pour obtenir une estimation de la courbe du poste HTA/BT. En exploitant la richesse des données des compteurs Linky, nous avons créé une manière de désagréger la courbe du poste à partir de cette mesure en courbes client. Cette idée de modélisation n'existe pas dans la littérature scientifique, ce qui fait de ce modèle quelque chose d'innovant et qui pourrait ouvrir des portes à la réflexion sur la manière d'utiliser les données des compteurs intelligents pour aider au dimensionnement du réseau, à l'intégration de nouveaux usages et à l'adaptation aux évolutions des comportements des consommateurs en général. Nous avons réalisé un modèle en deux étapes, chacune d'entre eux ayant été détaillé tout au long du chapitre. Une modélisation de construction simple, avec une utilisation minimale des données, ce qui facilite sa reproductibilité. De plus, cette caractéristique simple lui permet d'évoluer, ce qui montre

le potentiel de l'utilisation des données des compteurs intelligents. Nous avons pu parvenir à deux formulations de cette modélisation qui tiennent compte des comportements électriques de consommation des clients. Ainsi, cette modélisation de désagrégation devient un outil qui s'adapte aux comportements réels du réseau. La désagrégation proposée par le modèle 3 (« somme de gaussiennes-Pmax ») est plus propice à être utilisé dans des applications HTA grâce à son estimation précise de la courbe sur la maille du Poste HTA/BT. Quant au modèle 4 (« somme de gaussiennes-Pmax-énergies »), il est plus adapté aux applications BT car il maintient une cohérence énergétique avec les mesures Linky. Le modèle 4 offre une marge d'évolution pour réduire l'erreur de ses estimations de la courbe sur la maille du Poste HTA/BT. Ainsi, la dernière contribution scientifique (C3) relative à modélisation des courbes de charges des clients du réseau BT est présentée dans ce travail de thèse.

5 Conclusions et Perspectives

Arrivé au terme de ce manuscrit, un bilan des travaux effectués est présenté. Les perspectives de ces travaux sont ensuite discutées.

5.1 Synthèse des travaux effectués

Ce travail de thèse avait pour objectif de démontrer comment les données des compteurs de charge Linky pourraient contribuer à une meilleure connaissance des charges du réseau BT. Pour mieux comprendre cela, nous avons plongé dans le contexte actuel du réseau BT au chapitre 1. Nous avons ainsi vu la mutation sans précédent que sont en train de vivre les réseaux électriques qui accompagnent notre transition écologique. Cette dernière est vitale pour notre monde fortement impacté durant l'antropocène par notre utilisation importante des ressources fossiles. Ce changement implique une augmentation croissante de la consommation d'électricité pour réduire les émissions gaz à effet de serre. Ce changement implique également une production d'énergie de plus en plus importante à partir de sources renouvelables, qu'elles soient ou pas décentralisées, contribuant ainsi à un flux bidirectionnel dans le réseau. Pour essayer de résoudre partiellement les impacts négatifs de ces nouvelles sources et pour accompagner les investissements nécessaires dans cette transition écologique initiés par l'ouverture des marchés de l'énergie, le réseau électrique évolue et devient "intelligent". Le pilier de cette transformation du réseau électrique réside dans l'évolution du parc de compteurs grâce aux compteurs communicants.

Avec la compréhension de ce contexte de changements à la fois dans l'utilisation du réseau et dans le réseau lui-même, nous avons défini les trois contributions décrites de ce manuscrit. Pour illustrer, de manière générale, comment les données des compteurs communicants peuvent contribuer à une meilleure connaissance des charges du réseau BT, nous avons créé une nouvelle cartographie de l'utilisation des données de ces compteurs, dans le chapitre 2. Cette cartographie fait partie d'une des contributions de la thèse (C1). Elle a été fondée sur une recherche des principaux acteurs du sujet : les GRD et la recherche scientifique. Du point de vue des GRD, nous avons compris que l'évolution du parc de compteurs était initialement liée à trois facteurs : la libéralisation du marché de l'énergie, l'intérêt de fournir des factures d'énergie adaptées aux clients finaux et les aspects automatisation et télérelève. A partir de ce changement dans le parc de compteurs, nous avons pu comprendre les avantages de l'utilisation des données des compteurs intelligents constatés par ces GRD. Ces avantages ont été répartis en 4 catégories : client final, marché, technique et intégration des énergies renouvelables. Ces catégories ont servi de guide pour trier les 904 articles de recherche par thèmes. De cette manière, nous avons constitué notre cartographie de l'utilisation des données des compteurs intelligents. Celle-ci peut guider les GRD en phase de changement de leur parc de comptage pour utiliser les données afin de répondre aux enjeux de leur réseau. Elle peut également aider les chercheurs à trouver plus facilement des références pour leurs sujets de recherche.

À partir de notre nouvelle cartographie d'utilisation des données, nous pouvons comprendre qu'il y a un énorme espace pour discuter de la capacité d'accueil du réseau de distribution. Un sujet très important pour la discussion de ce thème est le foisonnement des charges. Ainsi, avec la contribution (C2), nous visons à approfondir ce sujet à travers l'estimation et l'analyse du foisonnement des charges en BT. Au chapitre 3, nous avons commencé par aborder ce que cela mesure réellement et clarifier la compréhension de son sens. Ensuite, nous avons parcouru la littérature sur le sujet, qui trouve les premières contributions au début du XXe siècle. À partir de cette littérature, nous pouvons comprendre que les articles se concentrent sur la relation entre la valeur du coefficient de foisonnement et le nombre de clients. Dans cette optique, notre recherche visait à mieux comprendre comment d'autres facteurs, en plus du nombre de clients, étaient importants pour la variation de ce coefficient. Notre objectif était de construire un modèle pour prédire ce coefficient à partir de ces autres facteurs. Grâce aux données des compteurs Linky et Pinky, nous avons pu enquêter sur ce sujet et construire un modèle avec une erreur de 12,9 % pour notre ensemble de données. De plus, nous avons constaté la cohérence de notre modèle avec les comportements de consommation classiques en France. Ensuite, ce modèle de prédiction a été comparé à un autre modèle conçu à partir d'une méthodologie classique (inspirée par la norme NF C14-100) et a obtenu un résultat 50 % meilleur en moyenne pour notre ensemble de données. Ces résultats nous montrent l'apport de l'utilisation des données des compteurs communicants pour l'estimation de la valeur du coefficient de foisonnement. Une meilleure estimation de ce coefficient entraîne une meilleure estimation de la pointe sur le poste HTA/BT, ce qui permet au planificateur d'avoir une meilleure compréhension de la capacité d'accueil du réseau, et par conséquent, de mieux orienter les investissements.

Enfin dans le chapitre 4, nous visons à aborder la modélisation de la courbe de charge des clients du réseau BT de manière innovante, dans le cadre de la contribution (C3). Ce qui nous pousse à explorer ce sujet est le fait que dans notre cartographie de la contribution (C1), la modélisation des charges des clients est réalisée à partir de méthodes d'estimation de la courbe de charge des appareils de type « *bottom-up* ». Ce type de modélisation demande des mesures temporelles fines, ce que tous les GRD n'ont pas nécessairement. Ainsi, notre modélisation, en plus d'être innovante, peut être plus facilement explorée. Cette modélisation est innovante car il n'a pas été trouvé d'approche similaire dans la littérature. Appelé modèle de désagrégation, il vise à obtenir, à partir de la courbe de charge du poste HTA/BT, K courbes correspondant aux groupes de clients présentant des comportements similaires. Pour ce faire, le modèle est divisé en deux étapes qui peuvent être réalisées grâce aux données des compteurs Pinky et Linky. La première étape consiste à segmenter les clients en K groupes à partir des données de puissance maximale mesurées par le compteur Linky, en écartant les

clients dont la consommation n'est pas représentative. La seconde étape consiste à désagréger la courbe de charge du poste mesurée par le compteur Pinky en K courbes. Dans cette deuxième étape, nous présentons différentes méthodes pour réaliser ce processus, de la plus simple à la plus complexe, en prenant en compte les caractéristiques réelles du réseau. En tenant compte des caractéristiques du réseau, nous avons abouti à la formulation de deux modèles. Le modèle 3 (« somme de gaussiennes-Pmax ») est plus propice à être utilisé dans des applications HTA car il offre une meilleure estimation de la courbe de charge du Poste HTA/BT. Quant au modèle 4 (« somme de gaussiennes-Pmax-énergies »), il est plus adapté pour des applications BT en raison de sa cohérence énergétique avec les mesures Linky. Cela démontre la pertinence de notre modèle de désagrégation. Avec cette modélisation, nous pourrions envisager de nouvelles solutions de dimensionnement du réseau, en augmentant sa capacité d'accueil et en activant des flexibilités.

