



HAL
open science

Contribution à la mise au point d'un pilotage énergétique décentralisé par prédiction

Luc Dufour

► **To cite this version:**

Luc Dufour. Contribution à la mise au point d'un pilotage énergétique décentralisé par prédiction. Génie civil. Ecole des Mines d'Albi-Carmaux, 2017. Français. NNT : 2017EMAC0004 . tel-01716841

HAL Id: tel-01716841

<https://theses.hal.science/tel-01716841>

Submitted on 24 Feb 2018

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

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



THÈSE

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par :

IMT - École nationale supérieure des Mines d'Albi-Carmaux

Présentée et soutenue par :

LUC DUFOUR

Le 20 Mars 2017

**Titre : Contribution à la mise au point d'un
pilote énergétique décentralisé par prédiction**

Ecole doctorale et discipline: MEGEP Energétique

Unités de recherches :

Centre RAPSODEE, CNRS UMR 5302, IMT - Mines Albi, France

Institut Informatique de Gestion, HES-SO Valais//Wallis, Sierre, Suisse

Directeurs de thèse :

Directeur de thèse : Bruno LADEVIE, CNRS UMR 5302, IMT - Mines Albi, France

Co-Directeur de thèse : Dominique GENOUD, Institut Informatique de Gestion, Sierre, Suisse

Jury :

Nadia MAIZI, Professeur et directrice du centre de mathématiques,
IMT - Mines Paris, Présidente du Jury

Michel BONVIN, Professeur, HES-SO Valais, Rapporteur

Jean HENNEBERT, Professeur, HES-SO Fribourg, Rapporteur

Frédéric BENABEN, Professeur, IMT - Mines Albi, Examineur

Jean-Jacques BEZIAN, Maître de recherche HDR, IMT - Mines Albi, Invité

Dominique GENOUD, Professeur, HES-SO Valais, co-Directeur

Bruno LADEVIE, Maître Assistant HDR, IMT - Mines Albi, Directeur

Summary

The world today is facing unprecedented environmental challenges and the global warming requires the states to react. The energy model which we inherit is today to be rethought. It is based on a variable power demand, constantly increasing and the supply of which depends on a point of production distant from the source of consumption. From an electrical point of view, the insertion of variable production from wind and photovoltaic increases pressure on the network.

This work presents a data-intensive solution to manage energy flux after a low transformer voltage named microgrid concept. A microgrid is an aggregation of building with a decentralized energy production and or not a storage system. These microgrids can be aggregate to create an intelligent virtual power plant.

However, many problems must be resolved to increase the part of these microgrid and the renewable resource in an energy mix. The physic model can not integrate and resolve in a short time the quickly variations. The intelligent district can be integrating a part of flexibility in their production with a storage system. This storage can be electrical with a battery or thermal with the heating and the hot water.

For a virtual power plant, the system can be autonomous when the price electricity prediction is low and increase the production provided on the market when the price electricity is high. For an energy supplier and with a decentralized production building distant of a low transformer voltage, regulations with a storage capacity enable a tension regulation. Finally, the auto-consumption becomes more and more interesting combined with a low electrical storage price and the result of the COP 21 in Paris engages the different country toward the energy transition. In these cases, a flexibility is crucial at the building level but this flexibility is possible if, and only if, the locally prediction are correct to manage the energy.

The main novelties of our approach is to provide an easy implemented and flexible solution to predict the consumption and the production at the building level based on the machine learning technique and tested on the real use cases in a residential and tertiary sector. A new evaluation of the consumption is realized: the point of view is energy and not only electrical. The energy consumption is decomposed between the heating consumption, the hot water consumption and the electrical devices consumption. A prediction every hour is provided for the heating and the hot water consumption to estimate the thermal storage capacity. A characterization of Electrical devices consumption is realized by a non-intrusive disaggregation from the global load curve. The heating and the hot water are identifying to provide a non-intrusive methodology of prediction. Every day, the heating, the hot water, the household appliances, the cooling and the standby are identified.

These predictions allow a shift of energy demand according the different predictions at $t + n$, n between 15 minutes and one day, and the system users (the final consumers and the aggregator). It is therefore to set up an information system non-intrusive, intelligent, robust and simple. To limit the intrusiveness of the systems, they are both adding intelligence as identify useful devices from a single measurement point but also collect and analyze information locally. The private data is not transmitted to the outside. Only the information's of energy predictions are sent at a higher level for the aggregation of the district data. The intelligence is also in learning method predictive: Support Vector Machine, Decision Tree Ensemble and Hidden Markov Model. Robustness is studied from a technology point of view (more communication protocol have been tested), technical (Several methods for collecting) and a data storage point of view (the limit frequency of collection). The ease of use results in a simple installation. Limit the frequency of collection and minimize the number of prediction model input data are essential points. One goal is to find generalization points to limit the number of sensors: identical load curve useful specific uses in a particular sector of activity, presence detection from the global burden of curve electricity. Evaluation of the results and relevance of the predictions made on the basis of reference models such as detailed thermal models for heating or SIA standards. These reference models need a priori information's about the different buildings: activity sector, construction year, surface, people number. The Information systems are implemented on real cases and are present at different scales: a house, a tertiary industrial district, an industrial site.

Résumé

Réchauffement climatique, accidents nucléaires, généralisation de l'électrification.... Le monde de l'énergie est face à des défis importants qui nous ramènent au fonctionnement global de notre société et au rapport des citoyens avec celle-ci : comment satisfaire les besoins en énergie d'une population de 9 milliards d'êtres humains en 2050, de façon économiquement viable et en minimisant l'impact sur l'environnement. En Europe, les États membres de l'Union européenne ont convenu d'un nouveau cadre pour l'énergie et le climat à l'horizon 2030, comprenant des objectifs stratégiques à l'échelle européenne pour la période 2020-2030. Ces objectifs visent à aider l'Union à mettre en place un système énergétique plus compétitif, plus sûr et plus durable et à réaliser son objectif à long terme de réduction des gaz à effet de serre fixé pour 2050. Les objectifs clés à l'horizon 2030 sont les suivants:

- ⇒ réduire d'au moins 40 % les émissions de gaz à effet de serre (par rapport aux émissions de 1990)
- ⇒ atteindre une part d'au moins 27 % d'énergies renouvelables
- ⇒ améliorer d'au moins 27 % l'efficacité énergétique

Pour répondre à nos engagements, la transition énergétique passera par une maîtrise de l'insertion de production d'énergie d'origine éolienne et photovoltaïque. D'autant plus que la majorité des installations solaires installées en France et en Suisse sont de faibles puissances, raccordées au réseau basse tension, lieu où l'information manque. Les modèles prédictifs historiques centralisés et paramétriques ont du mal à appréhender les variations brutales de productions et de consommations et un maillage plus fin devient utile pour comprendre les scénarios d'occupation du bâtiment.

La révolution Internet permet aujourd'hui une convergence entre le numérique et l'énergie. En Europe et depuis cinq ans, l'axe d'étude est celui de la maîtrise locale de l'électricité. Ainsi plusieurs quartiers intelligents ont été créés et les modèles utilisés de pilotage et de prédiction restent souvent la propriété des partenaires des projets. Les recherches s'attèlent à identifier la présence humaine de manière intrusive pour diminuer l'erreur de prédiction de consommation de chauffage et à interagir avec l'humain sous forme de jeu par exemple pour l'inciter à diminuer sa consommation. Un pilotage par l'eau chaude et à partir de batteries électriques permet un effacement partiel ou total d'une partie de ces quartiers connectés.

Notre objectif dans cette thèse est de **favoriser la maîtrise de la demande locale en énergie**. Cela passera par la mise en place d'un **bilan énergétique prédictif** pour l'heure suivante qui permettra d'avoir des possibilités de **scénarios de surproductions ou surconsommations** sur un quartier donné à **court terme au niveau de chaque flux énergétiques (t + 1 heure)**. Cela implique une prédiction fine des différents flux énergétiques d'un système en particulier le chauffage et l'eau chaude qui représentent le plus gros potentiel de flexibilité dans les bâtiments **[ANG-2004] et permettra ainsi de limiter le besoin en stockage électrique**.

Le besoin en énergie d'un système comme une maison est ainsi décomposé en un besoin en **chauffage**, en un besoin en **eau chaude sanitaire**, en **luminaires**, en besoin de **ventilation** et en **usages spécifiques électriques** utiles. Le système peut posséder une **production décentralisée** et un système de stockage ce qui augmentera sa capacité d'effacement.

Pour réaliser un bilan, nous devons calculer les différents flux énergétiques à l'intérieur de notre système : les déperditions par l'enveloppe et la ventilation, les gains internes solaires, des personnes et des appareils, le stockage, la production d'eau chaude sanitaire (ECS) et les usages spécifiques électriques utiles (USE). Sur certains de ces points, nous pouvons évaluer assez précisément et en fonction du temps les quantités d'énergie échangées. Pour les autres (ECS, USE, gains internes, stockage), la bibliographie nous donne que des méthodes globales et indépendantes du temps. Il n'est donc pas possible d'envisager une méthode correspondant au pas de temps souhaité.

Ceci impose la mise au point d'une méthode prédictive et apprenante dont nos modèles de simulation énergétique seront le point de référence. Il n'en reste pas moins que ces modèles permettent la compréhension du comportement énergétique du système.

L'outil devrait être non intrusif, personnalisé, robuste et simple. Pour limiter le caractère intrusif de l'outil, il s'agit à la fois d'ajouter de l'intelligence, comme par exemple l'identification des appareils utiles à partir d'un seul point de mesure mais aussi la collection et l'analyse d'informations locales. Les données privées ne sont pas transmises vers l'extérieur. Seules les informations de prédictions énergétiques sont envoyées à un niveau supérieur pour agrégation des données des quartiers. Nous testerons également différentes méthodes de prédictions : les méthodes paramétriques comme les régressions linéaires ou les méthodes non paramétriques par apprentissage comme l'utilisation des réseaux de neurones ou des arbres de décision.

La robustesse est étudiée d'un point de vue technologie (plusieurs protocoles de communication ont été testés), technique (plusieurs méthodes de collecte) et d'un point de vue du stockage de données (limiter la fréquence de collecte). Notre objectif est de limiter au maximum le nombre de capteurs et le nombre de données en transit tout en gardant un niveau de précision maximale par flux énergétique. Les études intrusives pour la consommation et pour la production atteignent des niveaux de précision de 90% en puissance. Cependant les résultats sont très différents selon les études et ne traitent que d'un flux à la fois.

Notre outil a été testé sur des cas réels dans le secteur résidentiel, tertiaire et industriel. Sur le secteur résidentiel, l'ensemble des flux énergétiques (chauffage, eau chaude, électrique, production décentralisée renouvelable) ont été caractérisés et prédits. Le taux d'exactitude des précisions des classifications atteint 98% +/-1.2% dans le cadre de la classification.

Ensuite, sur le secteur tertiaire, une prédiction électrique est réalisée chaque quart d'heure par secteur d'activité (bureaux, restauration, studio de télévision). Les résultats de classification peuvent atteindre 98.4% +/-0.6% dans le secteur d'activité bureaux avec un an de données d'entraînement. Enfin, nos études ont porté sur le secteur industriel (pharmaceutique) sur la prédiction des HVAC (gestion de la température, humidité et pression) à partir de la température extérieure et du planning de production avec des résultats de 98.5% +/- 1.6%.

A mon frère,

A mes parents,

À mes amis qui se reconnaîtront,

Remerciements

Je tenais à remercier tout particulièrement mon co-directeur de thèse de la HES-SO Valais, le professeur **Dominique Genoud**. Sa vision dans le domaine de l'énergie et la confiance qu'il m'a apportée sont les pierres angulaires de cette thèse. Sans lui, rien n'aurait été possible.

En me basant sur la réputation des élèves de l'école des mines d'Albi, j'ai choisi il y a plus de trois ans de présenter mon idée de doctorat à mon directeur de thèse, le professeur **Bruno Ladevie**. J'ai pu ainsi comprendre l'aura qu'il a auprès des élèves de l'école des mines d'Albi : cette personne m'a formé dans le domaine des bâtiments à énergie positive et son seul but est d'élever le plus haut possible ses élèves : un grand merci à lui.

Je tenais à remercier le professeur **Jean-Jacques Bézian** avec qui nous avons travaillé sur l'état de l'art technique et technologique du solaire dans le monde.

Je voudrais remercier particulièrement mes deux rapporteurs, **M. Michel Bonvin et M. Jean Hennebert** qui ont su me conseiller pour la rédaction de ce rapport et leurs questions pertinentes ont permis d'améliorer ce document.

Au fur et à mesure de mon arrivée au sein de la HES-SO Valais, l'équipe spécialiste de l'analyse de données s'est agrandie. Aujourd'hui, plus qu'une équipe, ils sont devenus mes amis. Leurs soutiens ont été indispensables. Merci à eux : **Milène Fauquex, Jérôme Treboux et Mariam Barque**.

Je veux également remercier mon ami **Ludovic Roussin** avec qui j'ai passé un nombre incroyable de moments et de soirées à parler d'énergie. Au-delà de la compréhension du réel besoin des fournisseurs et producteurs d'énergie, il m'a également transmis sa passion pour le domaine de l'énergie.

Je remercie également le responsable du pôle software engineering, le professeur **Yann Bocchi**. Son bureau a été toujours ouvert pour m'écouter et il a fallu parfois jouer d'astuces pour réussir cette thèse d'un point de vue administratif.

Pour continuer sur le côté administratif qui n'est pas mon fort, il me paraît indispensable de remercier l'équipe du secrétariat qui m'a aidé dans l'une des plus grandes problématiques de ma vie : l'administratif. Ainsi, je remercie la responsable de l'équipe, **Mme Anne-Christine Rey** et les deux secrétaires **Mme Dragana Zelenovic et Mme Darlène Gillioz**.

Enfin, je tiens à remercier particulièrement mes quatre stagiaires avec qui j'ai travaillé sur ce sujet. Sans eux et face à l'immensité de la base de données, cette thèse n'aurait pas pu être réalisée : **Calixte Mayoraz, Morgane Giorgi, Yohann Verneuil et Quentin Jeannes**.

Nomenclature

GES : Gaz à Effet de Serre

E : Besoin en énergie

i : Pas de temps de la prédiction

E t(i) : Besoin en énergie à un instant t(i)

j : Jour de l'année

k : Jour de la semaine

s : Système étudié

$E_{\text{Consommation_électrique_totale}}$: Consommation électrique globale en sortie du compteur d'électricité

$E_{\text{Chauffage}}$: Besoin en chauffage

$E_{\text{Eau chaude}}$: Besoin en eau chaude sanitaire

E_{USE} : Besoin en énergie lié aux usages spécifiques électriques

$E_{\text{Ventilation}}$: Besoin en énergie lié à la ventilation

$E_{\text{Luminaires}}$: Besoin en énergie lié aux luminaires

E_{ENRE} : Energie renouvelable électrique

E_{ENRT} : Energie renouvelable thermique

$E_{\text{stocké_batterie}}$: Energie stockée dans la batterie

$P_{\text{Chauffage}}$: Puissance électrique liée à la consommation de chauffage

$P_{\text{Eau chaude}}$: Puissance électrique liée à la consommation d'eau chaude

P_{USE} : Puissance électrique liée à la consommation liée aux usages spécifiques électriques

$P_{\text{Ventilation}}$: Puissance électrique liée à la consommation liée à la ventilation

$P_{\text{Luminaires}}$: Puissance électrique liée à la consommation liée aux luminaires

P_{ENRE} : Puissance électrique liée à la production d'énergie renouvelable électrique

$E_{\text{stocké_batterie}}$: Energie stockée dans la batterie

$E_{\text{stocké}}$: Energie thermique stockée dans la structure du bâtiment

$D_{\text{Ventilation}}$: Déperditions par ventilation

$D_{\text{Transmission}}$: Déperditions par transmission

$D_{\text{Eau chaude}}$: Déperditions par le système de production et de distribution d'eau chaude sanitaire

G_{Solaires} : Gain interne solaire

$G_{\text{Appareils}}$: Gain interne apporté par le fonctionnement des appareils

$G_{\text{Personnes}}$: Gain interne apporté par le métabolisme humain

S : Surface de référence énergétique

T_{Int} : Température intérieure au niveau de la surface de référence

T_{Ext} : Température extérieure au niveau de la surface de référence

H_V : Coefficient de déperditions par renouvellement d'air

H_T : Coefficient de déperditions par transmission à travers l'enveloppe

H_D : transmission directe vers l'extérieur, à travers l'enveloppe du bâtiment

H_S : Coefficient de déperditions par le sol

H_N : Coefficient de déperditions à travers les espaces non chauffés.

m_a : Masse d'air chauffé

c_a : Chaleur spécifique de l'air, à savoir environ $1000 \text{ J}/(\text{kg}\cdot\text{K})$

η_r : Rendement de récupération de chaleur sur l'air évacué

N : Nombre de personnes présentes au niveau la surface de référence

Publications

Journaux internationaux

2015

*Victoria Moreno, Luc Dufour, Antonio F Gómez Skarmeta, Antonio J Jara, Dominique Genoud, Bruno Ladevie and Jean-Jacques Bezian, **Big data: the key to energy efficiency in smart buildings**, Soft computing manuscript, 2015*

2014

*Antonio J Jara, Luc Dufour, Gianluca Rizzo, Dominique Genoud, Alexandre Cotting, Yann Bocchi and François Chabbey, **I-BAT : A Data-intensive Solution based on the Internet of Things to Predict Energy Behaviours in Microgrids**, International Journal of Data Warehousing and Mining (IJDWM), 2014*

Communications avec comité de relecture

2016

*Luc Dufour, Dominique Genoud, Stephane Genoud, Francesco Maria Cimmino, Bruno Ladevie and Jean-Jacques Bezian, **Economic interest of heating and hot water prediction system for residential district**, Advanced Information Networking and Applications (AINA-2016), Crans-Montana, Switzerland, March 23-25, 2016*

Best Paper Award

*Luc Dufour, Dominique Genoud, Bruno Ladevie and Jean-Jacques Bezian, **Heating and hot water industrial prediction system for residential district**, Advanced Information Networking and Applications (AINA-2016), Crans-Montana, Switzerland, March 23-25, 2016*

2015

*Luc Dufour, Dominique Genoud, Antonio J Jara, Jerome Treboux, Bruno Ladevie and Jean-Jacques Bezian, **A non-intrusive model to predict the flexible energy in a residential building**, IEEE WCNC : Wireless Communications and Networking Conference (WCNC-2015), New Orleans, USA, 2015*

*Victoria Moreno, Luc Dufour, Dominique Genoud, Bruno Ladevie and Jean-Jacques Bezian, **Exploiting IoT-based Sensed Data in Smart Buildings to Model its Energy Consumption**, International Conference on Communications (ICC-2015), London, UK, 2015.*

Best Paper Award

*Mariam Barque, Luc Dufour, Dominique Genoud, Arnaud Zufferey, Bruno Ladevie and Jean-Jacques Bezian, **Solar production prediction based on non linear meteo source adaptation**, Innovative*

Mobile and Internet Services in Ubiquitous Computing (IMIS-2015), Regional University of Blumenau, Bresil, 2015

2014

*Luc Dufour, Dominique Genoud, Gianluca Rizzo, Bruno Ladevie and Jean-Jacques Bezian, **Identification method for home electrical signal disaggregation**, International Conference on Sustainable Energy Technologies (SET 2014), 25-28.08.2014, Geneva, Switzerland*

*Luc Dufour, Dominique Genoud, Gianluca Rizzo, Bruno Ladevie, Jean-Jacques Bezian and Pierre Roduit, **Verification method for home electrical signal disaggregation**, Extending Seamlessly to the Internet of Things (ESIOT-2014), Birmingham, UK, 2014*

Introduction générale

Le monde actuel est face à des défis écologiques sans précédent et le réchauffement climatique oblige les états à limiter leurs dépendances à l'énergie fossile. Le modèle énergétique dont nous héritons est aujourd'hui à repenser. Il est basé sur une demande en énergie variable, en constante augmentation, dont la fourniture dépend d'un point de production éloigné de la source de consommation. Dans cette étude, Il s'agit de créer un outil pour prédire le quart d'heure et l'heure suivante l'ensemble des usages énergétiques d'un système. Le besoin en énergie d'un système comme une maison est décomposé en un besoin en chauffage, en un besoin en eau chaude sanitaire, en un besoin en luminaires, en besoin de ventilation et en usages spécifiques électriques utiles. Le système peut posséder une production décentralisée et un système de stockage ce qui augmentera sa capacité d'effacement.