5.2 Perspectives

Le travail réalisé dans ce manuscrit de thèse répond partiellement à la question de savoir quelles sont les apports des mesures des compteurs Linky (et de tout autre compteur communicant) à la connaissance des charges du réseau BT. En général, grâce à la cartographie et de manière plus approfondie avec les modèles de charge BT, nous avons montré comment les données mesurées par les compteurs communicants peuvent bénéficier au réseau BT.

5.2.1 Court terme

En ce qui concerne la thématique du foisonnement, une réflexion peut porter sur le développement de deux modèles de prédiction distincts au lieu d'un seul, un pour les postes HTA/BT sans clients industriels et un autre pour les postes HTA/BT avec des clients industriels. Étant donné que la proportion de postes HTA/BT avec des clients industriels est inférieure à celle des postes sans ces clients, cela pourrait permettre de mieux prendre en compte l'effet de ces clients dans l'estimation du coefficient de foisonnement. De plus, étant donné que les clients industriels ont tendance à avoir une consommation nettement plus élevée que les autres clients, cela pourrait expliquer en partie l'erreur de prédiction plus élevée pour les postes avec ce type de client. Une autre perspective de recherche sur ce sujet serait d'étudier comment la production décentralisée (en particulier l'installation de panneaux solaires sur les toits des bâtiments) affecte le coefficient de foisonnement. De plus, avec l'augmentation de

la production dans le réseau BT, il serait envisageable de calculer un coefficient de foisonnement de la production.

Concernant le thème de la modélisation des courbes de charges BT à partir de la désagrégation, une première réflexion concerne la difficulté d'évaluer les courbes désagrégées par notre modèle par rapport aux courbes de charge réelles des clients. Cela est dû au fait que, jusqu'à présent, il n'y a pas de poste HTA/BT où tous les clients sont abonnés à cette fonction, ce qui rend le travail d'évaluation plus complexe. Une façon de faire face à cela serait de construire un poste HTA/BT fictif composé uniquement de clients abonnés à cette fonction. La courbe de charge de ce poste fictif serait alors la somme des courbes de charge de tous les clients, et elle servirait de courbe à désagréger. De cette manière, il serait possible d'évaluer la sortie de notre modèle par rapport à ce qui est réellement observé dans le réseau. Un deuxième point de travail concerne le modèle 4. Il est nécessaire de comprendre pourquoi il éprouve des difficultés à maintenir une cohérence énergétique avec les mesures Linky pour les clusters à faible consommation énergétique. À partir de cela, il sera possible d'améliorer l'estimation de la courbe de charge du Poste HTA/BT.

5.2.2 Long terme

Sur le long terme, nous pouvons envisager d'enrichir la cartographie des cas proposée par (C1) de manière à ce qu'elle reste une ressource mise à jour pour visualiser les thèmes centraux de recherche à partir des données des compteurs intelligents. Par conséquent, cela montrerait également les zones peu explorées, pouvant orienter de nouvelles recherches dans ces domaines. Une autre perspective envisagée est l'évaluation de l'impact pour le GRD de l'utilisation des modèles proposés par les contributions (C2 et C3). À partir de cette évaluation, nous serons en mesure d'évoluer nos modèles afin de converger vers les différents besoins des GRDs.

Références

- [1] John Bellamy Foster (1992) The absolute general law of environmental degradation under capitalism, *Capitalism Nature Socialism*, 3:3, 77-81, DOI: [10.1080/10455759209358504](https://doi.org/10.1080/10455759209358504)
- [2] Moore, J. W. (2000). Environmental crises and the metabolic rift in world-historical perspective. *Organization & environment*, 13(2), 123-157.
- [3] World Commission on Environment and Development. (1987). *Our common future*. Oxford University Press.
- [4] Négociations internationales | Chiffres clés du climat. (n.d.). Chiffres Clés Du Climat. <https://www.statistiques.developpement-durable.gouv.fr/edition-numerique/chiffres-cles-du-climat/16-negociations-internationales>
- [5] « Production d'électricité et ses émissions de CO2. » <https://www.planete-energies.com/fr/media/article/production-deelectricite-ses-emissions-co2>
- [6] « Electricity consumption – Electricity Information: Overview – Analysis - IEA, » IEA. <https://www.iea.org/reports/electricity-information-overview/electricity-consumption>
- [7] « Bilan previsionnel 2050 futurs énergetiques. » RTE. <https://www.rte-france.com/analyses-tendances-et-prospectives/bilan-previsionnel-2050-futurs-energetiques>
- [8] G. Leach, "The energy transition," *Energy Policy*, vol. 20, no. 2, pp. 116–123, Feb. 1992, doi: 10.1016/0301-4215(92)90105-b.
- [9] LOI n° 2009-967 du 3 août 2009 de programmation relative à la mise en œuvre du Grenelle de l'environnement (1) - Légifrance. (n.d.). <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000020949548>
- [10] LOI n° 2010-788 du 12 juillet 2010 portant engagement national pour l'environnement (1) - Légifrance. (n.d.). <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000022470434>
- [11] LOI n° 2015-992 du 17 août 2015 relative à la transition énergétique pour la croissance verte (1) - Légifrance. (n.d.). <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000031044385>
- [12] éCO2mix - La production d'électricité par filière. (n.d.). RTE. <https://www.rte-france.com/eco2mix/la-production-deelectricite-par-filiere>
- [13] Open Data | Enedis. (n.d.). Enedis. <https://www.enedis.fr/open-data>
- [14] Petinrin, J. O., & Shaabanb, M. (2016). Impact of renewable generation on voltage control in distribution systems. *Renewable and Sustainable Energy Reviews*, 65, 770-783.
- [15] Émissions de gaz à effet de serre par activité | Insee. (n.d.). <https://www.insee.fr/fr/statistiques/2015759#tableau-figure1>

- [16] [Baromètre] Des immatriculations et parts de marché records pour les véhicules 100 % électriques en février 2023 - Avere-France. (2023, March 23). Avere-France. <https://www.aver-france.org/publication/barometre-des-immatriculations-et-parts-de-marche-records-pour-les-vehicules-100-electriques/>
- [17] Guille, C., & Gross, G. (2009). A conceptual framework for the vehicle-to-grid (V2G) implementation. *Energy policy*, 37(11), 4379-4390.
- [18] Consommation d'énergie par usage du résidentiel. (n.d.). Données Et Études Statistiques Pour Le Changement Climatique, L'énergie, L'environnement, Le Logement Et Les Transports. <https://www.statistiques.developpement-durable.gouv.fr/consommation-denergie-par-usage-du-residentiel>
- [19] *Smart grids* - IEA. (n.d.). IEA. <https://www.iea.org/energy-system/electricity/smart-grids>
- [20] Directive 96/92/CE du parlement européen et du conseil du 19 décembre 1996 concernant des règles communes pour le marché intérieur de l'électricité - Légifrance. (n.d.). <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000889133>
- [21] Botte, B., Cannatelli, V., & Rogai, S. (2005, June). The Telegestore project in ENEL's metering system. In CIREN 2005-18th International Conference and Exhibition on Electricity Distribution (pp. 1-4). IET.
- [22] Huang, Y., Grahn, E., Wallnerström, C. J., & Johansson, T. (2018). Smart meters in Sweden – Lesson learned and new regulations. ResearchGate. https://www.researchgate.net/publication/335940105_Smart_meters_in_Sweden_-_Lesson_learned_and_new_regulations
- [23] La CRE : favorable à la généralisation du compteur Linky. (n.d.). CRE. <https://www.cre.fr/documents/Presse/Communiqués-de-presse/la-cre-favorable-a-la-generalisation-du-compteur-linky>
- [24] Think Smartgrids. (2016, November 28). Les principaux démonstrateurs Smart Grids français : Think Smartgrids. <https://www.thinksmartgrids.fr/actualites/principaux-demonstrateurs-smart-grids-francais>
- [25] EUR-LEX - 32009L0072 - EN - EUR-LEX. (n.d.). <https://eur-lex.europa.eu/legal-content/fr/ALL/?uri=CELEX%3A32009L0072>
- [26] Your gateway to the EU, News, Highlights | European Union. (n.d.). European Union. <http://www.europa.eu/>
- [27] Chudjakov, N. (2010). Enel: Italy reaping first-mover benefits of smart meters. www.euractiv.com. <https://www.euractiv.com/section/climate-environment/interview/enel-italy-reaping-first-mover-benefits-of-smart-meters/>
- [28] The Circular Smart Meter: sustainable innovation by Enel. (2022, March 29). <https://www.enel.com/company/stories/articles/2022/03/circular-smart-meter-regenerated-plastic>

- [29] Piti, A., Di Stefano, L., Mardero, D., Boscagin, A., Kildani, M., & Ceneri, G. (2019). Smart Metering 2.0 enhancing a new customer experience. CIRED. <https://doi.org/10.34890/792>
- [30] Huang, Y., Grahn, E., Wallnerström, C. J., & Johansson, T. (2018). Smart meters in Sweden – Lesson learned and new regulations. ResearchGate. https://www.researchgate.net/publication/335940105_Smart_meters_in_Sweden_-_Lesson_learned_and_new_regulations
- [31] Smart Grid successful deployment in Sweden. (n.d.). <https://www.networkedenergy.com/en/success/details/12-networked-energy-services>
- [32] <https://www.mpo.cz/assets/cz/energetika/strategicke-a-koncepcni-dokumenty/narodni-akcni-plan-pro-chytre-site/2021/1/Prezentace-EON.pdf>
- [33] <https://www.eon.com/en/innovation/innovation-frontline/innovation-news/new-smart-service-in-sweden-elna-shows-eon-customers-the-true-potential-of-smart-meters.html>
- [34] Projecto Inovgrid: O futuro da energia | E-REDES. (n.d.). <https://www.e-redes.pt/pt-pt/transicao-energetica/inovacao-e-desenvolvimento/inovgrid>
- [35] Prata, R., Craveiro, H., Santos, C., & Quaresma, E. (2011). SmartGrid role in reducing electrical losses — The InovGrid experience. 11th International Conference on Electrical Power Quality and Utilisation. <https://doi.org/10.1109/epqu.2011.6128900>
- [36] Águas, A., Pereira, V. L. D. D. V., Roça, I., Jorge, L., Prata, R., Machado, J., Carvalho, P., & Ferreira, L. (2019). Data Analytics and Stochastic Simulation Methods for Risk-Controlled Network Planning: Validation Case Study. CIRED. <https://doi.org/10.34890/366>
- [37] Nunes, V., Carli, C., Rolim, C., Dordio, M., & Mota, T. (2019). DSO role in the deployment of Smart Cities solutions: the case of the Lisbon Urban Sharing Platform as a service provider. CIRED. <https://doi.org/10.34890/543>
- [38] Nós e a Transição | E-REDES. (n.d.). <https://www.e-redes.pt/pt-pt/sustentabilidade/nos-e-a-transicao>
- [39] ATEP - Homepage. (2021, July 19). Atep. <https://atep.e-redes.pt/index.php>
- [40] <https://www.boe.es/eli/es/rd/2006/12/29/1634>
- [41] <https://www.boe.es/eli/es/rd/2006/06/30/809>
- [42] The smart meter. (n.d.). Endesa. <https://www.endesa.com/en/the-e-face/electrical-grid/smart-meter>
- [43] Endesa. (2023, June 27). Knowing more (about your consumption) will liberate you. Endesa. <https://www.endesa.com/en/blogs/endesa-s-blog/light/understand-electricity-consume>
- [44] Iberdrola. (2021). SMART METERS. Iberdrola. <https://www.iberdrola.com/innovation/smart-meters>

- [45] Smart meters will help electric vehicles move into the mainstream. (n.d.). Financial Times - Paid Post by Iberdrola. <https://iberdrola.ft.com/smart-meters-will-help-electric-vehicles-move-into-the-mainstream>
- [46] Update on the rollout of smart meters - National Audit Office (NAO) report. (2023, June 22). National Audit Office (NAO). <https://www.nao.org.uk/reports/update-on-the-rollout-of-smart-meters/>
- [47] Should I get a smart meter? | Smart Energy GB. (n.d.). <https://www.smartenergygb.org/smart-meter-benefits/benefits-for-you#PrepayBFY>
- [48] How smart meters could help improve healthcare | Smart Energy GB. (n.d.). <https://www.smartenergygb.org/smart-meter-benefits/benefits-for-you/how-smart-meters-could-transform-health-and-social-care>
- [49] Van Aubel, P., & Poll, E. (2019). Smart metering in the Netherlands: What, how, and why. International Journal of Electrical Power & Energy Systems, 109, 719–725. <https://doi.org/10.1016/j.ijepes.2019.01.001>
- [50] Upgrading your energy meters | Liander. (n.d.). <https://www.liander.nl/smartmeter>
- [51] Voordelen | Stedin. (n.d.). Stedin. <https://www.stedin.net/slimme-meter/voordelen>
- [52] Voordelen van de slimme meter | Consument | Liander. (n.d.). https://www.liander.nl/consument/slimme_meter/why-smart-meter
- [53] Frequently asked questions (FAQs) - U.S. Energy Information Administration (EIA). (n.d.). <https://www.eia.gov/tools/faqs/faq.php?id=108&t=1>
- [54] Analyze residential power usage. (n.d.). https://www.pge.com/en_US/residential/save-energy-money/analyze-your-usage/your-usage/your-usage.page
- [55] SmartMeter FAQ. (n.d.). https://www.pge.com/en_US/residential/save-energy-money/analyze-your-usage/your-usage/view-and-share-your-data-with-smartmeter/smartmeter-faq.page
- [56] Sullivan, M. W., Bode, J., Kellow, B., Woehleke, S., & Eto, J. H. (2013). Using Residential AC Load Control in Grid Operations: PG&E's Ancillary Service Pilot. IEEE Transactions on Smart Grid, 4(2), 1162–1170. <https://doi.org/10.1109/tsg.2012.2233503>
- [57] Fighting The Switch? The Smart Meter Controversy. (2014, May 12). CBS News. <https://www.cbsnews.com/miami/news/fighting-the-switch-the-smart-meter-controversy/>
- [58] FPL unveils new “Power Delivery Diagnostic Center.” (n.d.). Florida Trend. <https://floridatrend.com/article/18280/fpl-launches-power-delivery-diagnostic-center--a-smart-grid-technology-center>
- [59] <https://www.fpl.com/smart-meters/pdf/energy-use.pdf>
- [60] <https://www.fpl.com/smart-meters/pdf/energy-smart-fact-sheet.pdf>