Dans le chapitre 1 sont décrits les modèles de référence utilisés pour simuler les différents usages énergétiques d'un système. Dans le chapitre 2, nous décrivons les méthodes de prédictions utilisées pour l'estimation de consommation et de production d'énergie à court terme. Dans le chapitre 3 est présenté le système d'information qui permet la collecte, le traitement, le stockage et le téléchargement des données. Le chapitre 4 décrit le traitement réalisé à l'intérieur de notre outil pour fournir les différentes variables d'entrée par usages énergétiques avant la modélisation. Enfin le chapitre 5 décrit les résultats obtenus en sortie des algorithmes décrits dans le chapitre 2.

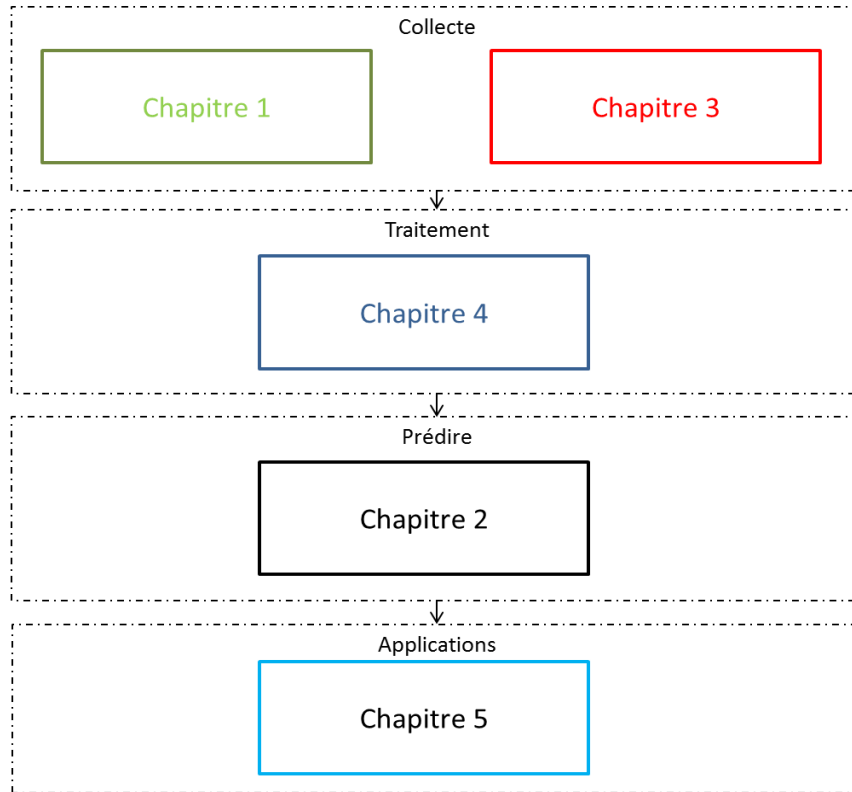


Figure 1: Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision

Les chapitres 1 et 3 décrivent la collecte des données dans cette étude. Le chapitre 4 représente le traitement de l'information. Enfin le chapitre 2 permet de fournir les différentes prédictions. Les chapitres 1 et 2 sont des revues bibliographiques et les chapitre 3 et 4 sont les outils utilisés pour fournir les résultats des prédictions dans le chapitre 5.

Sommaire

Résumé	I
Remerciements	VI
Nomenclature	VII
Publications	IX
Introduction générale	11
Préambule	16
Contexte	16
Une triple dépendance à l'énergie fossile.....	17
Limiter son usage dans la consommation énergétique globale.....	17
Limiter son usage dans la production d'électricité	21
Problématique.....	25
Limiter l'usage des énergies fossiles dans la stabilisation du réseau.....	25
Le big data au service de la transition énergétique	27
Limites des études précédentes.....	29
Objectifs de la thèse	31
Chapitre 1 : Caractérisation des usages énergétiques	34
1.1 Introduction.....	34
1.2 Définition du système	34
1.3 Notations.....	36
1.4 Le besoin en énergie de chauffage/climatisation	37
1.4.1 Déperditions.....	37
1.4.2 Etude des apports internes	38
1.4.3 Modèles simplifiés.....	41
1.5 Le besoin en eau chaude sanitaire	43
1.6 Les usages spécifiques électriques	44
1.7 Production local renouvelable	46
1.8 Stockage local de l'électricité	48
1.9 Conclusion	49

2	Chapitre 2 : Méthodes de prédiction des flux énergétiques	50
2.1	Introduction.....	50
2.2	Apprendre du passé pour prédire l'avenir	51
2.3	Méthodes linéaires.....	52
2.4	Méthodes non linéaires.....	52
2.4.1	Réseau de neurones	52
2.4.2	Les arbres de décision	55
2.4.3	Extensions liées au boosting et bagging.....	56
2.5	Calcul de l'erreur	57
2.5.1	Courbe ROC dans le cas de problème binaire	57
2.5.2	Méthodes statistiques.....	58
2.6	Conclusion	59
	Chapitre 3 : Mettre en place le système d'information	60
3.1	Introduction.....	60
3.2	Evaluer.....	62
3.2.1	Définition des acteurs.....	62
3.2.2	Définition des fonctionnalités	67
3.3	Collecter les données	69
3.3.1	Collecter les données au niveau des compteurs.....	69
3.3.2	Collecter les données météorologiques.....	72
3.3.3	Identifier la présence humaine	74
3.4	Stocker les données.....	76
3.5	Traiter et distribuer l'information	77
3.6	Conclusion	80
	Chapitre 4 : Traitement des données.....	81
4.1	Introduction.....	81
4.2	Extraction des caractéristiques depuis le compteur global	81
4.2.1	Préparer et extraire les statistiques de puissance	81
4.2.2	Identification du chauffage et de l'eau chaude.....	86
4.3	Créer son propre modèle de prédictions météorologiques.....	89

4.4	Classification.....	93
4.4.1	Apprentissage supervisé.....	93
4.4.2	Apprentissage non supervisé	94
4.5	Réduction du nombre de variables	98
4.6	Conclusion	99
Chapitre 5 : Présentation des résultats.....		101
5.1	Introduction.....	101
5.2	Prédire les besoins en énergie dans le secteur résidentiel	102
5.2.1	Présentation du cas d'étude.....	102
5.2.2	Prédire la consommation électrique globale	103
5.2.3	Prédire le besoin en chauffage.....	108
5.2.4	Prédire le besoin en eau chaude sanitaire	115
5.3	Prédire le besoin en énergie dans le secteur tertiaire	122
5.3.1	Présentation du cas d'étude.....	122
5.3.2	Prédire la consommation électrique globale	127
5.3.2.1	Secteur restauration.....	127
5.3.2.2	Secteur bureaux	130
5.3.3	Prédire la production solaire	134
5.4	Prédire les besoins en énergie dans le secteur industriel	138
5.4.1	Présentation du cas d'étude.....	138
5.4.2	Prédire le besoin en chauffage, ventilation, humidité et pression	140
Conclusion générale		143
Bibliographie.....		147
Table des illustrations.....		152
Table des tableaux		156
ANNEXES.....		Error! Bookmark not defined.

Préambule

Contexte

Hormis la crise qui débutât en 2008, nous notons une augmentation constante de la consommation d'énergie et des émissions de CO₂ dans le monde. Elle est en effet dépendante du PIB (Produit intérieur Brut) mondial et du taux d'accroissement naturel, tous deux en constante augmentation. Cela engendre une augmentation constante de la température sur terre et des phénomènes naturels dévastateurs (tornades, inondations, tsunami) qui ont poussé les états à réagir lors de la COP21 à Paris. Les Etats-Unis et la Chine, plus gros pollueurs mondiaux (Figure 2), ont signé un accord contraignant de réduction des gaz à effet de serre. La combustion d'énergies fossiles est la principale source des émissions de GES (gaz à effet de serre) dans le monde. En 2010, 22,7% des émissions mondiales de GES provenaient de la Chine, 15,6% des États-Unis et 10,9% de l'Union européenne d'après l'agence internationale de l'énergie. Rappelons que la transition énergétique est le remplacement des énergies fossiles (pétrole, charbon et gaz) par des énergies nouvelles renouvelables (éolien, solaire, marine, bois, déchet). Dans le domaine de la production électrique, les centrales à charbon et au pétrole sont de loin les plus émettrices.

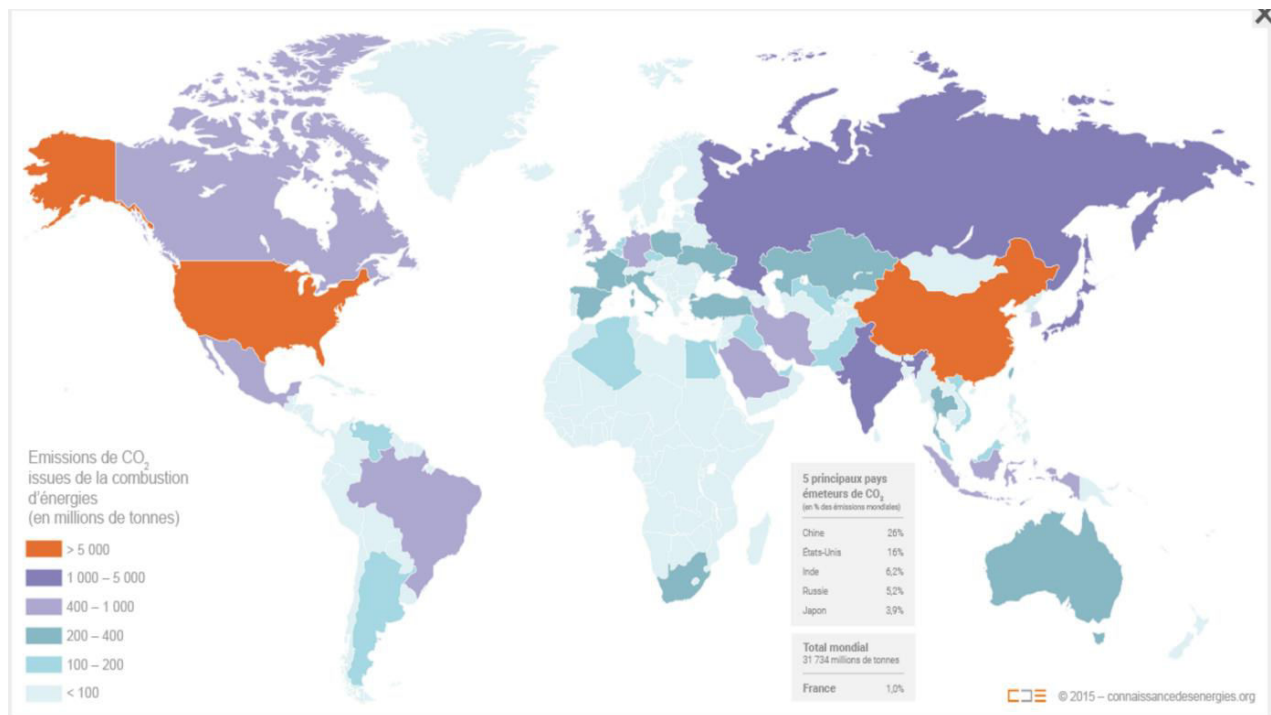


Figure 2: Emissions de CO₂ issues de la combustion d'énergie (en millions de tonnes)

Nous héritons d'un modèle énergétique centralisé qui passe par des étapes d'extraction, de prétraitement d'acheminement pour que cette énergie puisse être utilisée par les consommateurs finaux. L'énergie, nommée primaire, doit être extraite d'un point A (mines d'uranium, puits de pétrole,

puits de gaz,...). Elle est ensuite prétraitée sur place pour être acheminée à un point B par pipeline ou bateau. Elle est ensuite transformée sur le site du point B pour être acheminée au lieu de consommation. De nombreux conflits ont été et sont liés de près ou de loin à cet accès à l'énergie au point A. C'est pourquoi au-delà de la question environnementale, l'énergie fournie localement, par exemple par des énergies renouvelables, permet d'augmenter l'indépendance énergétique d'un pays (d'une région, d'un canton, d'une ville, d'un quartier).

Ainsi, le modèle énergétique dont nous héritons est aujourd'hui à repenser en particulier au niveau du réseau électrique. Selon l'agence internationale de l'énergie, 7% de l'énergie produite au sein des pays de l'OCDE (Organisation de Coopération et de Développement Économiques) disparaît en moyenne lors de sa distribution. Une autre limite du système centralisé est qu'il s'adapte difficilement aux fluctuations de la demande. Enfin, l'acceptation publique est de plus en plus faible pour la mise en place de grosses infrastructures de production d'énergie. Les réserves du public ne se limitent pas à l'énergie nucléaire mais aussi aux centrales thermiques à combustion fossile ou à la création de lignes à très haute tension.

De plus, nous pouvons citer la catastrophe de Fukushima qui a engendré un changement de politique énergétique dans certains pays notamment européens comme en Allemagne et en Suisse où un arrêt total du nucléaire est programmé. En Suisse, le Conseil Fédéral a annoncé le 25 mai 2011 une sortie progressive de l'énergie nucléaire programmée pour 2034.

Pour éviter que les contraintes réseaux ne deviennent un obstacle à la pénétration des énergies renouvelables, certains états font évoluer le cadre réglementaire de manière favorable aux solutions de stockage. Ces nouvelles législations offrent un cadre nécessaire au développement de la filière. Encore en gestation à ce jour, ces cadres réglementaires ont deux conséquences :

- ⇒ L'arrivée d'une nouvelle contrainte pour les producteurs d'énergies intermittentes,
- ⇒ La naissance d'une nouvelle opportunité

On peut citer la loi NOME qui prévoit la mise en place d'une obligation de capacité pour les producteurs d'électricité, leur imposant de justifier d'une capacité physique de production égale à la puissance nécessaire à la fourniture de leurs clients **[CRE-2015]**. Cette obligation devrait également s'assortir de la naissance d'un marché de capacité, ceci à l'horizon 2017. Une nouvelle valeur sera ainsi attribuée à la puissance rendue disponible par le stockage d'énergie, en procurant à ses détenteurs des certificats de capacité à valoriser sur un marché.

Une triple dépendance à l'énergie fossile

Limiter son usage dans la consommation énergétique globale

La répartition de la consommation d'énergie primaire de la France est stabilisée depuis le milieu des années 2000 **[MIN-2015]**, avec 42,6 % d'électricité primaire non renouvelable, environ 30 % de pétrole, 15 % de gaz, le reste se répartissant entre les énergies renouvelables et le charbon (Figures 3 et 4). Par rapport aux deux années précédentes, l'année 2015 se distingue par d'une part la nette retombée

du charbon, qui a atteint un plancher record, avec 3 Mtep (tonne équivalent pétrole) de moins qu'en 2013, et passe sous la barre des 4 % du bouquet énergétique primaire ; d'autre part, le renforcement de la part de l'électricité (y compris renouvelable), qui gagne un point par rapport à 2013, et a fourni 45,4 % de l'énergie consommée en France en 2014.

De plus, l'énergie primaire fournie par les énergies renouvelables thermiques et la valorisation des déchets consolide sa part, à 7 % du bouquet énergétique. En revanche, la part du gaz s'effrite continûment depuis 2010 tandis que la consommation primaire de pétrole, après dix ans de baisse presque ininterrompue, se redresse d'un Mtep par rapport à 2013, en données corrigées des variations climatiques.