- [61] Sanderson, B. (2016, June 6). Con Ed's new meters tell how much gas, electric you use. New York Post. <https://nypost.com/2016/06/04/con-eds-new-meters-tell-how-much-gas-electric-you-use/>
- [62] All about Con Edison Smart Meters | Con Edison. (n.d.). <https://www.coned.com/en/our-energy-future/our-energy-vision/where-we-are-going/smart-meters>
- [63] Smart Meter features and benefits | Con Edison. (n.d.). <https://www.coned.com/en/our-energy-future/our-energy-vision/where-we-are-going/smart-meters/how-will-a-smart-meter-help-me>
- [64] Canada Electricity Security Policy – Analysis - IEA. (n.d.). IEA. <https://www.iea.org/articles/canada-electricity-security-policy-2>
- [65] <https://www.bchydro.com/content/dam/BCHydro/customer-portal/documents/projects/smart-metering/smi-program-business-case.pdf>
- [66] https://www.capgemini.com/wp-content/uploads/2017/07/ss_The_Power_of_Smart_Metering.pdf
- [67] Who is my distributor? (2019, April 9). Australian Energy Regulator. <https://www.aer.gov.au/consumers/who-is-my-distributor>
- [68] Review announced into how electricity smart meters can deliver more customer benefits. (n.d.). AEMC. <https://www.aemc.gov.au/news-centre/media-releases/review-announced-how-electricity-smart-meters-can-deliver-more-customer>
- [69] Government of South Australia. (2023, February 23). Smart meters. <https://www.sa.gov.au/topics/energy-and-environment/meters/smart-meters>
- [70] Review announced into how electricity smart meters can deliver more customer benefits. (n.d.). AEMC. <https://www.aemc.gov.au/news-centre/media-releases/review-announced-how-electricity-smart-meters-can-deliver-more-customer>
- [71] Metering review for a smarter energy future. (n.d.). AEMC. <https://www.aemc.gov.au/news-centre/media-releases/metering-review-smarter-energy-future>
- [72] Endorsement for wholesale demand response - Ausgrid. (n.d.). <https://www.ausgrid.com.au/Connections/Connection-types/Endorsement-for-wholesale-demand-response>
- [73] https://www.aemc.gov.au/sites/default/files/documents/rule_change_submission_-_emo0040_-_intellihub_-_20210211.pdf
- [74] https://www.enecho.meti.go.jp/en/category/others/basic_plan/pdf/4th_strategic_energy_plan.pdf
- [75] Smart Meter Project | TEPCO. (n.d.). <https://www.tepco.co.jp/en/pg/development/domestic/smartmeter-e.html>

- [76] https://www.energypartnership-korea.org/fileadmin/user_upload/korea/media_elements/8-Study_2019_System_Integration_of_RE_and_Smart_Grids_in_Korea.pdf
- [77] Smart Meter, AMI development in Korea. (n.d.). Smart Meter, AMI Development in Korea. https://www.vinatech.com/winko.php?code=blog_e&v=eng&body=view&page=1&number=38&category=&keyfield=&key=
- [78] Son, W. (2018, September 30). AMI CASE Case06 / KOREA - ISGAN. ISGAN. <https://www.iea-isgan.org/ami-case-case06-korea/>
- [79] nationthailand. (2021, April 21). PEA eyes smart electricity grid for EEC provinces. Nationthailand. <https://www.nationthailand.com/business/30387012>
- [80] [https://www.pea.co.th/Webapplications/tor/Attachments/b1816cc7-350e-4ad9-9e34-2ca897d4b970/%E0%B8%81%E0%B8%9F%E0%B8%A0.%E0%B8%81%E0%B8%88%E0%B8%99.%E0%B8%9B\(%E0%B8%A1\)-003-2561_3%20%E0%B8%A3%E0%B8%B2%E0%B8%A2%E0%B8%81%E0%B8%B2%E0%B8%A3.pdf](https://www.pea.co.th/Webapplications/tor/Attachments/b1816cc7-350e-4ad9-9e34-2ca897d4b970/%E0%B8%81%E0%B8%9F%E0%B8%A0.%E0%B8%81%E0%B8%88%E0%B8%99.%E0%B8%9B(%E0%B8%A1)-003-2561_3%20%E0%B8%A3%E0%B8%B2%E0%B8%A2%E0%B8%81%E0%B8%B2%E0%B8%A3.pdf)
- [81] Post, J. (2020, June 17). PLN to install 79 million smart meters in Indonesia after billing fiasco. The Jakarta Post. <https://www.thejakartapost.com/news/2020/06/17/pln-to-install-79-million-smart-meters-in-indonesia-after-billing-fiasco.html>
- [82] U.S. International Trade Administration. (n.d.). Indonesia Smart Grid Market. International Trade Administration | Trade.gov. <https://www.trade.gov/market-intelligence/indonesia-smart-grid-market>
- [83] https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2023/04/loT-and-Essential-Utility-Services-Indonesia-market-case-study.pdf?utm_source=website&utm_medium=resourcepage&utm_campaign=loTEUSIndonesia
- [84] Energy Research – Pecan Street Inc. (2022, October 27). Pecan Street Inc. <http://www.pecanstreet.org/energy/>
- [85] Ratnam, E. L., Weller, S. R., Kellett, C. M., & Murray, A. T. (2015). Residential load and rooftop PV generation: an Australian distribution network dataset. International Journal of Sustainable Energy, 36(8), 787–806. <https://doi.org/10.1080/14786451.2015.1100196>
- [86] Irish Social Science Data Archive. (n.d.). ISSDA | Commission for Energy Regulation (CER). <http://www.ucd.ie/issda/data/commissionforenergyregulationcer/>
- [87] UK Data Service › Study. (n.d.). <https://discover.ukdataservice.ac.uk/catalogue?sn=7857>
- [88] Miller, C., & Meggers, F. (2017). The Building Data Genome Project: An open, public data set from non-residential building electrical meters. Energy Procedia, 122, 439–444. <https://doi.org/10.1016/j.egypro.2017.07.400>

- [89] B. Yildiz, J. Bilbao, J. Dore, et A. Sproul, « Recent advances in the analysis of residential electricity consumption and applications of smart meter data », *APPLIED ENERGY*, vol. 208, p. 402-427, déc. 2017, doi: 10.1016/j.apenergy.2017.10.014.
- [90] W. Kong, Z. Dong, J. Ma, D. Hill, J. Zhao, et F. Luo, « An Extensible Approach for Non-Intrusive Load Disaggregation With Smart Meter Data », *IEEE TRANSACTIONS ON SMART GRID*, vol. 9, n° 4, p. 3362-3372, juill. 2018, doi: 10.1109/TSG.2016.2631238.
- [91] Z. Guo, Z. Wang, et A. Kashani, « Home Appliance Load Modeling From Aggregated Smart Meter Data », *IEEE TRANSACTIONS ON POWER SYSTEMS*, vol. 30, n° 1, p. 254-262, janv. 2015, doi: 10.1109/TPWRS.2014.2327041.
- [92] M. Weiss, A. Helfenstein, F. Mattern, T. Staake, et IEEE, « Leveraging smart meter data to recognize home appliances », présenté à 2012 IEEE INTERNATIONAL CONFERENCE ON PERVASIVE COMPUTING AND COMMUNICATIONS (PERCOM), 2012, p. 190-197.
- [93] K. Armel, A. Gupta, G. Shrimali, et A. Albert, « Is disaggregation the holy grail of energy efficiency? The case of electricity », *ENERGY POLICY*, vol. 52, p. 213-234, janv. 2013, doi: 10.1016/j.enpol.2012.08.062.
- [94] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, et A. Rizzoli, « A Hybrid Signature-based Iterative Disaggregation algorithm for Non-Intrusive Load Monitoring », *APPLIED ENERGY*, vol. 185, p. 331-344, janv. 2017, doi: 10.1016/j.apenergy.2016.10.040.
- [95] D. Egarter, V. Bhuvana, et W. Elmenreich, « PALDi: Online Load Disaggregation via Particle Filtering », *IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT*, vol. 64, n° 2, p. 467-477, févr. 2015, doi: 10.1109/TIM.2014.2344373.
- [96] F. Quilumba, W. Lee, H. Huang, D. Wang, et R. Szabados, « Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities », *IEEE TRANSACTIONS ON SMART GRID*, vol. 6, n° 2, p. 911-918, mars 2015, doi: 10.1109/TSG.2014.2364233.
- [97] A. Albert et R. Rajagopal, « Smart Meter Driven Segmentation: What Your Consumption Says About You », *IEEE TRANSACTIONS ON POWER SYSTEMS*, vol. 28, n° 4, p. 4019-4030, nov. 2013, doi: 10.1109/TPWRS.2013.2266122.
- [98] J. Kwac, J. Flora, et R. Rajagopal, « Household Energy Consumption Segmentation Using Hourly Data », *IEEE TRANSACTIONS ON SMART GRID*, vol. 5, n° 1, p. 420-430, janv. 2014, doi: 10.1109/TSG.2013.2278477.
- [99] S. Haben, C. Singleton, et P. Grindrod, « Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data », *IEEE TRANSACTIONS ON SMART GRID*, vol. 7, n° 1, p. 136-144, janv. 2016, doi: 10.1109/TSG.2015.2409786.

- [100] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, et C. Kang, « Probabilistic individual load forecasting using pinball loss guided LSTM », *APPLIED ENERGY*, vol. 235, p. 10-20, févr. 2019, doi: 10.1016/j.apenergy.2018.10.078.
- [101] H. Shi, M. Xu, et R. Li, « Deep Learning for Household Load Forecasting-A Novel Pooling Deep RNN », *IEEE TRANSACTIONS ON SMART GRID*, vol. 9, n° 5, p. 5271-5280, sept. 2018, doi: 10.1109/TSG.2017.2686012.
- [102] S. Singh et A. Yassine, « Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting », *ENERGIES*, vol. 11, n° 2, févr. 2018, doi: 10.3390/en11020452.
- [103] M. Ribeiro, K. Grolinger, H. ElYamany, W. Higashino, et M. Capretz, « Transfer learning with seasonal and trend adjustment for cross-building energy forecasting », *ENERGY AND BUILDINGS*, vol. 165, p. 352-363, avr. 2018, doi: 10.1016/j.enbuild.2018.01.034.
- [104] L. Candanedo, V. Feldheim, et D. Deramaix, « Data driven prediction models of energy use of appliances in a low-energy house », *ENERGY AND BUILDINGS*, vol. 140, p. 81-97, avr. 2017, doi: 10.1016/j.enbuild.2017.01.083.
- [105] J. Rhodes, W. Cole, C. Upshaw, T. Edgar, et M. Webber, « Clustering analysis of residential electricity demand profiles », *APPLIED ENERGY*, vol. 135, p. 461-471, déc. 2014, doi: 10.1016/j.apenergy.2014.08.111.
- [106] J. Kwac et R. Rajagopal, « Data-Driven Targeting of Customers for Demand Response », *IEEE TRANSACTIONS ON SMART GRID*, vol. 7, n° 5, p. 2199-2207, sept. 2016, doi: 10.1109/TSG.2015.2480841.
- [107] Y. Huang, H. Tian, et L. Wang, « Demand response for home energy management system », *INTERNATIONAL JOURNAL OF ELECTRICAL POWER & ENERGY SYSTEMS*, vol. 73, p. 448-455, déc. 2015, doi: 10.1016/j.ijepes.2015.05.032.
- [108] M. Dyson, S. Borgeson, M. Tabone, et D. Callaway, « Using smart meter data to estimate demand response potential, with application to solar energy integration », *ENERGY POLICY*, vol. 73, p. 607-619, oct. 2014, doi: 10.1016/j.enpol.2014.05.053.
- [109] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, et S. Zonouz, « A Multi-Sensor Energy Theft Detection Framework for Advanced Metering Infrastructures », *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, vol. 31, n° 7, p. 1319-1330, juill. 2013, doi: 10.1109/JSAC.2013.130714.
- [110] K. Zheng, Q. Chen, Y. Wang, C. Kang, et Q. Xia, « A Novel Combined Data-Driven Approach for Electricity Theft Detection », *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, vol. 15, n° 3, p. 1809-1819, mars 2019, doi: 10.1109/TII.2018.2873814.
- [111] M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, et A. Gómez-Expósito, « Detection of Non-Technical Losses Using Smart Meter Data and Supervised Learning », *IEEE TRANSACTIONS ON SMART GRID*, vol. 10, n° 3, p. 2661-2670, mai 2019, doi: 10.1109/TSG.2018.2807925.

- [112] T. Ahmad, « Non-technical loss analysis and prevention using smart meters », *RENEWABLE & SUSTAINABLE ENERGY REVIEWS*, vol. 72, p. 573-589, mai 2017, doi: 10.1016/j.rser.2017.01.100.
- [113] J. Chou et I. Yutami, « Smart meter adoption and deployment strategy for residential buildings in Indonesia », *APPLIED ENERGY*, vol. 128, p. 336-349, sept. 2014, doi: 10.1016/j.apenergy.2014.04.083.
- [114] J. Chou, C. Kim, T. Ung, I. Yutami, G. Lin, et H. Son, « Cross-country review of smart grid adoption in residential buildings », *RENEWABLE & SUSTAINABLE ENERGY REVIEWS*, vol. 48, p. 192-213, août 2015, doi: 10.1016/j.rser.2015.03.055.
- [115] J. Nielsen, G. Madueño, N. Pratas, R. Sorensen, C. Stefanovic, et P. Popovski, « WHAT CAN WIRELESS CELLULAR TECHNOLOGIES DO ABOUT THE UPCOMING SMART METERING TRAFFIC? », *IEEE COMMUNICATIONS MAGAZINE*, vol. 53, p. 41-47, sept. 2015.
- [116] K. Ahuja et A. Khosla, « Network selection criterion for ubiquitous communication provisioning in smart cities for smart energy system », *JOURNAL OF NETWORK AND COMPUTER APPLICATIONS*, vol. 127, p. 82-91, févr. 2019, doi: 10.1016/j.jnca.2018.11.011.
- [117] T. Beck, H. Kondziella, G. Huard, et T. Bruckner, « Assessing the influence of the temporal resolution of electrical load and PV generation profiles on self-consumption and sizing of PV-battery systems », *APPLIED ENERGY*, vol. 173, p. 331-342, juill. 2016, doi: 10.1016/j.apenergy.2016.04.050.
- [118] S. Schopfer, V. Tiefenbeck, et T. Staake, « Economic assessment of photovoltaic battery systems based on household load profiles », *APPLIED ENERGY*, vol. 223, p. 229-248, août 2018, doi: 10.1016/j.apenergy.2018.03.185.
- [119] E. Barbour et M. González, « Projecting battery adoption in the prosumer era », *APPLIED ENERGY*, vol. 215, p. 356-370, avr. 2018, doi: 10.1016/j.apenergy.2018.01.056.
- [120] M. Neaimeh et al., « A probabilistic approach to combining smart meter and electric vehicle charging data to investigate distribution network impacts », *APPLIED ENERGY*, vol. 157, p. 688-698, nov. 2015, doi: 10.1016/j.apenergy.2015.01.144.
- [121] H. Çimen, N. Çetinkaya, J. Vasquez, et J. Guerrero, « A Microgrid Energy Management System Based on Non-Intrusive Load Monitoring via Multitask Learning », *IEEE TRANSACTIONS ON SMART GRID*, vol. 12, n° 2, p. 977-987, mars 2021, doi: 10.1109/TSG.2020.3027491.
- [122] J. Ponocko et J. Milanovic, « Forecasting Demand Flexibility of Aggregated Residential Load Using Smart Meter Data », *IEEE TRANSACTIONS ON POWER SYSTEMS*, vol. 33, n° 5, p. 5446-5455, sept. 2018, doi: 10.1109/TPWRS.2018.2799903.
- [123] C. Feng, Y. Wang, K. Zheng, et Q. Chen, « Smart Meter Data-Driven Customizing Price Design for Retailers », *IEEE TRANSACTIONS ON SMART GRID*, vol. 11, n° 3, p. 2043-2054, mai 2020, doi: 10.1109/TSG.2019.2946341.