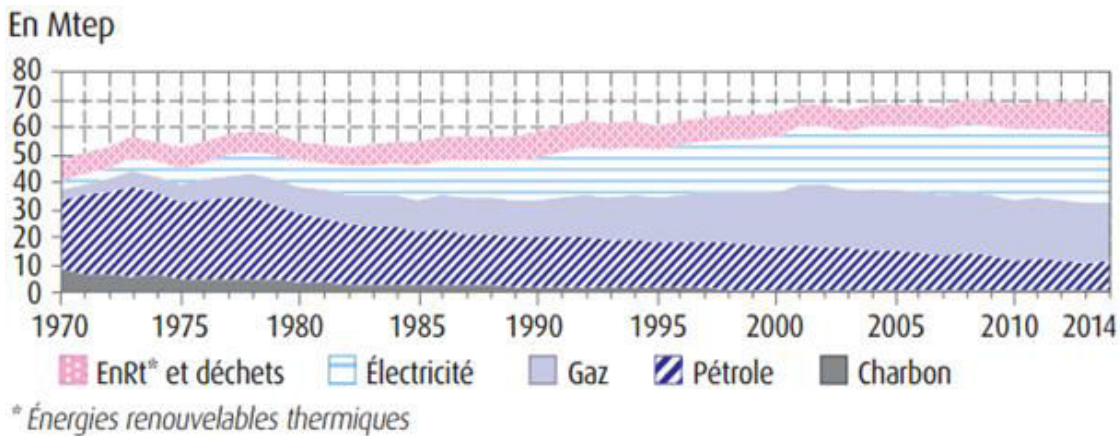
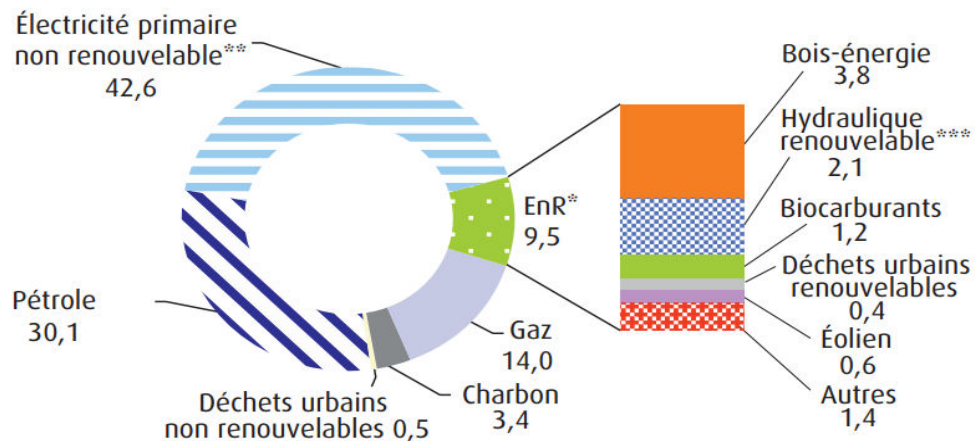


Figure 3 : Evolution de la consommation des différentes sources d'énergie en France [MIN-2015]

Données corrigées des variations climatiques (256,6 Mtep en 2014)
En %



* EnR : énergies renouvelables (électriques et thermiques, voir « Méthodologie - définitions » p. 47).
** Comprend la production nucléaire, déduction faite du solde exportateur d'électricité (pour simplifier, le solde exportateur est retranché de l'électricité nucléaire) et la production hydraulique par pompage.
*** Hydraulique hors pompage.

Figure 4 : Répartition de la consommation des différentes sources d'énergie en France, [MIN-2015]

A titre d'illustration, en Suisse et depuis la fin de la seconde guerre mondiale, la figure 5 décrit une explosion de la consommation énergétique en particulier au niveau de l'électricité et des carburants [OFF-2015].

Consommation finale d'énergie par agent énergétique

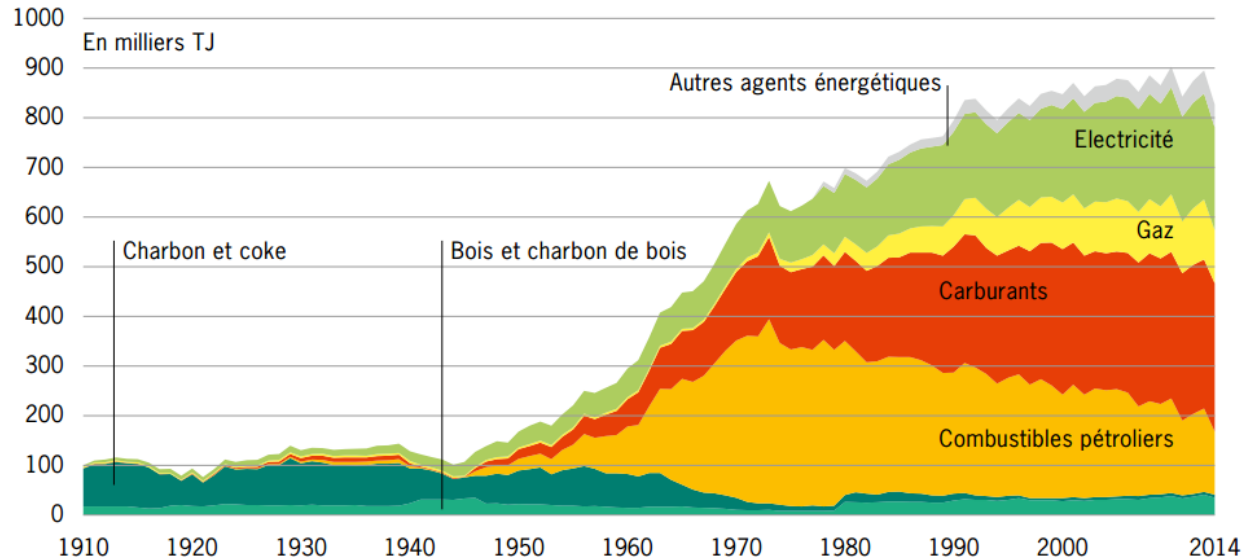


Figure 5 : Evolution de la consommation des différentes sources d'énergie en Suisse, 2014, [OFF-2014]

Maîtriser la consommation finale des bâtiments

Dans les pays industrialisés, les bâtiments consomment presque la moitié de l'énergie utilisée par la société et en conséquence sont source d'une partie non négligeable de la pollution (Figure 5). Cette énergie est utilisée pour de nombreux usages, notamment :

- ⇒ le chauffage et/ou le refroidissement, pour assurer un climat intérieur confortable
- ⇒ la circulation de fluides tels que l'air (ventilation), l'eau (eau chaude, chauffage)
- ⇒ les transports (ascenseurs)
- ⇒ l'éclairage
- ⇒ les communications (téléphone, radio, télévision)
- ⇒ la production de biens (fabriques, cuisines, couture, etc.)

Au cours de la période 1973-2014, la part de l'industrie incluant la sidérurgie dans la consommation finale d'énergie diminue fortement (36 % à moins de 20 %). Le secteur résidentiel-tertiaire passe de 42 % à plus de 45 %, alors que le secteur des transports progresse de moins de 20 % à environ 33 %. Enfin, la part de l'agriculture s'établit à 3 % en 2014. Le **secteur résidentiel et tertiaire** est le **premier poste de la consommation d'énergie finale**. En France, comme décrit dans la figure 9, ce secteur représente 43% de l'énergie finale totale. En 2014, la consommation d'énergie dans le résidentiel-tertiaire s'élève à 67,7 Mtep.

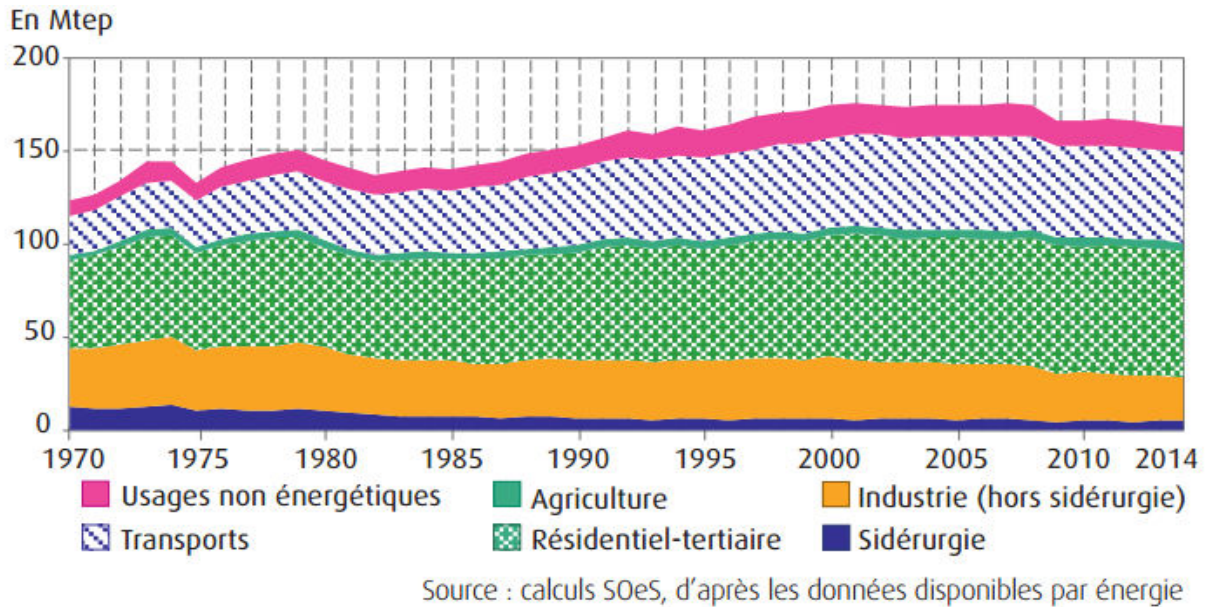


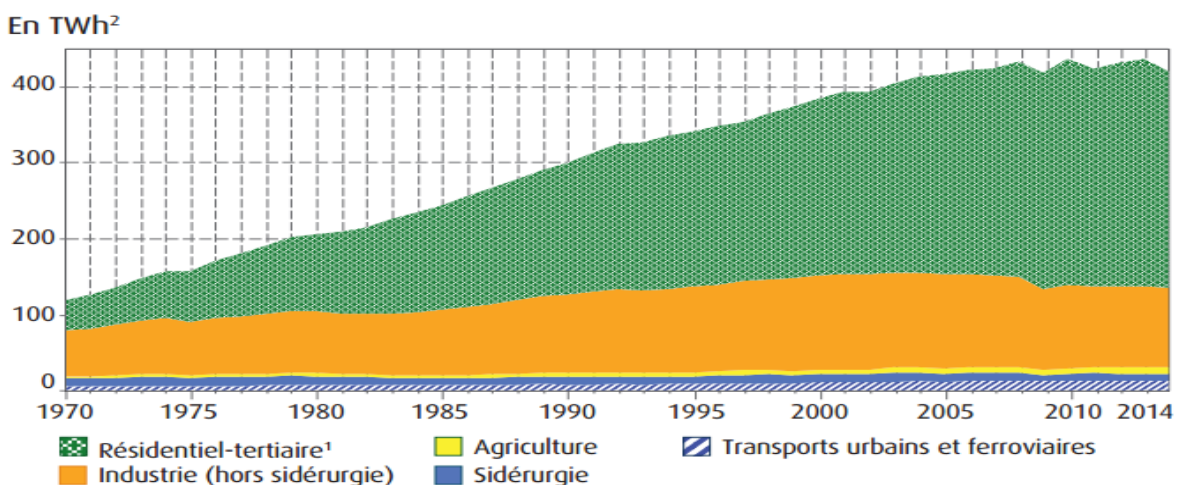
Figure 6 : Evolution de la consommation d'énergie finale par secteur en France de 1970 à 2014

La consommation se maintient ainsi à son niveau du début des années 2000, alors qu'elle augmentait régulièrement dans les années 1990 (+ 1,1 % par an en moyenne). Si nous prenons l'exemple des logements résidentiels et des bureaux, leurs consommations ont augmenté de 30% au cours des 30 dernières années liées à une :

- ⇒ croissance continue du parc de bâtiments
- ⇒ augmentation de la surface occupée par personne
- ⇒ augmentation des exigences de confort
- ⇒ apparition de nouveaux usages spécifiques électriques

Si nous prenons le secteur tertiaire, la consommation en énergie finale a augmenté de 27% entre 1990 et 2007 pour atteindre 223 TWh soit 15% de la consommation nationale.

Si nous regardons spécifiquement le cas de la consommation d'électricité, les secteurs tertiaires et résidentiels représentent la plus forte croissance pendant les cinquante dernières années pour être aujourd'hui la plus grande part (Figure 7) de la consommation d'électricité.



Sources : calculs SOeS, d'après l'enquête sur le transport et la distribution d'électricité ; RTE ; ERDF ; Rica

Figure 7 : Evolution de la consommation d'électricité finale par secteur en France corrigée des variations climatiques

En Suisse, la répartition par secteur de la consommation finale d'énergie en 2014 est la suivante : le secteur des transports représente 37,7 %, les ménages 26,5 %, l'industrie 19,0 %, les services : 15,8 % et l'agriculture 0,9 %.

Les deux tiers de l'énergie sont consommés par le résidentiel-tertiaire et sont consacrés au **chauffage**, 20 % à l'électricité spécifique (éclairage, climatisation...) et 14 % à l'eau chaude sanitaire et la cuisson. **L'électricité (40 %) est l'énergie prépondérante** dans la consommation totale, devant le gaz naturel (23 %), les produits pétroliers (20 %) et le bois énergie (16 %) [OFF-2015]. C'est pourquoi nous nous sommes également focalisés sur la prédiction de la consommation électrique totale.

En France, les normes de construction actuelles imposent aux bâtiments neufs de consommer peu d'énergie (consommation de référence fixée à 50 kWh/m² annuelle pour les Bâtiments Basse Consommation et ceux respectant la Règlementation Thermique 2012, mais le parc immobilier se renouvelle lentement avec environ 1% de bâtiment neuf par an d'après l'INSEE (Institut National de la statistique et des Etudes Economiques) en 2012 [ARB-2013]. La consommation annuelle moyenne tous des bâtiments tertiaires est de 211 kWh/m² ce qui est loin des standards actuels de construction.

Dans la nouvelle loi de transition énergétique en France, les bâtiments privés résidentiels dont la consommation en énergie primaire sera supérieure à 330 kWh/m² devront faire l'objet de travaux de rénovation avant 2025. Une des actions préconisées pour atteindre cet objectif est la mise en place de stratégies de Maitrise de la Demande en Energie (MDE), avec notamment l'amélioration de l'exploitation des systèmes de chauffage et refroidissement.

Limiter son usage dans la production d'électricité

Le nucléaire est la seule technologie qui est diminué sa production (- 1,1 % / an). Sa part de marché était de 15,7 % en 2004 ; elle baisse à 10,8 % en 2013. La production mondiale d'électricité renouvelable (y compris celle des grands barrages hydroélectriques) a augmenté sa part de marché de 17,5 % à 21,7 %. Les renouvelables compensent presque la diminution du nucléaire (production

électrique totale non-fossile de 32,5 % en 2013 comparée à 33,6 % en 2004). L'hydroélectricité est la première des énergies renouvelable avec 975 TWh de plus produite en 2013 par rapport à 2003 comme le montrent les figures 8 et 9. La production annuelle hydroélectrique mondiale croît de 3,4 %/ an en moyenne sur dix ans. Elle est supérieure à la croissance mondiale de la production électrique (3,1 % par an) [OBS-2013].

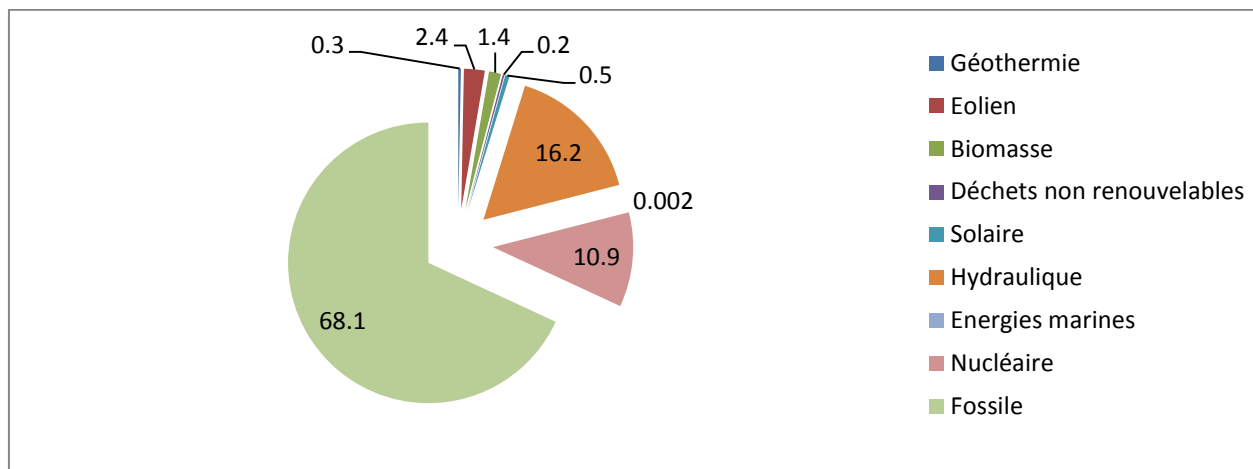


Figure 8 : Structure de la production d'électricité mondiale en 2013, [OBS-2013]

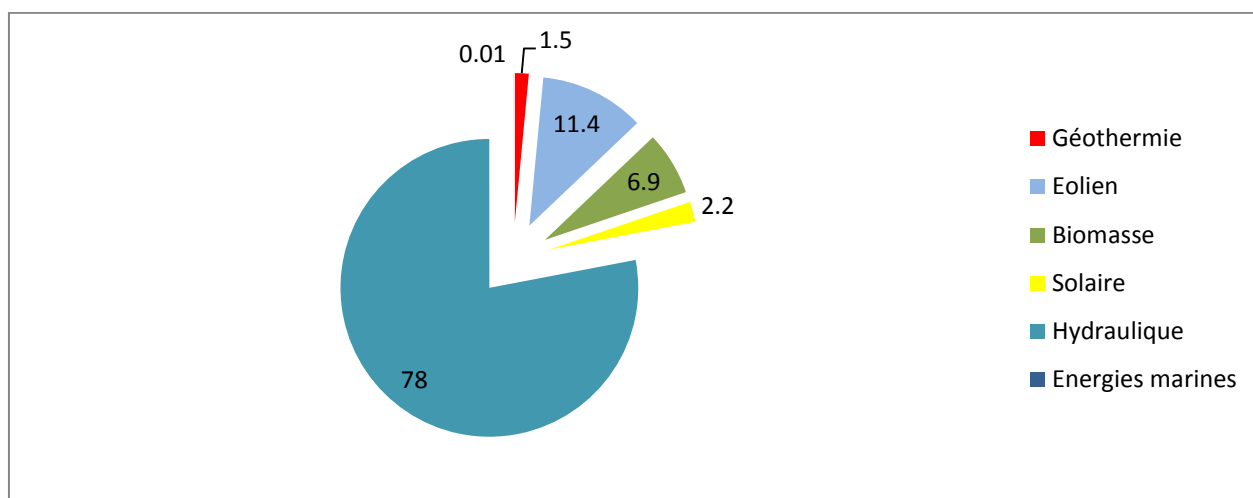


Figure 9 : Structure de la production électrique d'origine renouvelable mondiale en 2013, [OBS-2013]

Les énergies renouvelables non-hydroélectriques augmentent leur production au rythme de 15,6% par an, soit une production en 2013 équivalente à 3,67 fois la production de 2004. Leur part de marché passe de 1,91 % en 2004 à 5,34 % en 2013.