- [124] R. Miller, L. Golab, et C. Rosenberg, « Modelling weather effects for impact analysis of residential time-of-use electricity pricing », *ENERGY POLICY*, vol. 105, p. 534-546, juin 2017, doi: 10.1016/j.enpol.2017.03.015.
- [125] M. Nicolson, G. M. Huebner, and D. Shipworth, « Are consumers willing to switch to smart time of use electricity tariffs? The importance of loss-aversion and electric vehicle ownership », *Energy Research & Social Science*, vol. 23, pp. 82–96, Jan. 2017, doi: 10.1016/j.erss.2016.12.001.
- [126] A. Pimm, T. Cockerill, et P. Taylor, « Time-of-use and time-of-export tariffs for home batteries: Effects on low voltage distribution networks », *JOURNAL OF ENERGY STORAGE*, vol. 18, p. 447-458, août 2018, doi: 10.1016/j.est.2018.06.008.
- [127] P. Simshauser et D. Downer, « On the Inequity of Flat-rate Electricity Tariffs », *ENERGY JOURNAL*, vol. 37, n° 3, p. 199-229, juill. 2016, doi: 10.5547/01956574.37.3.psim.
- [128] A. Pità, G. Verticale, C. Rottondi, A. Capone, et L. Lo Schiavo, « The Role of Smart Meters in Enabling Real-Time Energy Services for Households: The Italian Case », *ENERGIES*, vol. 10, n° 2, févr. 2017, doi: 10.3390/en10020199.
- [129] K. Hopf, M. Sodenkamp, et T. Staake, « Enhancing energy efficiency in the residential sector with smart meter data analytics », *ELECTRONIC MARKETS*, vol. 28, n° 4, p. 453-473, nov. 2018, doi: 10.1007/s12525-018-0290-9.
- [130] F. Luo, G. Ranzi, W. Kong, Z. Dong, S. Wang, et J. Zhao, « Non-intrusive energy saving appliance recommender system for smart grid residential users », *IET GENERATION TRANSMISSION & DISTRIBUTION*, vol. 11, n° 7, p. 1786-1793, mai 2017, doi: 10.1049/iet-gtd.2016.1615.
- [131] F. Luo, G. Ranzi, X. Wang, et Z. Dong, « Social Information Filtering-Based Electricity Retail Plan Recommender System for Smart Grid End Users », *IEEE TRANSACTIONS ON SMART GRID*, vol. 10, n° 1, p. 95-104, janv. 2019, doi: 10.1109/TSG.2017.2732346.
- [132] M. Cai et al., « Impact of HVAC Set Point Adjustment on Energy Savings and Peak Load Reductions in Buildings », présenté à 2018 IEEE INTERNATIONAL SMART CITIES CONFERENCE (ISC2), 2018.
- [133] W. Gans, A. Alberini, et A. Longo, « Smart meter devices and the effect of feedback on residential electricity consumption: Evidence from a natural experiment in Northern Ireland », *ENERGY ECONOMICS*, vol. 36, p. 729-743, mars 2013, doi: 10.1016/j.eneco.2012.11.022.
- [134] H. B. Gear, "Diversity factor," *Proceedings of the American Institute of Electrical Engineers*, vol. 29, no. 8, pp. 1365–1374, Aug. 1910, doi: 10.1109/paiee.1910.6660566.
- [135] C. Bary, "Coincidence-Factor relationships of Electric-Service-Load characteristics," *Transactions of the American Institute of Electrical Engineers*, vol. 64, no. 9, pp. 623–629, Sep. 1945, doi: 10.1109/t-aiee.1945.5059190.

- [136] A. Sargent, R. P. Broadwater, J. C. Thompson and J. Nazarko, 1994, "Estimation of diversity and kWhRto-peak-kW factors from load research data ", IEEE Transactions on Power Systems. vol. 9(3), 1450-1456.
- [137] J. Nazarko, R. P. Broadwater and N. I. Tawalbeh, 1998, "Identification of statistical properties of diversity and conversion factors from load research data" MELECON '98. 9th Mediterranean Electrotechnical Conference, vol.1, 217-220.
- [138] V. P. Chatlani, D. J. Tylavsky, D. C. Montgomery and M. Dyer, 2007, "Statistical Properties of Diversity Factors for Probabilistic Loading of Distribution Transformers" 39th North American Power Symposium, 555-561.
- [139] G. T. Heydt, "Distribution Transformer Loading: probabilistic modeling and diversity factor," IEEE Transactions on Power Delivery, vol. 38, no. 2, pp. 842–849, Apr. 2023, doi: 10.1109/tpwrd.2022.3199999.
- [140] J. Lee, M. Bariya, and D. Callaway, 2022. "Peak Load Estimation with the Generalized Extreme Value Distribution" Berkeley Education Technical Report
- [141] J. Dickert and P. Schegner, 2010, "Residential load models for network planning purposes," 2010 Modern Electric Power Systems, pp. 1-6.
- [142] S. Velandar, Energidistributionens ekonomiska lagar, Stockholm, 1947 (en suédois)
- [143] V. Neimane, "On Development Planning of Electricity Distribution Networks," Doctoral Dissertation, 2001
- [144] Guilherme Ramos Milis, Christophe Gay, Marie-Cécile Alvarez-Hérault, Bruno Gourguechon, Raphaël Caire. USE OF LINKY SMART METER DATA TO ENHANCE THE DIVERSITY FACTOR ASSESSMENT IN REAL NETWORKS. 27th International Conference on Electricity Distribution CIRED 2023, Jun 2023, ROME, Italy. (hal-04137757)
- [145] Réglementation Thermique 2012 : un saut énergétique pour les bâtiments, 2011 [Online]. Disponible : <https://www.ecologie.gouv.fr/sites/default/files/RT%202012%20-%20un%20saut%20C3%A9nerg%20pour%20les%20b%C3%A2timents%20neufs%20-%20Avril%202011.pdf>
- [146] M. R. Segal, 2004, "Machine learning benchmarks and random forest regression".
- [147] A. L. Boulesteix, S. Janitza, J. Kruppa and I. R. König, 2012, "Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics" Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discover, pp. 493-507.
- [148] S. M. Lundberg and S. I. Lee, 2017, "A unified approach to interpreting model predictions", Advances in neural information processing systems, 30.
- [149] "Smart Grids Innovation Challenge Country Report 2019", MI-IC1 Smart Grids. https://www.mi-ic1smartgrids.net/our-work/activities/country_report_2019/

- [150] "NF C14-100," Afnor EDITIONS. <https://www.boutique.afnor.org/fr-fr/norme/nf-c14100/installations-de-branchement-a-basse-tension/fa156644/897#AreasStoreProductsSummaryView>
- [151] "Reconstituer les flux de responsabilité d'équilibre - RTE Portail Services," *Portail Services RTE*. <https://www.services-rte.com/fr/decouvrez-nos-offres-de-services/le-role-des-gestionnaires-de-res/reconstituez-les-flux-re.html>
- [152] Enedis, « Règles de gestion mises en œuvre par Enedis pour le traitement des données dans le processus de Reconstitution des Flux ». Disponible : https://www.enedis.fr/sites/default/files/import/Enedis-NOI-CF_103E.pdf
- [153] 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.
- [154] A. Zahedi, "A review of drivers, benefits, and challenges in integrating renewable energy sources into electricity grid," *Renewable & Sustainable Energy Reviews*, vol. 15, no. 9, pp. 4775–4779, Dec. 2011, doi: 10.1016/j.rser.2011.07.074.
- [155] E. L. Codjo, B. B. Zad, J. Toubeau, F. Bart, and F. Vallée, "Machine Learning-Based Classification of electrical low voltage cable degradation," *Energies*, vol. 14, no. 10, p. 2852, May 2021, doi: 10.3390/en14102852.
- [156] A. Grandjean, J. Adnot, and G. Binet, "A review and an analysis of the residential electric load curve models", *Renewable and Sustainable Energy Reviews*, 2012.
- [157] R. Bartels, D. G. Fiebig, M. Garben, and R. Lumsdaine "An end-use electricity load simulation model", *Utilities Policy*, 2(1) :71 – 82, 1992.
- [158] L. G. Swan and V. I. Ugursa, "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.
- [159] R. Bro and A. K. Smilde, 2014, "Principal component analysis" *Analytical methods*, 2812-2831.
- [160] E. C. Levy, "Complex-curve fitting," in *IRE Transactions on Automatic Control*, vol. AC-4, no. 1, pp. 37-43, May 1959, doi: 10.1109/TAC.1959.6429401.
- [161] M. Newville, A. Ingargiola, T. Stensitzki, and D. B. Allen, "LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python," *Astrophysics Source Code Library*, Sep. 2014, doi: 10.5281/zenodo.11813
- [162] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol.16, no. 5, pp. 1190–1208, Sept. 1995
- [163] E. Julien, J. -P. Cosperec and P. Feron, "LV network planning tools at EDF," *12th International Conference on Electricity Distribution, 1993. CIREN*, Birmingham, UK, 1993, pp. 6.10/1-6.10/5 vol.6.