En France comme en Suisse et comme dans la plupart des pays industrialisés, l'énergie hydraulique a atteint son maximum de capacité [MIN-2015], [OFF-2015]. Des améliorations d'un point de vue technique et technologique sont encore possibles mais ils ne pourront pas compenser l'arrêt du nucléaire pour la Suisse, la baisse pour la France et le remplacement en général des énergies fossiles pour produire par exemple de la chaleur ou de l'eau chaude dans les bâtiments.

En Suisse, l'électricité est principalement produite à partir du nucléaire, d'hydraulique à accumulation, et d'hydraulique au fil de l'eau (Figure 10).

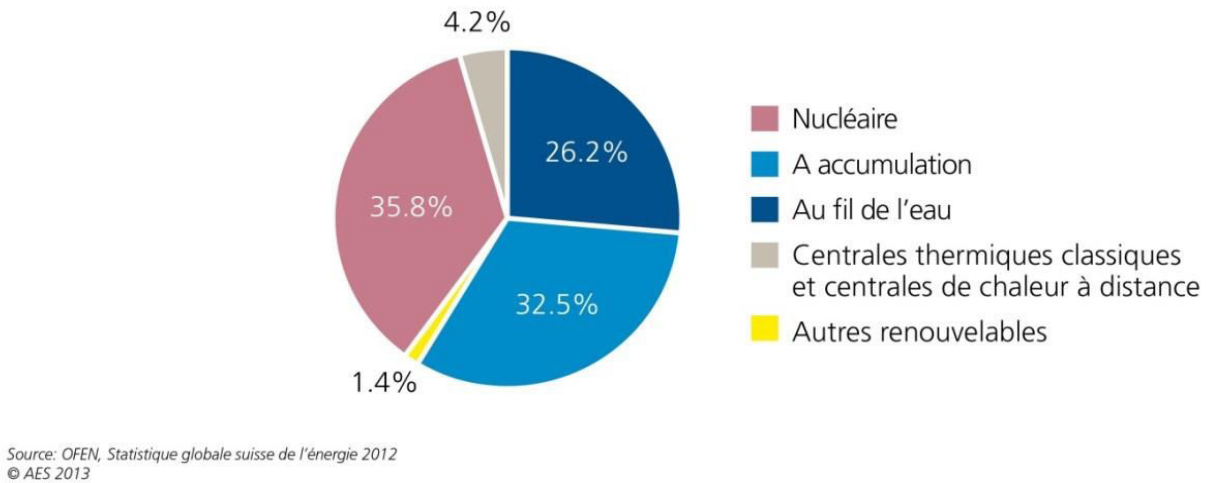


Figure 10 : Part des différentes sources de production dans le mix électrique Suisse en 2012

La Suisse a cependant marqué une volonté politique de sortir du nucléaire, elle qui produit plus d'un tiers de la consommation électrique finale [OFF-2015]. Il est alors nécessaire de mettre en place des systèmes permettant de combler ce manque, en augmentant l'efficacité énergétique des systèmes et en produisant plus d'électricité d'origine renouvelable.

L'augmentation du renouvelable passera par le développement de l'énergie solaire grâce à la compression des coûts de fabrication. Cependant ces énergies renouvelables sont totalement dépendantes des variations climatiques et leur variation de production augmente les coûts de raccordement et ceux liés aux erreurs de prédictions.

A titre d'exemple, la figure 10 décrit les taux de croissance des différentes énergies entre 2013 et 2015 dans le monde. Les énergies éoliennes terrestres et solaires connaissent le plus fort taux de croissance dans le monde et l'année 2015 est la première où le nucléaire décroît et est compensé par l'essor des énergies renouvelables.

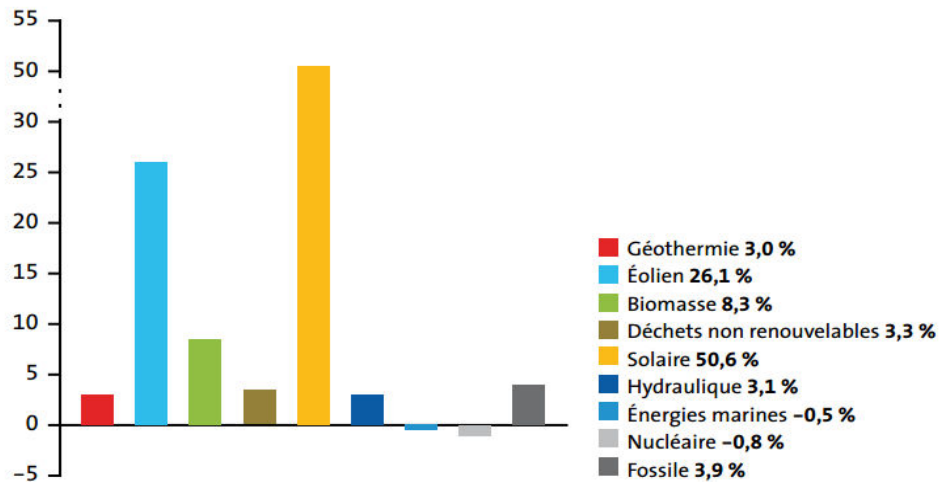


Figure 11 : Taux de croissance des différentes énergies entre 2013 et 2015, [MIN-2015]

En Suisse, plusieurs scénarios ont été réalisés pour une échéance à 2050. Le premier scénario est le scénario le plus plausible, où la réglementation aujourd'hui en vigueur est respectée et favorise l'insertion des énergies renouvelables dans le mix énergétique suisse.

Le deuxième est l'introduction de plus en plus importante des énergies renouvelables, sources d'énergies délocalisées, sur le réseau électrique.

Enfin, le scénario 3 décrit une forte croissance de l'énergie solaire sur le réseau qui devient de loin la première des sources de production d'énergie. Le graphique suivant montre l'évolution des énergies renouvelables en Suisse d'ici 2050 selon trois scénarios évalués par l'Association des Electriciens Suisses (Figure 12).

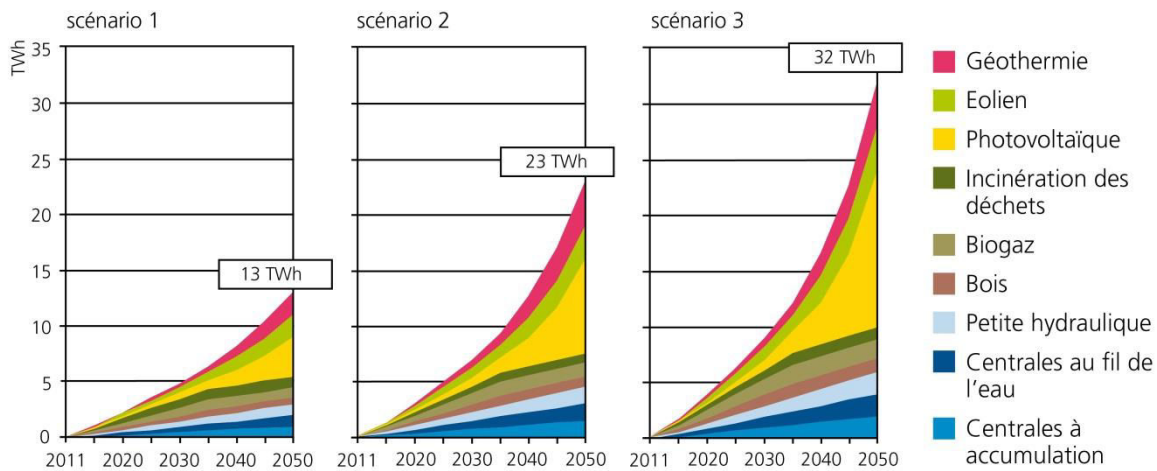


Figure 12 : Projections de la part des différentes énergies renouvelables selon trois scénarios pour la Suisse [AES-2013]

Quel que soit le scénario qui tend à se produire, le solaire photovoltaïque est amené à être de plus en plus développé sur le territoire, créant de multiples sources de production d'énergie intermittentes délocalisées. Ces énergies sont intégrées sur le réseau au niveau basse tension (400 Volts)

lorsqu'il s'agit de petites centrales de production ou moyenne tension (20000 Volts) pour des centrales plus importantes.

Leur intégration au réseau est donc difficile dans la mesure où ce dernier a été dimensionné de façon à acheminer de l'électricité depuis des centrales de grande production d'électricité au consommateur alimenté en 230 Volts, comme l'illustre le schéma ci-dessous fourni par l'Association des Electriciens Suisses (Figure 13).

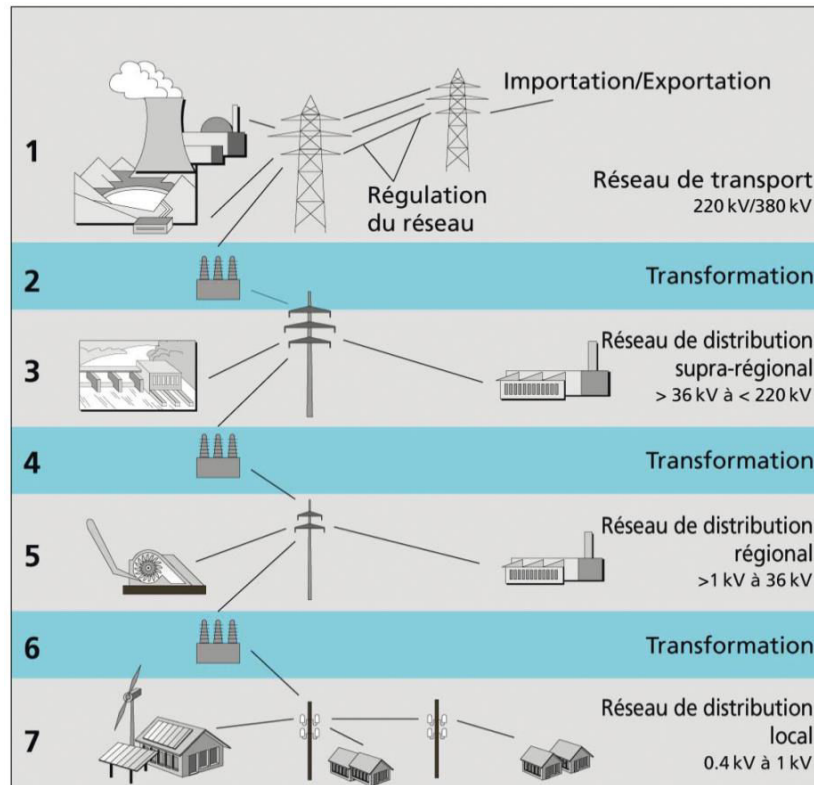


Figure 13 : Réseau de production et d'acheminement d'électricité en Suisse, [AES-2013]

Problématique

Limiter l'usage des énergies fossiles dans la stabilisation du réseau

Vu du réseau électrique, il est nécessaire à tout moment de garantir l'équilibre entre la production et la demande. Ces sources d'électricité d'origine renouvelable sont, pour la plupart, à production intermittente. De plus, les prédictions liées à ces productions renouvelables sont totalement dépendantes des données météorologiques et engendrent des erreurs importantes (Figure 14).

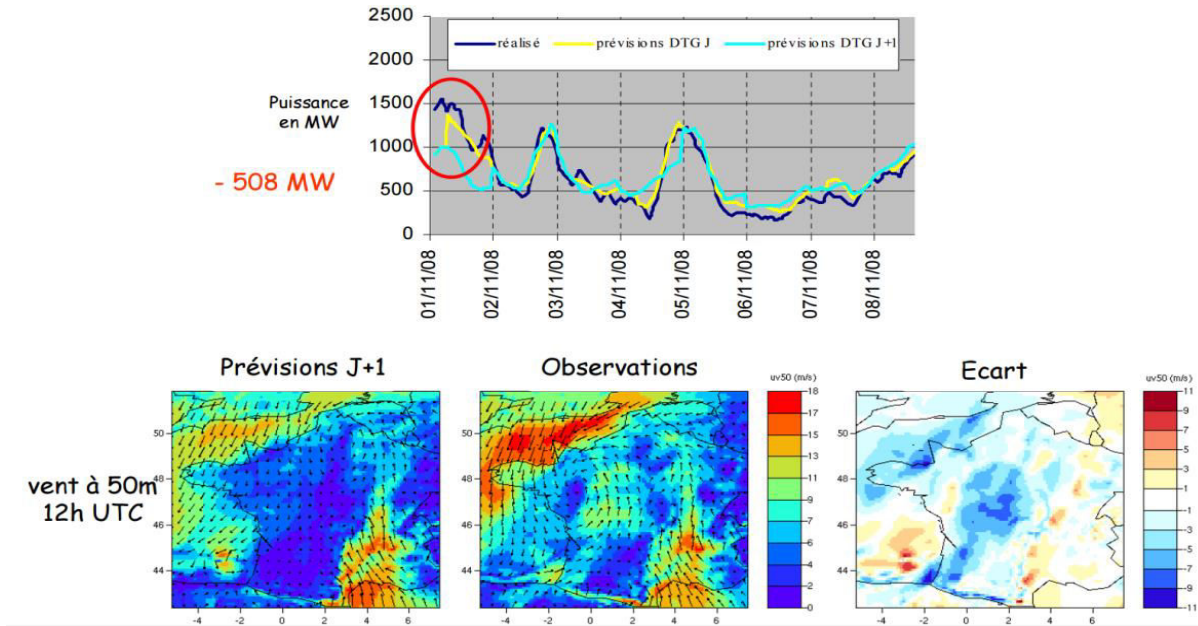


Figure 14 : Erreur de prédiction sur la production éolienne due à une erreur de prévisions météorologiques, [EDF-2013]

Face aux différentes variations de la demande et de la production des énergies renouvelables, les moyens de production d'énergie n'ont pas la même réactivité pour assurer l'équilibre du réseau électrique afin de combler de brutales variations de consommations et de productions d'origine renouvelable.

L'hydraulique est aujourd'hui la source de production qui permet de répondre le plus rapidement aux fluctuations du réseau électrique, de la minute jusqu'au quart d'heure. Les moyens thermiques ont besoin d'un horizon allant du quart d'heure à l'heure correspondant au délai d'allumage des dispositifs. Enfin la gestion des stocks et les différents achats selon le type de marché se passent de $t + 15$ minutes jusqu'à $t + 1$ semaine (Figure 15).

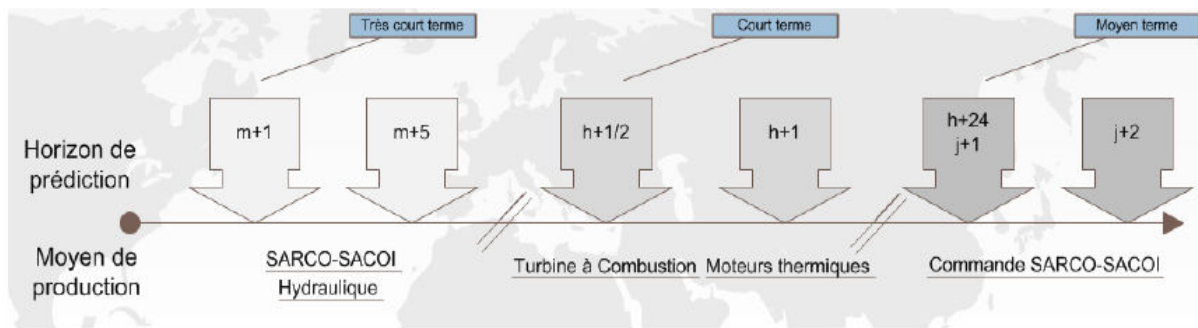


Figure 15 : Moyens de productions utilisés en fonction des ajustements possibles par rapport aux horizons de prédiction [EDF-2013]

Ainsi, différentes solutions sont mises en œuvre qui dépendent principalement de leurs réactivités à être utilisées :

- ⇒ Le recours à des **moyens flexibles de production de pointe**. Les centrales à gaz, au charbon ou au fioul sont les options les plus utilisées aujourd'hui puisqu'elles permettent une réactivité de quelques minutes à l'heure mais elles sont fortement émettrices de CO₂.
- ⇒ Les **interconnexions** sur les réseaux de transport et de distribution entre les régions et les pays européens permettent l'échange d'électricité ponctuelle dans un réseau électrique Européen. Cependant, les projets de nouvelles lignes sont coûteux et se heurtent à des problèmes d'acceptabilité, comme par exemple la mise en place de la ligne France-Espagne, avec un coût final sept fois supérieur aux prévisions et 14 ans de travaux ou encore la ligne Chamoson Ulrichen en Suisse.
- ⇒ La **maitrise de la demande en énergie (MDE)**. Elle permet d'effacer une partie de la consommation des industries et des ménages grâce à des contrats d'effacement. Dans un premier temps, seules les entreprises les plus consommatrices d'électricité sont intégrées au système de délestage et le secteur tertiaire et résidentiel n'est utilisé qu'en dernier recours. Actuellement, le délestage de charges reste une mesure de dernier recours en cas de surcharges en cascade, d'écroulement du plan de fréquence ou du plan de tension [RTE-2004].
- ⇒ Le **stockage d'énergie** est une des solutions transverse et complémentaire soit à partir de centrales solaires thermiques ou plus localement par des batteries électriques. Le bâtiment devient également un moyen de stockage à travers son inertie thermique et sa capacité de stockage lié au pilotage de l'eau chaude sanitaire.

Ces différentes solutions sont utilisées quand la balance entre la consommation et la production n'est plus assurée sur le réseau électrique qui est contrôlé par un suivi continu au niveau du réseau haute et moyenne tension. Pour anticiper leurs utilisations, des prédictions horaires sont réalisées pour le jour suivant avec un suivi à la minute ou au quart d'heure. Puisque le réseau électrique est historiquement centralisé dans les pays industrialisés, les prédictions le sont également.

Cependant, la **majorité des installations photovoltaïques sont de faible puissance et raccordées au réseau basse tension**, lieu où l'**information manque et éloigné des points de productions** centralisés. C'est pourquoi la collecte de données et le pilotage à ce niveau du réseau électrique deviennent nécessaires.

Le big data au service de la transition énergétique

Pour pallier les variations de productions d'énergie d'origine renouvelable et aux variations de consommations, la notion de « réseaux intelligents » SmartGrid est née à partir de la volonté de réguler l'énergie en particulier électrique au niveau du réseau moyenne et basse tension [DAV-2010] [GUE-2013].

L'idée de SmartGrid consiste à créer un sous-réseau dans lequel une connexion est créée entre les sources de production d'énergie et les puits de consommation d'énergie, de façon à équilibrer au niveau de ce sous-réseau les pics de production et de consommation. La notion de microgrid rejoint ce concept en se plaçant au niveau d'un transformateur 400 V [SHU-12] [RIV-2013]. Un transformateur

400 V permet d'adapter la tension en provenance du réseau à une tension d'alimentation pour un groupement d'appareils. Un transformateur 400V peut être connecté à une seule entreprise ou industrie si elle utilise beaucoup d'électricité, à un regroupement d'entreprises ou à un quartier résidentiel.

En Europe, nous pouvons citer le projet IssyGrid ou Nice Grid (Figure 16) [NIC-2014]. L'architecture des systèmes de stockage se décompose en 3 niveaux pour une capacité totale de 1,5 MW. Une batterie de 1.1 MW fait le lien entre le réseau RTE, (Réseaux et transport d'électricité) et EDF (Electricien de France), 3 batteries de 33 kW sont installées sur le réseau basse tension du quartier et des batteries de 4.6 kW chez les clients résidentiels. Ces batteries sont contrôlées par un agrégateur (Network Battery Aggregator) qui permet de collecter les paramètres d'état de chaque batterie, de proposer des offres de flexibilité au NEM (Network Energy Manager) et de répartir de façon optimale la décision du NEM. L'ensemble du système permet de piloter à distance les sources de flexibilité (stockage et report de consommation) à partir d'un centre d'optimisation tenant compte de l'état du réseau au niveau local et régional ainsi que des prévisions de consommation et de production.

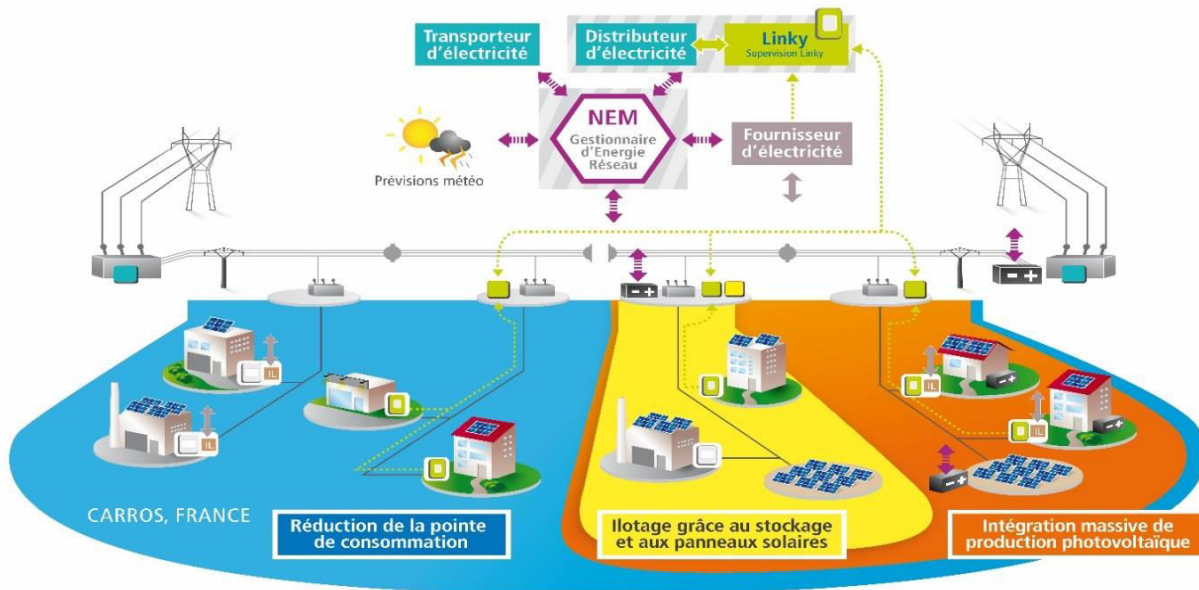


Figure 16 : Représentation du microgrid Nice Grid avec les différents niveaux de pilotage, [NIC-2014]

La plus grande batterie couplée à une production solaire et éolienne et connectée au réseau se situe en Chine et associe 140 MW d'énergie solaire et éolienne associée à une batterie Phosphate-ion de 6 MW et de 36 MWh [ABB-2013].

En Suisse, la plus grande batterie sur le réseau électrique a été inaugurée en 2012 à Dieketon dans le canton de Zurich. La technologie Li-ion utilisée permet de stocker 250 kWh d'énergie et à une puissance de 1 MW. Le dispositif permet d'exploiter les centrales renouvelables en évitant les problèmes de chute de tension ou de surtension et offre également la possibilité de fonctionner en mode îloté. A l'EPFL (Ecole Polytechnique Fédérale de Lausanne) en Suisse, des travaux portent dans le domaine de la simulation de MicroGrids en utilisant des scénarios prédéfinis pour le pilotage ainsi que des composants électroniques capables de simuler en temps réel le fonctionnement d'un SmartGrid. Nous pouvons également citer le MicroGrid de la HES-SO au TechnoPôle de Sierre détaillé dans le chapitre 3 qui

possède à la fois une station solaire, des mesures à la seconde de la consommation et de la production et une nouvelle technologie de batterie de l'entreprise Leclanché de 25 kWh qui permet de lisser les pics de puissance **[IBA-2013]**.

Le microgrid de Kythnos (Grèce) qui a été mis en place en 2010 et associe une batterie de 53 kWh et 10 kW de production PV réparti sur les toits des 12 maisons partageant le microgrid. Chaque maison est équipée d'un ILC (Intelligent load controller) permettant de contrôler l'état du réseau et 2 interrupteurs reliés à des équipements délestables choisis par le propriétaire de la maison. Le système de stockage dispose de son propre agent permettant une gestion optimale des batteries. Ce système multi-agent dispose d'un contrôleur central qui récolte l'ensemble des décisions collectées, surveille l'équilibre offre = demande et joue le rôle de l'intelligence décisionnelle en utilisant les prévisions de production et de consommation.

Enfin, nous pouvons citer le projet Européen ENTROPY **[ENT-2015]** mené à l'Institut d'Informatique et de Gestion (IIG) de la HES-SO qui va proposer une plateforme basée sur le principe de jeu interactif (Serious Gaming). Celle-ci exploite les données issues de divers capteurs (luminosité, température, présence...) et permettra de diminuer la consommation d'énergie de façon interactive dans les entreprises.

Limites des études précédentes

Des dizaines de microgrids ont été créés depuis plus de cinq ans pour piloter les consommations d'énergie des bâtiments. Pour la plupart de ces projets, les fournisseurs et producteurs d'électricité nationaux sont partenaires et les solutions de pilotage menées reposent sur des solutions fermées dépendant de solutions propriétaires. Aujourd'hui, les compétences de pilotage des réseaux électriques sont la propriété intellectuelle des seuls fournisseurs et producteurs d'électricité. Nous voulons élargir ces connaissances à l'ensemble des acteurs de la transition énergétique.

De plus, le pilotage de ces quartiers intelligents repose sur des batteries électriques avec une vision purement électrique du problème. Le problème de la fabrication, de la durée de vie et la dépollution en fin de vie des batteries engagent également les acteurs de la transition énergétique. Le problème de la régulation du réseau doit être pris d'un point de vue énergétique et inclure l'eau chaude et le chauffage pour compléter l'offre de stockage.

Nous remarquons également une tendance à un pilotage centralisé où l'ensemble des informations d'un quartier est centralisé pour ensuite permettre le pilotage des différents éléments connectés au système d'information. Les données, en particulier liées à la consommation, ne sont pas obligées de transiter par un niveau supérieur d'analyse. Cette analyse pourrait se faire localement avec comme seul envoi les données utiles au pilotage (prédiction de la consommation électrique, capacité d'effacement du système (prédiction du besoin en eau chaude, prédiction du besoin en chauffage, stockage électrique, prédiction de la production).

Les études prédictives de pilotage ne travaillent actuellement que sur un seul flux : les besoins en chauffage ou la production solaire décentralisée. Un outil et une méthodologie d'approche multi-flux (consommation électrique globale, chauffage/refroidissement, eau chaude sanitaire et production décentralisée) permettraient d'être plus réactif car plus précis.

Au niveau de la prédiction de consommations de chauffage et de climatisation, les premières études pour comprendre le comportement du bâtiment dans un environnement donné, datent des années 1970. Les facteurs entrant en compte dans la consommation énergétique des bâtiments sont nombreux : facteurs météorologiques, architecture du bâtiment, matériaux de construction mais aussi le comportement des usagers et le contexte socio-économique.

Aujourd'hui il demeure encore une grande **incertitude** qui porte **sur les usages, le comportement des occupants et le besoin en données locales**. Les axes de recherche proposés aujourd'hui s'attèlent d'une part à identifier la présence humaine de manière intrusive (port d'un bracelet par exemple) pour diminuer l'erreur de prédiction et d'autre part à interagir avec l'humain pour l'inciter à diminuer sa consommation par le jeu ce qui complexifie le système d'information à installer.

Les modèles de prédiction de consommation sont très importants, les échelles de temps sont multiples et les prédictions sont réalisées à l'heure, à la journée, au mois, à l'année, etc. **[REB-2015] [MOU-2016] [CHE-2016]**.

Dans **[ZHA-2016]**, la série temporelle de consommation est prédite à partir d'une méthode hybride qui utilise une régression suivie d'une méthode d'apprentissage avec l'algorithme SVM pour Support Vector Machine. Des désagrégations de courbes de charges globales sont également réalisées pour identifier les appareils électriques pour une meilleure prédiction de la consommation électrique totale **[KOL-2012] [PAR-2011]**. D'autres études utilisent des **modèles réduits** pour simplifier les modèles thermiques détaillés. Le bâtiment est modélisé par un circuit électrique équivalent simple (RC, R2C2, R3C2, R6C2) **[JAZ-2002] [ARB-2015]**.

Au niveau de la prédiction de production solaire décentralisée, les différentes études concernent un large choix d'installations photovoltaïques et de type de données, (étude sur 12 cellules PV en laboratoire ou de 14 000m² de panneaux) **[ZAM-2014] [GRA-2016] [GUL-2016]**. D'après la bibliographie, les paramètres pouvant influencer la prédiction sont tellement nombreux, qu'il est difficile d'utiliser les résultats issus d'autres études. Ainsi, il peut être intéressant d'essayer des méthodes ou des paramètres même s'ils n'ont pas forcément fait leurs preuves dans d'autres études. Les données de prédiction météorologique sont parfois corrigées lors d'une première phase de test afin d'améliorer la prédiction de production des panneaux. Les données proviennent d'une seule station météo locale ou régionale.

De plus, les quartiers intelligents utilisent le plus souvent le secteur résidentiel ou des campus universitaires comme bâtiment d'étude. Il serait intéressant de généraliser une méthodologie de pilotage par prédiction à tous les secteurs d'activité.

Enfin, l'apport des objets connectés peut en partie répondre aux problématiques de prédictions de la consommation énergétique par détection de la présence humaine. Cependant, l'intrusion de nouveaux objets complexifie le système d'information dans son ensemble, du transit des données à la maintenance de ces objets. De plus, ces objets connectés communiquent avec l'extérieur et leurs données peuvent être envoyées et stockées dans des pays tiers.

Pour imaginer un grand déploiement des quartiers intelligents, la solution doit être la plus simple possible, robuste et généralisable en minimisant le système d'information, c'est-à-dire le nombre de capteurs et de données en transit.

Objectifs de la thèse

Cette thèse concerne l'utilisation des systèmes d'information avec ce que l'on appelle aujourd'hui le « Big Data » et les domaines de la prédiction par des méthodes par apprentissage dans le domaine énergétique (Figure 17).

Dans ces travaux, il s'agit d'implémenter une intelligence embarquée dans un bâtiment qui contrôlerait une partie de l'énergie selon des scénarios prédéfinis et validés par l'utilisateur final en concertation avec le nouveau fournisseur d'énergie appelé agrégateur. Il s'agit d'augmenter la flexibilité de la demande locale en énergie.

Notre niveau d'étude est **le bâtiment** dans le cas d'un quartier résidentiel ou **des parties de bâtiment** (un restaurant d'entreprise à l'intérieur d'un complexe de bureaux par exemple) (Figure 18).

Notre premier objectif est de réaliser un **bilan énergétique prédictif** pour l'heure suivante. Cela permettra d'avoir des possibilités de **scénarios de surproductions ou de surconsommations à court terme pour chaque flux énergétiques (t + 1 heure) au niveau d'un quartier**. Cela implique une prédiction fine des différents flux énergétiques d'un système, en particulier le chauffage et l'eau chaude, qui représentent le plus gros potentiel de flexibilité dans les bâtiments [ANG-2004] et permettra ainsi de **compléter l'offre de stockage**.

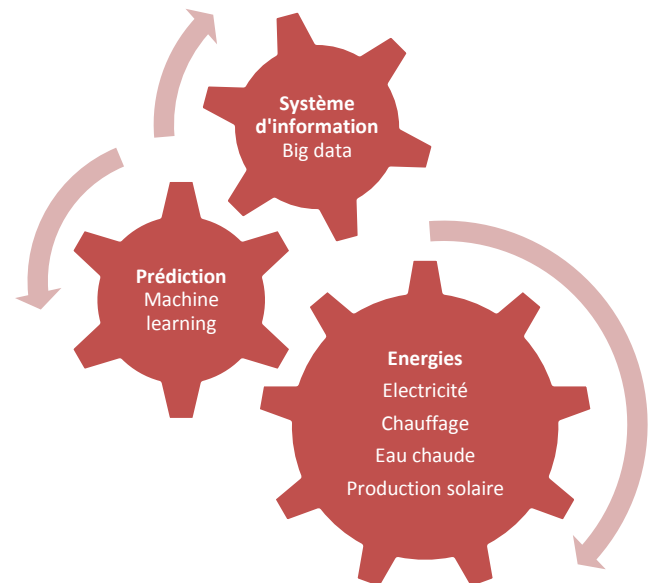


Figure 17 : Domaines d'application de la thèse pour la mise en place d'un pilotage énergétique décentralisé

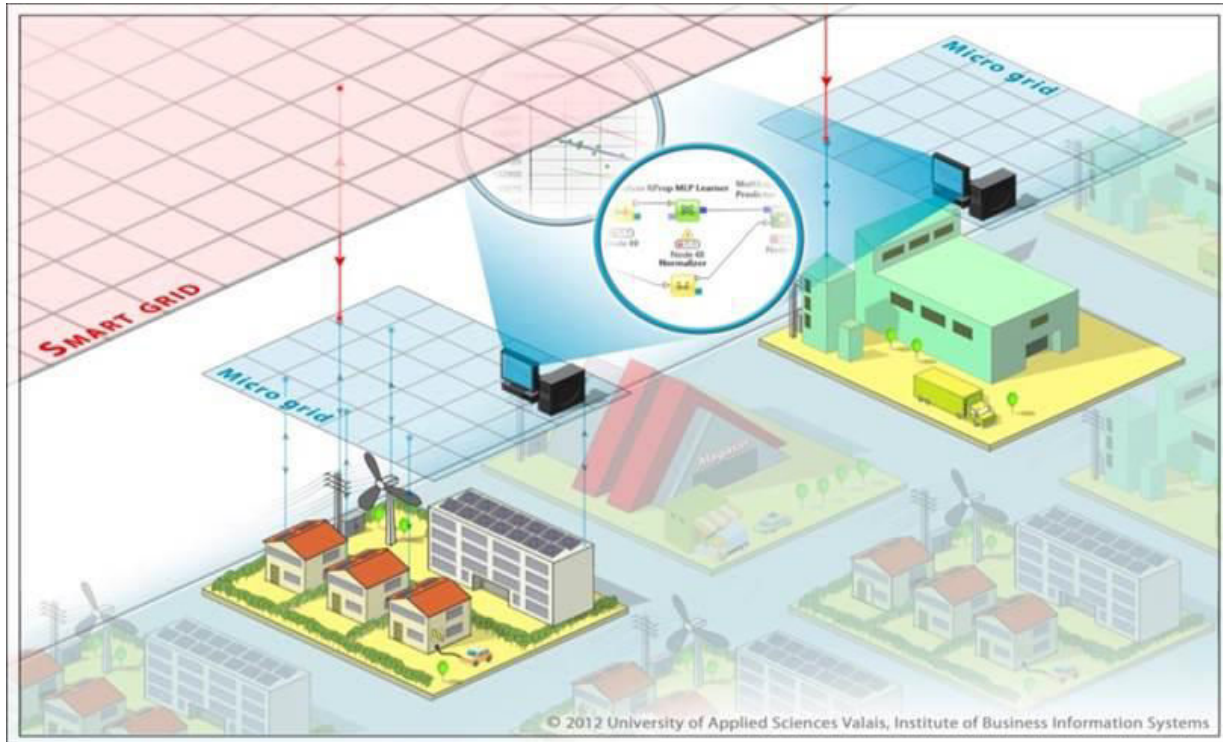


Figure 18 : Représentation d'un microgrid et de son insertion au smartgrid [IBA-2012]

Le besoin en énergie d'un système comme une maison est ainsi décomposé en un besoin en **chauffage**, en un besoin en **eau chaude sanitaire**, en **luminaires**, en besoin de **ventilation** et en **usages spécifiques électriques** utiles. Le système peut posséder une **production décentralisée** et un système de stockage ce qui augmentera sa capacité d'effacement.