- [164] C. DE GOUELLO, « Le système d'information GDO-BT et la détection des contraintes sur le réseau BT rural. Limites de BATECA. Avancées de BAGHEERA ». Rapport pour la DIGEC. CIRED. Paris, (1996c).
- [165] Franck Nadaud. Hétérogénéité spatiale d'un service de réseau, équité et efficacité collective : la distribution rurale d'électricité et la maîtrise de la demande. Economies et finances. Ecole des Hautes Etudes en Sciences Sociales (EHESS), 2005. Français.
- [166] Ni Ding. Load models for operation and planning of electricity distribution networks with metering data. Engineering Sciences [physics]. Université de Grenoble, 2012. English
- [167] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," *16th IEEE International Conference on Tools with Artificial Intelligence*, Boca Raton, FL, USA, 2004, pp. 576-584, doi: 10.1109/ICTAI.2004.50.
- [168] Likas, A., Vlassis, N., & Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, 36(2), 451–461. [https://doi.org/10.1016/s0031-3203\(02\)00060-2](https://doi.org/10.1016/s0031-3203(02)00060-2)
- [169] Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior Ville Satopa † , Jeannie Albrecht† , David Irwin‡ , and Barath Raghavan§ †Williams College, Williamstown, MA ‡University of Massachusetts Amherst, Amherst, MA § International Computer Science Institute, Berkeley, CA
- [170] Scipy.Signal.find_Peaks — SciPy v1.11.3 Manual. (n.d.). https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html

6 Annexes

6.1 Annexes 1 : Relatives au chapitre 3

6.1.1 Métriques d'évaluation des erreurs utilisées dans des modèles de prévision

Le Tableau 6-1 donne une description des principales métriques utilisées pour l'évaluation des erreurs des modèles de prévision.

	MAE	MSE	RMSE	MAPE	SMAPE
D E S C R I P T I O N	La MAE est la moyenne des différences absolues entre les prédictions et les valeurs réelles	Le MSE est la moyenne des différences quadratiques entre les prédictions et les valeurs réelles	Le RMSE est la racine carrée du MSE, revenant à l'échelle originale des données	Le MAPE calcule la moyenne des pourcentages absolus d'erreur par rapport aux valeurs réelles	Le SMAPE est une métrique symétrique qui exprime l'erreur en termes de pourcentage, prenant en compte la magnitude des valeurs prédites et réelles.
F O R M U L E	$\frac{1}{N} \times \sum_{n=1}^N P_n - R_n $	$\frac{1}{N} \times \sum_{n=1}^N (P_n - R_n)^2$	\sqrt{MSE}	$\frac{1}{N} \times \sum_{n=1}^N \frac{ P_n - R_n }{ R_n }$	$\frac{1}{N} \times \sum_{n=1}^N \frac{ P_n - R_n }{(R_n + P_n)/2}$

Tableau 6-1 : Métriques d'évaluation des erreurs

6.1.2 Charte de chaleur avec tous les indicateurs

La Figure 6-1 montre la carte thermique pour tous les indicateurs décrits dans le Tableau 3-1. Même s'ils sont fortement corrélés, les 3 mesures de température ont été conservées car elles fournissent des informations complémentaires.

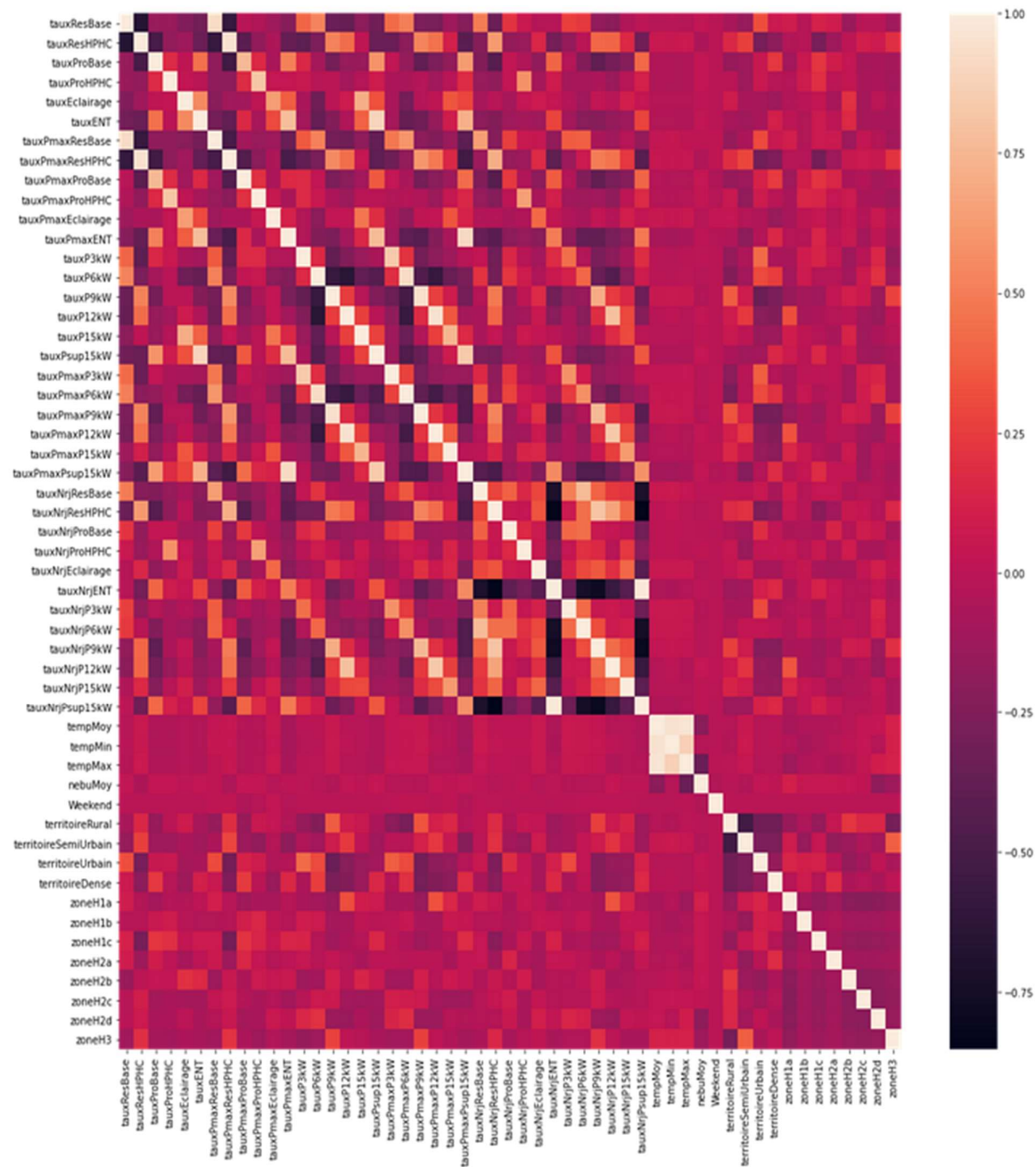


Figure 6-1 : Charte de chaleur

6.2 Annexes 2 : Relatives au chapitre 4

6.2.1 Les histogrammes des occurrences de puissances maximales et les points de maxima du Poste Exemple pour la saison d'hiver