Hormis l'existence pour le chauffage de logiciels puissants mais qui demandent beaucoup d'informations a priori sur le système pour être exécuté, il existe des incertitudes fortes sur la prédiction locale des flux énergétiques d'un système et ainsi une forte différence entre les simulations et la réalité au niveau des consommations et des productions. C'est pourquoi deux niveaux d'aide au pilotage ont été réalisés : un niveau d'anticipation et un niveau réactif. Le premier niveau s'appuie sur les méthodes existantes connues dans le monde énergétique (utilisation de logiciels de simulation thermique détaillée, de production solaire, normes SIA **[SIA-380]**) pour construire des simulations horaires. Cela nous permettra de construire des modèles de référence à partir desquels des études énergétiques peuvent être réalisés et servir de modèles de prédictions de base quand le système n'a pas encore récolté de données.

Le second niveau utilisera des méthodes d'apprentissage pour construire des modèles basés sur le passé qui caractérise le présent pour prédire le futur. Les modèles de prédiction de consommation sont très importants, les échelles de temps sont multiples, les prédictions sont réalisées à l'heure, à la journée, au mois, à l'année et les modèles d'apprentissage utilisés sont des réseaux de neurones **[REB-2015] [MOU-2016] [CHE-2016]**.

Les différents modèles de prédictions pour ces différents flux devront être généralisables : des méthodes paramétriques et non paramétriques seront testées, comme les réseaux de neurones, mais également des méthodes plus récentes comme les ensembles d'arbre de décision. Des études multi-échelles et multi-secteurs sur des cas réels seront réalisées.

Le deuxième objectif est de connaître les paramètres qui permettent de **construire un système d'information local** afin de piloter énergétiquement le système par charges. Nous voulons réduire au maximum le nombre de capteurs installés, le nombre de données stockées et en transit dans le système d'information. Nous voulons créer un **système non intrusif**, c'est-à-dire qu'**aucun capteur n'est présent dans le système considéré**.

Les mesures se font au niveau des différents compteurs globaux d'énergie de consommations et de productions. Dans notre cas, le compteur global d'énergie lié à la consommation est le compteur d'électricité. A partir de cette seule information, nous essayerons de tirer le maximum de connaissances, en particulier l'identification des appareils électriques de manière non intrusive. Cela permet au système, muni d'un seul appareil de mesure au niveau du compteur global, de fournir **un outil de suivi de la consommation par usages** pour le consommateur final. Il permet également d'**augmenter la précision de la consommation électrique** en détectant les habitudes de consommation (par exemple, allumage de l'électroménager et du chauffage la nuit).

De plus, si le système d'étude permet un effacement du chauffage et de l'eau chaude, cela permet de **limiter le nombre de données** en transit et stockées. Seules des informations de pilotage de type ON-OFF seraient envoyées.

Pour permettre le pilotage énergétique par prédiction, nous devons répondre aux questions énoncées ci-dessous par flux énergétique étudié :

- ⇒ Quelles **données** ?
- ⇒ À quelles **fréquences de collecte** ?
- ⇒ Avec quel **système d'information (appareil, stockage...)** ?
- ⇒ Avec quelles **méthodes de prédictions** ?
- ⇒ Avec quels **modèles mathématiques** ?
- ⇒ Pour quels **niveaux de prédictions** ?
- ⇒ Avec quel **historique** ?
- ⇒ Sur quel **horizon** ?

Pour permettre de répondre à ces questions, nous sommes partis des mesures locales à la seconde des différents flux énergétiques. Les mesures météorologiques ont été prises à partir d'une cartographie de stations météorologiques entourant nos différents sites étudiés.

Chapitre 1 : Caractérisation des usages énergétiques

1.1 Introduction

L'objectif de ce chapitre est d'extraire des méthodes paramétriques, les caractéristiques indispensables à la prédiction de flux énergétiques. Cela nous permettra de définir les capteurs et les données d'entrée par modèle et de définir si besoin les informations a priori nécessaires pour réaliser la prédiction. Cela introduit le premier niveau d'aide à la décision qui caractérise par des modèles de référence les différents flux énergétiques. Ces modèles sont utilisés aujourd'hui par les cabinets d'étude qui les utilisent pour proposer des actions de réductions du besoin en chauffage notamment.

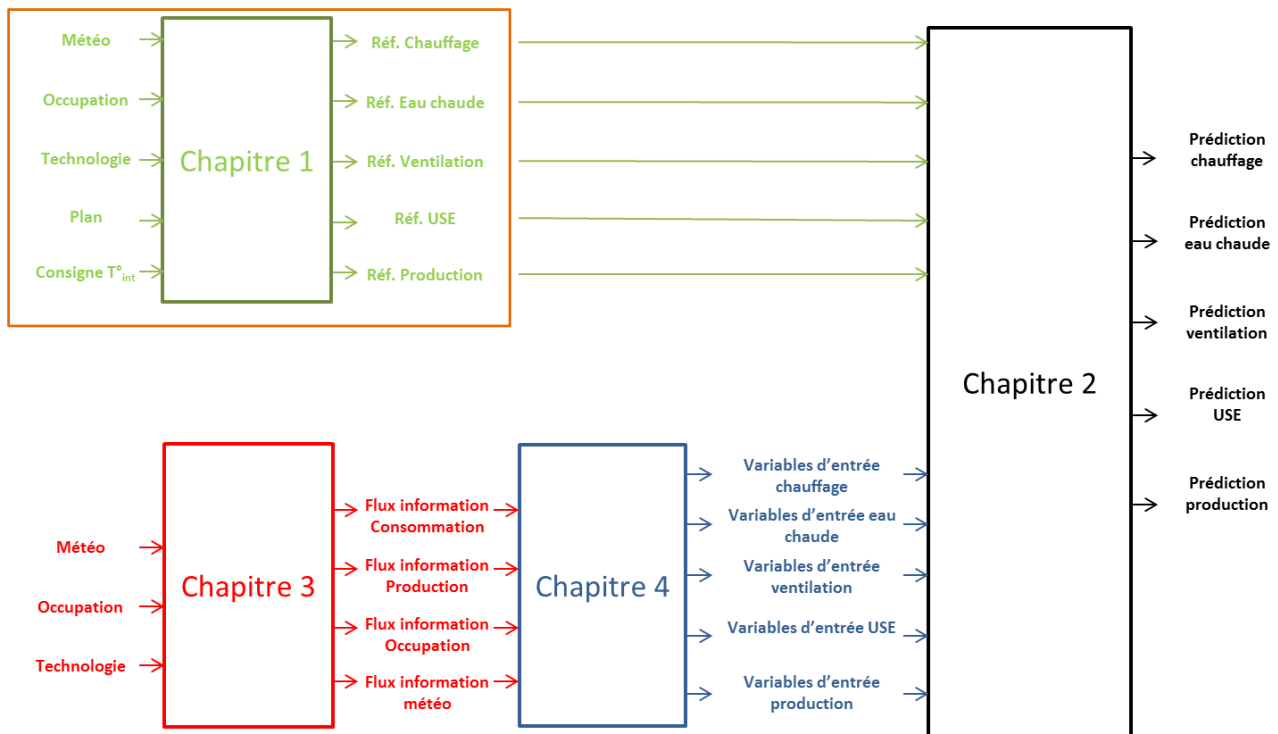


Figure 19 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision

1.2 Définition du système

Avant de calculer le bilan, il convient de **délimiter le système étudié** dans l'espace et dans le temps, et de définir les utilisations de l'énergie et les vecteurs énergétiques que l'on va considérer. La **limite de notre système d'étude est au niveau d'un bâtiment ou un ensemble de bâtiments** situé après le **réseau basse tension 400 V** relié à un compteur d'énergie.

A l'intérieur de ce système d'étude, plusieurs sous-systèmes peuvent exister **déterminés par une surface de référence** énergétique [SIA-380] liée au compteur d'énergie électrique. Ce compteur peut être relié à un bâtiment complet (maison individuelle) ou à une surface donnée (locaux de bureaux à l'intérieur d'un bâtiment).

La **délimitation spatiale** consiste à définir les frontières du domaine étudié, au travers desquelles passent les flux d'énergie à calculer. Pour un bâtiment, cette frontière est généralement constituée par :

- ⇒ l'enveloppe du bâtiment (Figure 18)
- ⇒ les compteurs d'entrée des sources d'énergie de réseau (électricité, gaz, chauffage à distance)
- ⇒ les entrées des combustibles (pétrole, charbon, bois)
- ⇒ les surfaces de captage d'énergie solaire
- ⇒ les raccordements d'entrée de l'eau froide
- ⇒ les raccordements de sortie des égouts

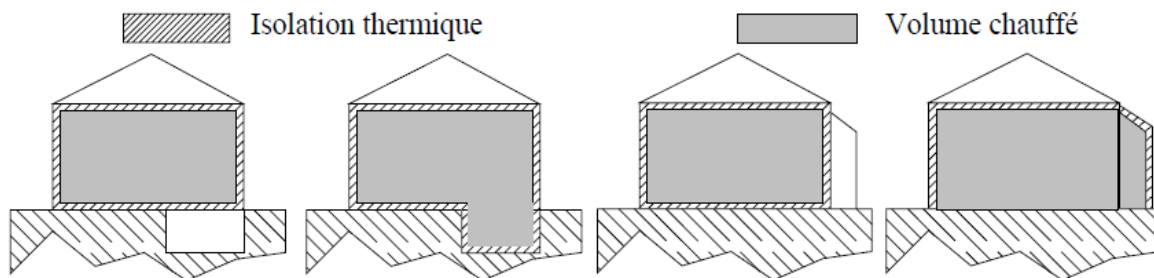


Figure 20 : Représentation de l'enveloppe du bâtiment à prendre en considération pour un bilan énergétique

La **délimitation par utilisation et par vecteur** permet de définir le système énergétique dont on s'occupe. Ces divers systèmes interagissent entre eux : ce qui peut être une perte pour un système (par exemple les pertes thermiques du four de cuisson) peut être un gain pour l'autre (ici pour le chauffage pendant l'hiver). Sans délimiter le système, on ne pourra pas définir les gains et les pertes.

Dans le bâtiment, ces systèmes sont notamment:

- ⇒ le système de chauffage
- ⇒ le système de climatisation
- ⇒ l'eau chaude
- ⇒ la cuisson
- ⇒ l'électroménager
- ⇒ l'éclairage
- ⇒ les transports et télécommunications

Les **vecteurs énergétiques** à considérer sont:

- ⇒ les combustibles (mazout, charbon, gaz, bois)
- ⇒ la chaleur à distance

- ⇒ l'électricité
- ⇒ le soleil
- ⇒ la chaleur humaine et animale

1.3 Notations

Chaque heure, notre outil doit fournir pour l'heure suivante le besoin énergétique d'un ou de plusieurs systèmes noté S , allant de 1 à n :

$$E t(i) = \sum_{S=1}^{S=n} E t(i) \quad (1.1)$$

Nous décomposons le besoin en énergie d'un système en un besoin en chauffage, un besoin en eau chaude sanitaire, un besoin en luminaires, un besoin en ventilation et un besoin en USE pour usages spécifiques électriques :

$$E t(i) = (E_{\text{Chauffage}} + E_{\text{Eau chaude}} + E_{\text{USE}} + E_{\text{Ventilation}} + E_{\text{Luminaires}}) t(i) \quad (1.2)$$

Selon les systèmes, la consommation électrique globale peut rassembler plusieurs de ces énergies.

Nous définissons la consommation électrique globale en sortie du compteur qui est égale à la somme des différents usages U électriques à l'intérieur du système étudié :

$$E_{\text{Consommation_électrique_totale}} t(i) = \sum_{U=1}^{U=d} E_{\text{Usages_électriques}} t(i) \quad (1.3)$$

Le système peut être relié à un onduleur qui ajuste la consommation et la production décentralisée. Avec $E_{\text{ENRE}}(t)$ l'énergie produite par la production décentralisée renouvelable électrique, nous avons :

$$E t(i) = (E_{\text{Chauffage}} + E_{\text{Eau chaude}} + E_{\text{USE}} + E_{\text{Ventilation}} + E_{\text{Luminaires}} - E_{\text{ENRE}}) t(i) \quad (1.4)$$

Le système peut posséder des panneaux thermiques pour la production d'eau chaude. Avec E_{ENRT} la source d'énergie décentralisée renouvelable thermique, nous avons :

$$E t(i) = (E_{\text{Chauffage}} + E_{\text{Eau chaude}} + E_{\text{USE}} + E_{\text{Ventilation}} + E_{\text{Luminaires}} - E_{\text{ENRE}} - E_{\text{ENRT}}) t(i) \quad (1.5)$$

Une partie de ces usages énergétiques peut être pilotée basée sur des prédictions par vecteur énergétique :

$$E t(i) = (+/-E_{\text{Chauffage}} +/- E_{\text{Eau chaude}} +/- E_{\text{USE}} +/- E_{\text{Ventilation}} +/- E_{\text{Luminaires}} - E_{\text{ENRE}} - E_{\text{ENRT}}) t(i) \quad (1.6)$$

Enfin, le système peut posséder un stockage électrique par batterie par exemple :

$$E t(i) = (+/-E_{\text{Chauffage}} +/- E_{\text{Eau chaude}} +/- E_{\text{USE}} +/- E_{\text{Ventilation}} +/- E_{\text{Luminaires}} +/- E_{\text{Stocké Batterie}} - E_{\text{ENRE}} - E_{\text{ENRT}}) t(i) \quad (1.7)$$

1.4 Le besoin en énergie de chauffage/climatisation

Le bilan énergétique à chaque pas de temps est une balance des entrées et des sorties d'énergie du bâtiment pendant une période de temps donnée. Ce bilan doit être équilibré, par conservation de l'énergie. Le bilan énergétique détaille toutes les pertes et tous les gains, les sommes des gains et des pertes étant égales si la période de consommation est suffisamment grande (par exemple une année, voire un mois s'il n'existe pas de capacité de stockage particulièrement grande).

L'énergie en chauffage et climatisation est la différence entre les déperditions liées à la ventilation, à la transmission et par le système d'eau chaude et les gains liés à l'apport solaire, la chaleur émise par les personnes et une énergie stockée par le bâtiment qui représente l'inertie thermique du bâtiment ou un système de stockage. Il est positif ou négatif selon la période (hiver ou été) et la température extérieure et intérieure.

Le besoin de chauffage se calcule par la différence entre les déperditions par l'enveloppe et par la ventilation, et les gains liés à la chaleur dégagée par les occupants, l'apport des gains solaires, la chaleur dégagée des appareils en fonctionnement et par l'apport d'énergie stockée dans le système (Figure 21) donné par l'équation :

$$E_{\text{Chauffage}} t(i) = \left(\begin{matrix} D_{\text{ventilation}} + D_{\text{Transmission}} + D_{\text{Eau chaude}} \\ -G_{\text{Solaires}} - G_{\text{Appareils}} - G_{\text{Personnes}} - E_{\text{Stocké}} \end{matrix} \right) t(i) \quad (1.8)$$

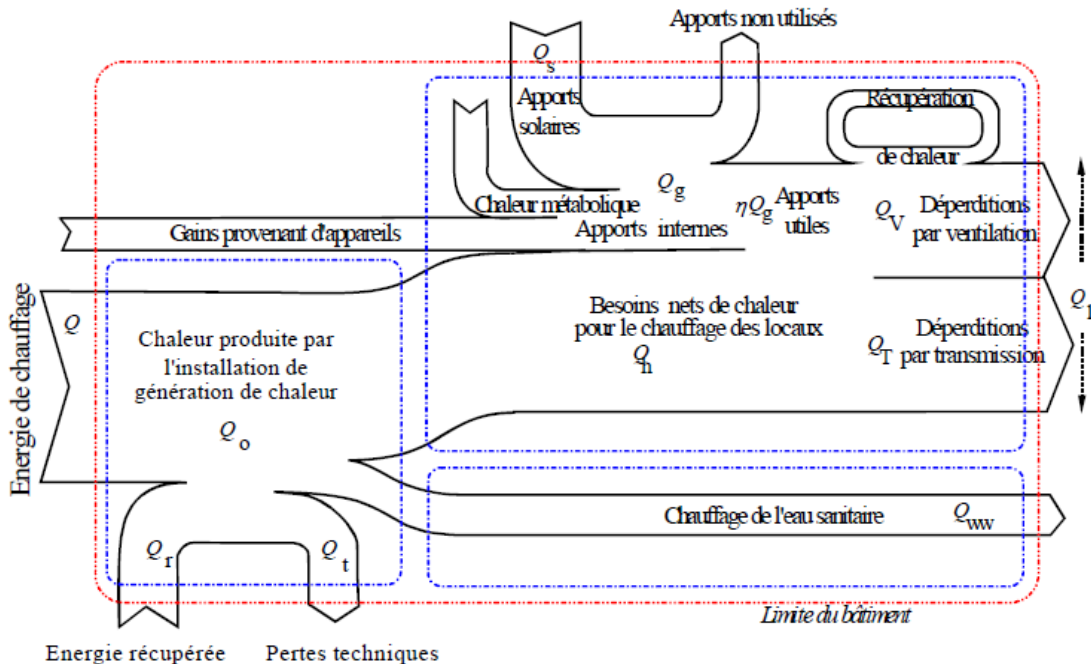


Figure 21 : Flux globaux d'énergie pour caractériser le besoin en chauffage [LAD-2013]

1.4.1 Déperditions

Les déperditions par l'enveloppe se situent au niveau du sol, des zones en contact direct avec l'extérieur, et par l'intermédiaire de zones non chauffées (Figure 22).

Les déperditions liées à une surface S peuvent s'exprimer en fonction de la différence entre les températures intérieures et extérieures (Equation 1.9) :

$$Q = H(T_{int} - T_{ext}) \quad (1.9)$$

- ⇒ T_{ext} : Température extérieure
- ⇒ T_{int} : Température intérieure
- ⇒ H : Coefficient de déperditions

$$H = H_T + H_V \quad (1.10)$$

- ⇒ H_V : Coefficient de déperditions par renouvellement d'air
- ⇒ H_T : Coefficient de déperditions par transmission à travers l'enveloppe

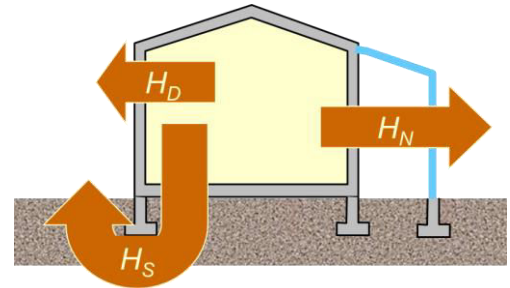


Figure 22 : Types de transmissions possibles des flux énergétiques

Le coefficient de déperditions par transmission H_T , se calcule par:

$$H_T = H_D + H_S + H_N \quad (1.11)$$

- ⇒ H_D : Transmission directe vers l'extérieur, à travers l'enveloppe du bâtiment
- ⇒ H_S : Déperditions par le sol
- ⇒ H_N : Déperditions à travers les espaces non chauffés

Les déperditions par ventilation dépendent de la masse d'air chauffé m_a , de la chaleur spécifique de l'air c_a et du rendement de récupération de chaleur sur l'air évacué η_r :

$$Q = m_a c_a (T_{int} - T_{ext})(1 - \eta_r) \quad (1.12)$$

- ⇒ m_a : Masse d'air chauffé
- ⇒ c_a : Chaleur spécifique de l'air, à savoir environ 1000 J/(kg·K)
- ⇒ η_r : Rendement de récupération de chaleur sur l'air évacué.

1.4.2 Etude des apports internes

Pour pouvoir calculer précisément les gains internes, il faut connaître les puissances dégagées par chaque apport. Les sources dissipent l'énergie de manière radiative et convective dans des proportions variables selon leur origine. Le gain de chaleur par rayonnement est partiellement absorbé par les parois et les objets environnants, mais n'affecte l'air de la pièce qu'après un certain temps [ASH-1997].

La part convective du gain de chaleur est transmise directement à l'air de la pièce par convection naturelle ou forcée sans retard. Les méthodes de calcul des besoins en chauffage se contentent de valeurs globales. Par exemple, dans les calculs réglementaires, ces apports sont évalués à 100 Wh par jour et m^2 de surface habitable [TEC-1998].

Gains dus aux occupants

Les gains de chaleur apportés par les personnes dépendent de la puissance dégagée par personne pour une surface donnée :

$$G_{Personnes} = p \frac{P h}{24} = A \frac{P h}{24D} \quad (1.13)$$

Avec

- ⇒ p : Nombre de personnes présentes dans la zone chauffée
- ⇒ P : Puissance dégagée par personne
- ⇒ h : Temps de présence par jour
- ⇒ A : Surface brute de plancher chauffé occupée par personne
- ⇒ D : Surface disponible par personne

Les normes SIA définissent des taux de présence différents selon les secteurs d'activité, qui, associées à une surface par personne, permettent de définir une énergie à partir d'une puissance spécifique. Pour un modèle de référence sans information sur le planning de production par exemple, on pourrait admettre une activité moyenne et classer les habitants en deux catégories: adultes et enfants. La norme SIA 380/1 propose les nombres présentés dans le tableau 1.

Secteur d'activité	Occupation (m2/personne)	Présence (heure/jour)	Puissance (Watt)
Logement collectif	40	12	70
Logement individuel	60	12	70
Bureau	20	6	80
Restaurant	5	4	70
Ecole	10	3	100

Tableau 1 : Puissance thermique dégagée par les habitants, norme SIA 380/1

Gains dus aux appareils

Il n'existe pas de méthode de calcul universelle pour calculer les apports internes apportés par les appareils. Dans [HOS-1998], les auteurs ont proposé d'utiliser un radiomètre scanneur pour mesurer la part radiante de la chaleur provenant des équipements des bureaux mais cette méthode dépend des conditions expérimentales (température de la pièce, mouvement de l'air, température des parois, variation de la température de la pièce, choix de la position et du nombre des points de mesure) et ne donne pas le total des gains internes provenant de ces appareils. [WIL-1991] ont mesuré la chaleur générée par les équipements des bureaux et ont comparé ces valeurs aux puissances données par le fabricant. Les études menées ont montré qu'il faut appliquer un facteur d'usage à la puissance donnée par le fabricant. Ce facteur varie entre 0.3 et 0.79. La valeur généralement utilisée est égale à 0.5, surtout pour les équipements dont le cycle nécessite le maintien d'une température constante.

$$P_a t(i) = P_a f_e t(i) \quad (1.14)$$

Gains solaires

Ces gains représentent la chaleur obtenue par transformation du rayonnement solaire entré par les menuiseries ou capté par des systèmes spécifiques passifs ne faisant pas partie de l'installation de chauffage. Dans le bâtiment, les surfaces de captage utilisées pour le chauffage solaire passif sont :

- ⇒ Les surfaces transparentes des fenêtres et portes
- ⇒ Le sol et les murs des serres et vérandas
- ⇒ Les parties opaques (façades, toiture), surtout si elles sont recouvertes d'une isolation transparente ou d'un vitrage (murs Trombe)

Les apports solaires dépendent de l'ensoleillement, de l'orientation des surfaces réceptrices, de l'ombrage permanent et des caractéristiques de transmission et d'absorption solaires des surfaces réceptrices. Pour une période de calcul donnée, les apports solaires s'obtiennent en sommant les apports de chaque surface de captage :

$$Q_s = \sum_j Q_{sj} \quad (1.15)$$

$$Q_s = \sum_j Q_{sj} = \sum_j I_{sj} \sum_n A_{snj} \quad (1.16)$$

où la première somme s'effectue sur toutes les orientations j , et la seconde sur toutes les surfaces n qui captent le rayonnement solaire, avec,

- ⇒ I_{sj} : Irradiance solaire, à savoir l'énergie totale, sur la période de calcul, du rayonnement solaire global incident sur une surface unitaire ayant l'orientation j
- ⇒ A_{snj} : Aire réceptrice équivalente de la surface n ayant l'orientation j

1.4.3 Modèles simplifiés

Modèles statiques

Ils étudient le comportement du bâtiment en régime permanent : la température intérieure est égale à la température de consigne et les sollicitations sont constantes dans le temps. Il y a équilibre thermique entre les échanges avec l'intérieur du bâtiment apportés par le chauffage et l'extérieur. Les modèles statistiques expriment la puissance appelée par le chauffage en fonction des données météorologiques donnés par l'équation du bilan thermique.

Dans une optique de simplification due à un manque de données, l'approximation de l'équilibre thermique consiste à considérer tous les apports calorifiques autres que le chauffage comme constants. L'équation se résume dans ce cas à :

$$Q t(i) = a + b (L (T_{int} - T_{ext}) t(i) \quad (1.17)$$

L'équation peut être simplifiée en considérant une température moyenne intérieure constante :

$$Q t(i) = \alpha + \beta T_{ext} t(i) \quad (1.18)$$

Les coefficients α et β sont estimés par régression linéaire à partir d'un nuage de points correspondant aux données journalières, hebdomadaires ou mensuelles collectées au niveau du bâtiment. Il est possible d'enrichir ce modèle en tenant compte des flux solaires donnés par la droite de régression linéaire :

$$Q t(i) = a + bT_{ext} + cI \quad (1.19)$$

Avec

- ⇒ Q : Puissance du système de chauffage,
- ⇒ I : Ensoleillement.
- ⇒ Les coefficients a, b et c sont calculés par la méthode de régression aux moindres carrés.

Modèles dynamiques

Ils permettent de prendre en compte le régime transitoire du bâtiment en fonction des sollicitations. Ils permettent à la fois de mieux caractériser le comportement énergétique du bâtiment et d'insérer l'humain à l'intérieur de ce bâtiment. Deux familles se distinguent : les modèles de signature énergétique et les modèles physiques.

Les modèles de signature énergétique introduisent la réponse du bâtiment aux différentes sollicitations

D'autres études utilisent des modèles réduits pour simplifier les modèles thermiques détaillés. Le bâtiment est modélisé par un circuit électrique équivalent simple (RC, R2C2, R3C2, R7C2) [JAZ-2002] [ARB-2013] pour modéliser le besoin en chauffage et refroidissement.

Pour les **modèles thermiques détaillés**, il s'agit d'étudier physiquement le comportement du bâtiment. Cette modélisation se base sur une connaissance parfaite du système et des différents scénarios d'utilisation et d'excitation. Un bilan énergétique est calculé heure par heure. Pour faire cette simulation, il faut rentrer dans le modèle les informations relatives à la physique du bâtiment :

- ⇒ le type de bâtiment (habitation, commerce, industrie)
- ⇒ le lieu géographique
- ⇒ l'orientation
- ⇒ les masques
- ⇒ la surface
- ⇒ les éléments de constructions
- ⇒ les menuiseries
- ⇒ les apports internes (personnes et équipements)
- ⇒ la perméabilité à l'air
- ⇒ les températures minimale et maximale de consigne (si besoin)
- ⇒ la limite maximale de puissance

Afin de pouvoir étudier quelle est la part de contribution énergétique des apports internes (occupants, puissance dissipée par les équipements électriques et solaires) sur une plus longue période

que précédemment, il est nécessaire de faire pour cela 5 simulations distinctes [IZA-2006]. Ces 5 simulations correspondent à un plan d'expérience où l'on appliquera le théorème de superposition. La méthode consiste à faire une simulation indépendante de chaque apport interne et d'étudier leur contribution par rapport à une simulation complète (Tableau 2).

Scénarios	Consigne de chauffage	Apports solaires	Apports occupants	Apports équipements électriques
Simulations sans apports	√			
Simulations apports solaires		√		
Simulations apports occupants			√	
Simulations apports électriques				√
Simulations tout apports	√	√	√	√

Tableau 2 : Plan d'expérience réalisé sur les différentes simulations de chauffage [IZA-2006]

Plusieurs logiciels permettent de faire ce type de bilan (bSol [BSO-2015], Pleiades COMFIE [PLE-2015], Objective-a [OBJ-2015]) en utilisant des schémas numériques stationnaires ou instationnaires. Contenu de la contrainte prédictive de notre étude, nous utiliserons les simulations thermiques dynamiques. En l'occurrence, notre choix s'est porté sur bSol et Pleiades COMFIE pour avoir des simulations horaires.

1.5 Le besoin en eau chaude sanitaire

Les sources d'énergie utilisées pour la production d'eau chaude sont la production instantanée par combustion (gaz, fioul domestique) avec une régulation de la température d'eau chaude contrôlée par le débit de gaz ou de production à accumulation thermique. Il s'agit alors d'un thermostat qui commande le fonctionnement du générateur. La consommation d'énergie pour l'eau chaude sanitaire est calculée à partir des besoins en énergie nécessaires pour fournir l'eau chaude au point de puisage divisés par le rendement du système. Les variables qui influencent le besoin en eau chaude sanitaire sont la fréquence d'utilisation en particulier dans l'habitat résidentiel et la température de consigne (généralement de 40°C pour un lavabo et de 60°C pour une machine à laver). Pour simplifier, certaines méthodes utilisent des valeurs basées sur des études statistiques qui dépendent du nombre d'occupants et de la surface de la zone chauffée [SAVE-2000].

Equipements par logement	Besoin énergétique (kWh/Jour)
1 évier	$1.3 + 0.3 * n$
1 évier + 1 lavabo	$2.5 + 0.9 * n$
1 évier + 1 lavabo + 1 douche	$3.5 + 0.9 * n$
1 évier + 1 lavabo + 1 petite baignoire	$2.0 + 1.3 * n$
1 évier + 1 lavabo + 1 grande baignoire	$3.8 + 1.3 * n$
1 évier + 1 lavabo + 1 grande baignoire +	$3.8 + 1.8 * n$

2^e cabinet de toilette

Tableau 3 : Besoin en eau chaude sanitaire selon le nombre d'équipements associé à un nombre de personnes n [ANG-2004]

1.6 Les usages spécifiques électriques

Nous nous attacherons à définir les appareils qui représentent la plus grande part dans la consommation énergétique du système. Dans le secteur résidentiel, le chauffage et l'eau chaude, en particulier sur les bâtiments les plus récents, représentent les plus gros consommateurs d'énergie.

L'objectif à partir de mesures au niveau du compteur global est de fournir **un outil de suivi de la consommation par usage** pour le consommateur final. Il permet également d'**augmenter la précision de la consommation électrique** en détectant par exemple les habitudes de consommation (par exemple, allumage de l'électroménager et du chauffage la nuit).

De plus, si le système d'étude permet un effacement du chauffage et de l'eau chaude, cela permet de **limiter le nombre de données en transit** et stockées. Seules des informations de pilotage de type ON-OFF seraient envoyées.

Les premiers travaux pour l'identification et la classification des équipements remontent aux années 80 par [HAR-1992]. Depuis, de nombreux papiers, thèses et brevets sont venus s'ajouter à ces travaux, montrant un grand intérêt à la fois de la communauté scientifique mais également des producteurs et fournisseurs d'électricité. Les principaux travaux menés sur la caractérisation des composantes constitutives de la consommation générale d'une installation électrique se font à partir de la courbe de charge générale (courant / tension / puissance).

Nous pouvons citer les travaux menés aux Etats-Unis (MIT (Massachusetts Institute of Technology), Université de Californie), en France (EDF-R & D) et au Canada (Université de Concordia). Les méthodes proposées se basent sur les travaux menés par George Hart dans les années 1980 qui a proposé certaines techniques de décodage en s'appuyant sur une modélisation par machines à états finis. Chaque appareil est modélisé par une machine à états, une recherche du plus court chemin permettant d'effectuer le décodage des événements. Pour classifier ces différents usages électriques, le travail du M.I.T [HAR-1992] propose une classification en fonction du profil de la courbe de charge individuelle de chaque appareil. Cette classification est la suivante :

- ⇒ Les usages à **consommation permanente** avec une puissance constante, 24 heures sur 24 et 7 jours sur 7 (Alarme incendie, Modem, Téléphone fixe).
- ⇒ Les **usages ON/OFF** : cette catégorie est constituée d'appareils possédant deux états seulement. Un état OFF où ils ne consomment pas, un état ON où ils consomment une puissance approximativement constante. De nombreux appareils courants dans le secteur résidentiel appartiennent à cette catégorie (bouilloire, ampoule électrique, réfrigérateur, congélateur, grille-pain, machine à café, micro-ondes, four, chauffage, eau chaude sanitaire).
- ⇒ Les usages à **nombre d'états fini** qui incluent les usages passant par plusieurs états distincts où ils consomment une puissance constante (machine à laver, sèche-linge, lave-vaisselle).

- ⇒ Les usages à **nombre d'états infini** : cette catégorie inclut les usages dont la puissance consommée peut varier continûment, en général en fonction d'un réglage utilisateur (les aspirateurs à puissance variable).

La méthode développée au M.I.T utilise la puissance active et la puissance réactive collectées à la seconde au niveau d'un point de mesure (Figure 23). La seconde est choisie pour limiter le risque de perte d'évènements liés à l'enclenchement ou l'arrêt de machine en simultanément. Cette méthode peut se résumer en cinq étapes :

- ⇒ Définition des différents évènements à partir de la puissance active
- ⇒ Classification à partir de deux variables : la puissance active et la puissance réactive
- ⇒ Regroupement des évènements ON/OFF par symétrie dans le plan puissance active/puissance réactive
- ⇒ Réduction du nombre d'évènements non classifiés
- ⇒ Association d'un nom à chaque usage identifié

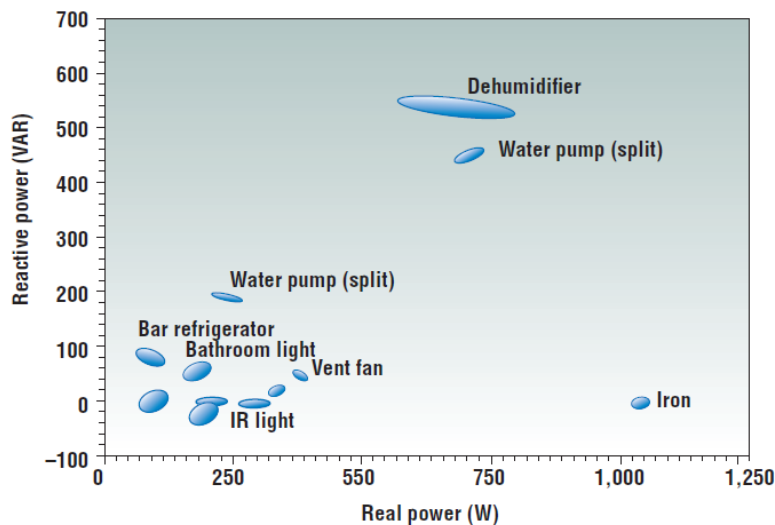


Figure 23 : Représentation des évènements détectés dans le plan Puissance Active en abscisse, Puissance Réactive en ordonnée pour une maison à gauche, deux maisons à droite, Etats Unis, [HAR-1992]

Dans [MAR-2000], l'auteur propose d'utiliser des connaissances a priori sur les durées moyenne et maximale de fonctionnement de chaque usage en combinaison avec un algorithme de détection de changements dans la puissance active ce qui permet de diminuer le taux d'erreur de l'algorithme de classification. Les travaux de [BAR-2004] proposent un algorithme identifiant les changements les plus fréquents puis construisant des machines à états finis hypothétiques à partir de ces changements. Cet algorithme présente des points communs avec celui du M.I.T, mais traite l'ensemble du problème de manière globale, ce qui permet de détecter conjointement les évènements ON et OFF d'un même

appareil, ou d'ignorer les changements peu fréquents pour se focaliser sur les contributions les plus importantes à la courbe de charge totale.

Dans **[GUE-2009]**, une étude par charges est réalisée à partir de données à la seconde et permet de soutirer des informations a priori sur les appareils comme les puissances minimales ou maximales par appareils. Dernièrement, plusieurs auteurs se sont intéressés à la modélisation de courbes de charge par des chaînes de Markov cachées **[PAR-2012]**, **[KIM-2011]**.

Cependant, la généralisation du problème d'identification des appareils est complexe puisque les appareils et les habitudes de consommation diffèrent entre les maisons. Plusieurs campagnes de mesures basées sur des centaines de logements ont mis en lumière des facteurs allant de 1 à 20 pour la consommation énergétique d'un même appareil dans des logements différents **[ENE-2004]** **[ADE-2010]**.