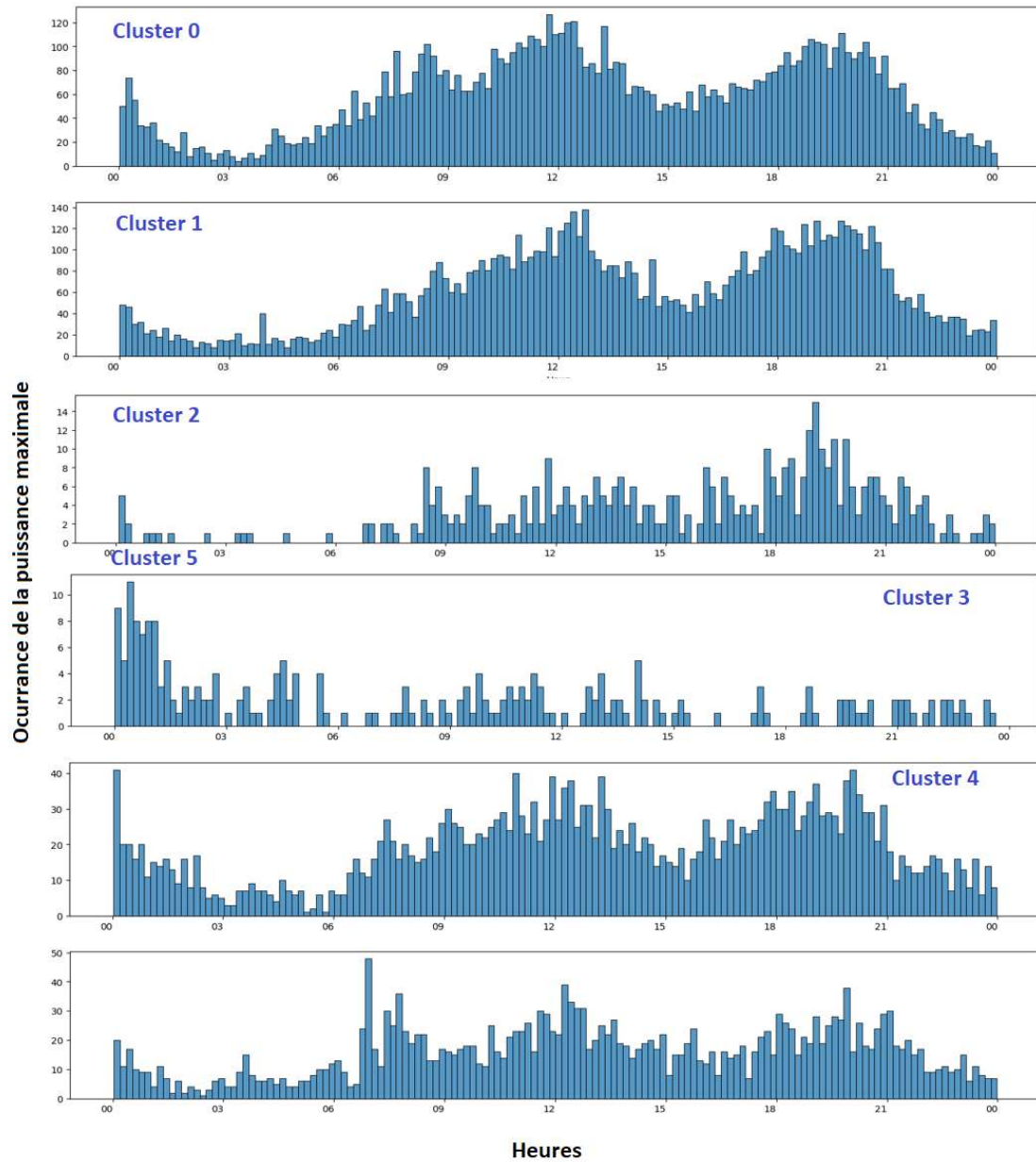


Figure 6-2 : Occurances de la puissance maximale pour le clients du Poste Exemple dans la saison d'hiver

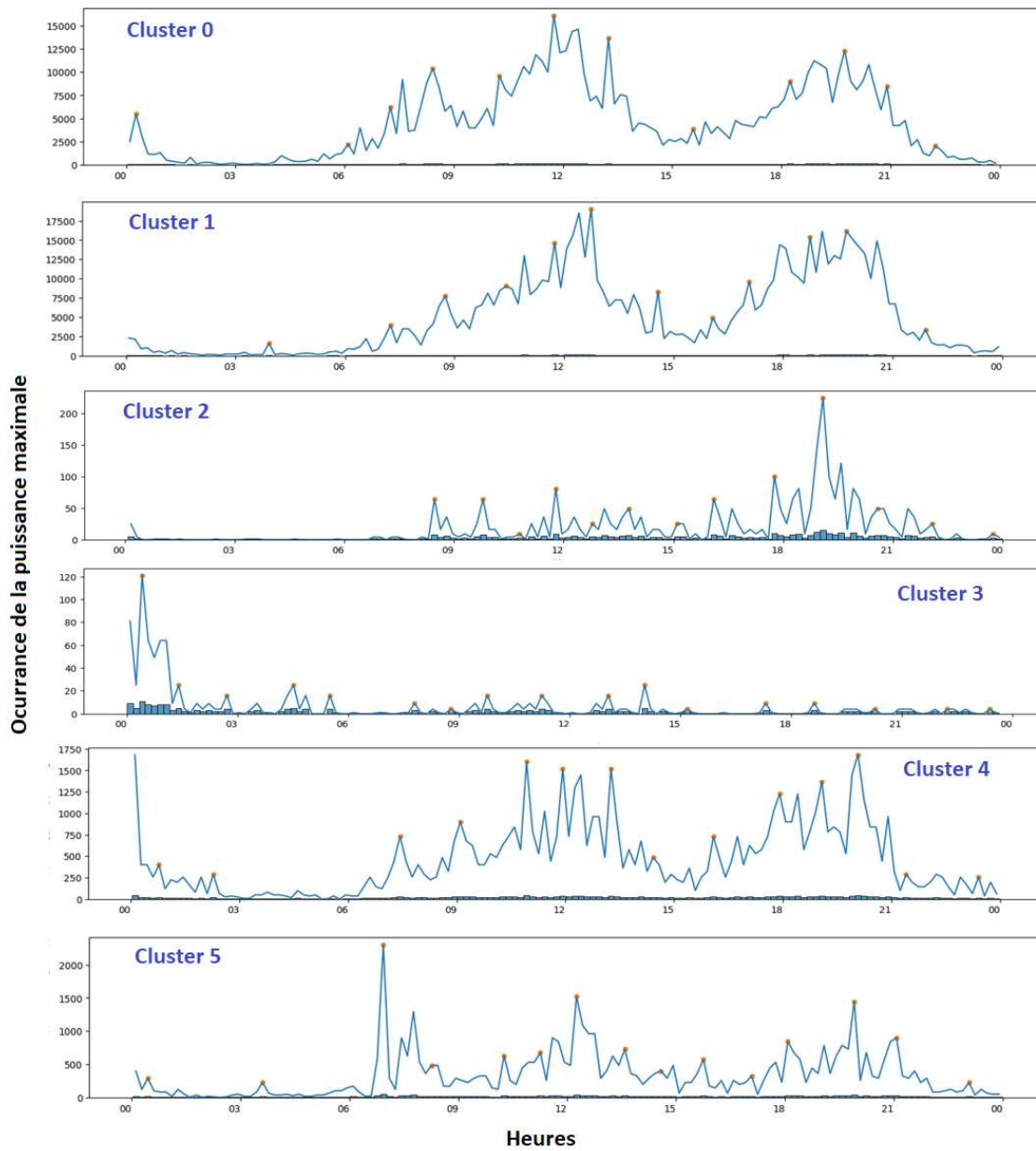


Figure 6-3 : Points de maxima des clusters du Poste Exemple dans la saison d'hiver

6.2.2 Les erreurs du modèle 4 avec différentes valeurs de Lambda

Le Tableau 6-2 montre les erreurs du modèle 4 avec différentes valeurs de lambda.

	Lambda = 1	Lambda = 0,1	Lambda = 0,01	Lambda = 0,001	Lambda = 0,0001	Lambda = 0,00001
SMAPE minimal (%)	17,52	14,45	9,72	7,41	2.06	5,03
SMAPE moyen (%)	36,17	32,88	26,13	19,19	4.86	8,75
SMAPE maximal (%)	63,64	58,22	49,63	37,30	11.30	29,40

Tableau 6-2 : Les erreurs du modèle 4 pour le Poste Exemple avec différentes valeurs de lambda