Si nous voulons fournir un outil non intrusif de désagrégation de courbe de charges, il nous faudra créer nos bases de données d'apprentissage d'appareils électriques. Si nous sommes à une fréquence d'échantillonnage d'une seconde, l'utilisation de la puissance active et réactive permet d'augmenter significativement la précision d'identification **[DUF-2015]**.

Pour des raisons de données sur la puissance active et réactive, nous nous intéresserons dans cette étude à l'identification du chauffage et de l'eau chaude qui représentent à la fois le plus gros potentiel de flexibilité et la plus grande part de consommation dans le secteur résidentiel. Ces usages énergétiques sont fournis par une pompe à chaleur dans notre maison test. Il nous faudra ainsi créer une base de données pour cet appareil en collectant les puissances active et réactive à la seconde.

Pour permettre l'identification des appareils électriques avec comme point de mesure le compteur global, nous nous sommes basés sur nos propres campagnes de mesure et sur la base de données collectées par la HES-SO Fribourg **[RID-2015]**.

1.7 Production local renouvelable

Energie solaire

Un des exemples d'utilisation de l'énergie solaire pour chauffer un bâtiment est d'utiliser des capteurs passifs. On va placer une vitre devant les murs pour chauffer l'air entre la vitre et le mur par convection naturelle (technique présente principalement aux Etats-Unis). Nous pouvons nommer les murs TROMBE **[TRO-2015]** (Figure 24).



Figure 24 : Types d'installations différentes de récupération du rayonnement solaire pour produire de la chaleur [TRO-2015]

Pour cette technologie, le marché est dominé à plus de 70% par la Chine. Les Etats-Unis, l'Allemagne et la Turquie avec chacune entre 3 et 5% de la part du marché global de l'électricité solaire pour chauffer le bâtiment. Au niveau mondial et dopé par la Chine, nous pouvons noter une augmentation de 400 % de la capacité thermique utilisée sur les 10 dernières années d'après l'agence internationale de l'énergie.

Dans un second point, nous pouvons citer les installations photovoltaïques qui sont majoritairement **installées sur le réseau basse tension**. Elles transforment le rayonnement solaire en électricité, soit en courant continu. Ce dernier est transformé en courant alternatif grâce à un onduleur, puis injecté dans le réseau public. Le photovoltaïque (PV) est une technologie-clé pour assurer un approvisionnement énergétique durable.

Aujourd'hui nous pouvons considérer les capteurs photovoltaïques (électricité) comme une technologie mature. Une des innovations aujourd'hui est portée par le couplage entre l'électricité et la chaleur avec des panneaux photovoltaïques et thermiques (Figure 25). Les capteurs peuvent être fixes ou suivre la course du soleil. Les cellules solaires sont refroidies par un fluide.

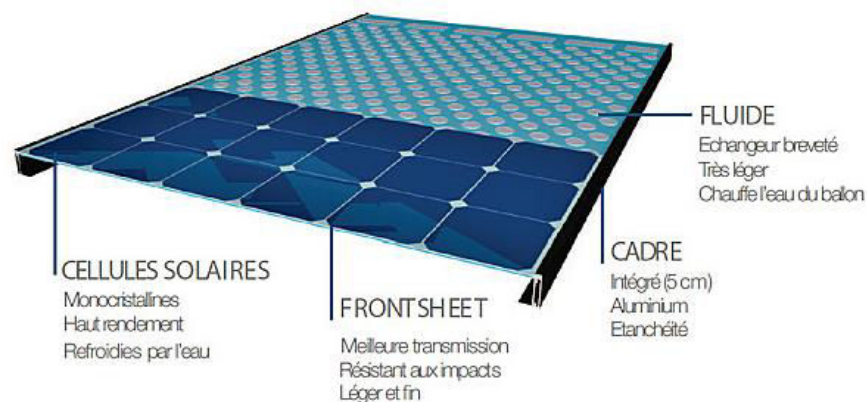


Figure 25 : Nouveau type de capteurs solaires qui intègre l'électricité et la chaleur [DUA-2014]

Pour les bâtiments individuels performants, la surface de captage est suffisante pour permettre une autonomie énergétique annuelle avec stockage de la chaleur (pour l'eau chaude sanitaire) et de l'électricité (stockage ou réseau électrique).

De nouveaux plans économiques rendent l'autoconsommation plus intéressante avec par exemple la diminution des prix des batteries électriques. Les freins au développement de l'énergie solaire locale restent une gestion optimale locale de l'offre et de la demande, une acquisition et le traitement en temps réel des données de consommations et de productions et des prévisions fiables à court terme.

Pour les réseaux électriques, il ne peut pas actuellement avoir plus de 30 % de sources de productions intermittentes mais il est possible de dépasser ce seuil à partir du stockage local d'après l'agence internationale de l'énergie.

Production d'électricité d'origine éolienne

L'intégration de systèmes éoliens dans le bâtiment est plus difficile que le photovoltaïque à cause de son exigence de régularité sur la vitesse du vent. Pour pallier aux variations de production, les éoliennes sont éloignées géographiquement mais sont présentes sur un même réseau qui, par effet de foisonnement, permettra d'assurer une plus grande régularité dans la production.

Dans le monde, en 2015, l'énergie éolienne a couvert 3% de la consommation d'électricité mondiale totale grâce à une puissance de 432'000 MW. [GWE-2015]. En Europe, fin 2015, la production continentale de l'énergie éolienne, correspondait à environ 11,4% de la consommation d'électricité totale [EWE-2015]. En Suisse, les 34 éoliennes installées totalisent 60 Mégawatts (MW) de puissance électrique et ont produit 110 mio. de kilowattheures (kWh) en 2015. C'est la consommation équivalente à 30'000 ménages suisses ou 0,2 % de la consommation d'électricité totale. En France on dénombrait 10'400 MW, en Allemagne 45'000 MW et en Italie 9'000 MW.

Le point faible de l'énergie solaire et de l'énergie éolienne est leur caractère intermittent et leur dépendance aux éléments météorologiques. Cela rend difficile leurs utilisations sans autre source d'énergie complémentaire ou sans un système de stockage, à moins d'avoir recours à un système de pilotage local innovant.

1.8 Stockage local de l'électricité

Les matériaux utilisés pour la construction du bâtiment possèdent une capacité thermique nommé **inertie thermique**. Le terme du stockage passif est introduit dans [HEN-2007] pour désigner la capacité d'accumuler de l'énergie sous forme thermique dans l'enveloppe du bâtiment. Les termes passifs et actifs sont également introduits. Le stockage thermique actif se trouve dans le système de production de l'eau chaude sanitaire ou le stockage de froid dans les systèmes de climatisation [HEN-2007]. Les bâtiments peuvent être ainsi séparés par leur inertie thermique. Cette inertie doit être prise en compte dans la mise en œuvre d'une régulation. Elle constitue une caractéristique physique du bâtiment.

Les batteries permettent le stockage de l'électricité pour les systèmes isolés et servent de conditionneur de puissance. La batterie peut permettre également de restituer la nuit le courant accumulé durant la journée. Par contre, les accumulateurs électriques sont généralement chers,

encombrants (environ 30kg pour 1kWh stocké) avec une durée de vie et un nombre de cycles associés encore faibles et posent des problèmes de pollution (métaux lourds) lors de leur fin de vie.

Nous pouvons cependant citer les nouvelles technologies de batterie sur la base de module Lithium Titanate Leclanché SRB1021 [LEC-2015] et les différentes applications réseau qui peuvent être réalisées et combinées avec une électronique de puissance pour contrôler les flux d'énergie.

1.9 Conclusion

Nous souhaitons un système de prédiction local. Pour réaliser un bilan, nous devons calculer les différents flux énergétiques à l'intérieur de notre système :

- ⇒ Déperditions par l'enveloppe et la ventilation
- ⇒ Gains internes solaires, des personnes et des appareils
- ⇒ Stockage
- ⇒ Production d'ECS
- ⇒ Usages spécifiques électriques utiles

Sur certains de ces points, nous pouvons évaluer assez précisément et en fonction du temps les quantités d'énergie échangées. Pour les autres (ECS, USE, gains internes, stockage), la bibliographie nous fournit des méthodes globales et indépendantes du temps. Il n'est donc pas possible à ce stade d'envisager une méthode correspondant au pas de temps souhaité.

Ceci impose la mise au point d'une méthode prédictive et apprenante dont nos modèles de simulation énergétique seront le point de référence. Il n'en reste pas moins que ces modèles permettent la compréhension du comportement énergétique du système.

Chapitre 2 : Méthodes de prédiction des flux énergétiques

2.1 Introduction

Notre outil doit être capable de fournir des prédictions et des classifications de chaque flux énergétique. L'objectif final est le pilotage des flux énergétiques et la stabilisation du réseau électrique au niveau local. Cela demande une sûreté dans la qualité de la prédiction des différents flux énergétiques d'autant plus au niveau d'un système tel qu'une maison

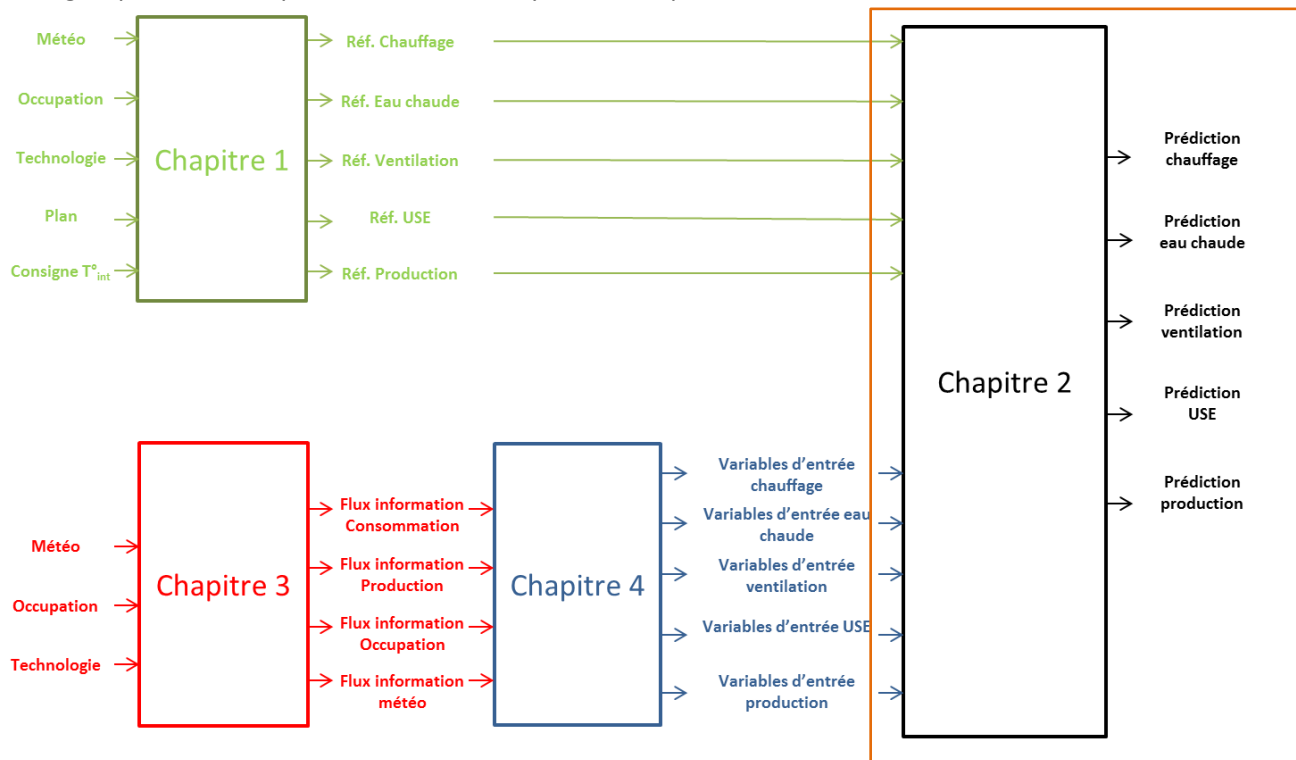


Figure 26 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision

Comme décrit dans le chapitre 3 pour la définition des acteurs, dans certains cas, l'objectif n'est pas d'avoir seulement une information la plus précise en termes de prédictions énergétiques mais d'avoir par exemple un ensemble de charges à éteindre à un instant donné.

Les différents flux énergétiques seront caractérisés par des classes représentées par une gaussienne qu'il s'agira de prédire pour l'heure suivante.

Par exemple, deux classes allumées et éteintes suffiraient à caractériser l'eau chaude et le chauffage pour l'heure suivante ou encore une caractérisation précise du pic de puissance liée à la consommation électrique si l'entreprise connectée est soumise à ce système de facturation. Une étude spécifique du pic de puissance lié à la production permettrait également de prévenir les moments les moins fréquents mais les plus importants pour la stabilisation du micro-réseau.

Si la classe est connue au préalable et que l'opération de classement consiste à analyser les caractéristiques des individus pour les placer dans une classe, la méthode est dite «supervisée». Dans le cas contraire, la méthode est «non-supervisée», ce vocabulaire étant issu de l'apprentissage automatique tiré des modèles non linéaires.

2.2 Apprendre du passé pour prédire l'avenir

La consommation d'électricité est historiquement enregistrée à l'aide de compteurs électromécaniques à simple ou double tarif. La lecture des compteurs s'effectue chez le consommateur, une ou deux fois par année.

Aujourd'hui, les compteurs numériques sont capables de relever la consommation, la production ainsi que tout paramètre lié à l'état du réseau à chaque seconde près. En France, 35000 compteurs intelligents seront installés d'ici 2021.

En Suisse romande, le projet Green E-Value [GRE-2014] propose un portail accessible aux propriétaires des immeubles et aux locataires qui leur fournit les informations de consommation et propose des conseils d'économie d'énergie. Dans les locaux communs, des écrans indiquent les consommations moyennes des logements et permettent aux locataires de s'autoévaluer et d'améliorer leur comportement.

Ainsi les données relatives à la consommation et à la production sont ou seront collectées dans quelques années sur des millions de logements. Associées à des bases de données météorologiques déjà existantes depuis des dizaines d'années comme par Météo France ou Météo Suisse, il sera possible de créer des bases de données conséquentes relatives à un système comme une maison. Ces bases de données créées dans le cadre de cette thèse sont développées dans le chapitre 3. Cela permet à partir d'extraction de connaissances développées dans le chapitre 4 de modéliser sur plusieurs mois ou sur plusieurs années le comportement énergétique d'un système. Dans notre cas, une modélisation est propre à chaque flux et un modèle spécifique est créé pour modéliser le besoin en chauffage, en eau chaude sanitaire, en USE et la production solaire.

Pour chacun de ces flux, une base de données est créée et une partie de ces données est utilisée pour **entraîner un modèle** linéaire ou non linéaire. Une autre partie de ces données est utilisée pour **tester le modèle** entraîné. Il est possible également de rapprendre les modèles au bout d'une période prédéfinie.

2.3 Méthodes linéaires

L'une des méthodes les plus classiques est une **régression linéaire**. Avec y la variable aléatoire réelle à expliquer (variable endogène, dépendante ou réponse) et x la variable explicative ou effet fixe (exogène). Elle est toujours utilisée pour des prédictions de la consommation d'énergie totale [ZAM-2013].

Nous avons :

$$y = f(x) = \beta_0 + \beta_1 x + \varepsilon \quad (2.1)$$

En statistique, les modèles ARMA (modèles autorégressifs et moyenne mobile), ou aussi modèles de Box-Jenkins, sont les principaux modèles de séries temporelles. Les modèles ARIMA permettent de traiter les séries non stationnaires après avoir déterminé le niveau d'intégration. Le modèle ARIMA ou ARMA souffre d'une lacune majeure puisqu'il est incapable de traiter plusieurs variables.

Pour contourner ce problème, il faut pouvoir généraliser le modèle ARIMA dans le cas à plusieurs variables décrit dans [SIM-1980] nommé VAR pour (Vector Auto Regressive). Mais, contrairement au modèle structurel à plusieurs variables, dans les modèles VAR, toutes les variables sont endogènes. Ces modèles (ARIMA et VAR) ne permettent de traiter que des phénomènes qui sont linéaires ou approximativement (par exemple le PIB) mais ne permettent pas d'identifier les propriétés des phénomènes qui sont non linéaires.

2.4 Méthodes non linéaires

2.4.1 Réseau de neurones

Un réseau de neurones artificiels est un ensemble d'algorithmes dont la conception est à l'origine très schématiquement inspirée du fonctionnement des neurones biologiques. Nous retrouvons plusieurs types de réseau de neurones qui ont chacun leurs spécificités, avantages et inconvénients.

Un réseau de neurones est en général composé d'une succession de couches dont chacune prend ses entrées sur les sorties de la précédente. À chaque synapse est associé un poids synaptique, de sorte que les neurones de la couche précédente sont multipliés par ce poids, puis additionnés par les neurones de niveau i . Le neurone reçoit des neurones en amont un certain nombre de valeurs via ses connexions synaptiques, et il produit une certaine valeur en utilisant une fonction d'activation appelée aussi fonction de combinaison. Cette fonction peut être formalisée comme étant une fonction vecteur-à-scalaire, notamment :

- ⇒ Les réseaux de type **MLP** (multi-layer perceptron) calculent une combinaison linéaire des entrées, c'est-à-dire que la fonction de combinaison renvoie le produit scalaire entre le vecteur des entrées et le vecteur des poids synaptiques.

- ⇒ Les réseaux de type **RBF** (radial basis function) calculent la distance entre les entrées, c'est-à-dire que la fonction de combinaison renvoie la norme euclidienne du vecteur issu de la différence vectorielle entre les vecteurs d'entrées.

Nous pouvons citer les plus utilisés dans le domaine de la prédiction énergétique comme la famille des ANN (Artificial Neural Network) avec l'un des plus polaires, le **MLP (Multi Layer Perceptron)** [BER-2010].

Dans le **MLP**, les neurones d'une couche sont reliés à la totalité des neurones des couches adjacentes. Ces liaisons sont soumises à un coefficient altérant l'effet de l'information sur le neurone de destination (Figure 27). Le poids de chacune de ces liaisons est l'élément clef du fonctionnement du réseau : la mise en place d'un Perceptron multicouche pour résoudre un problème passe donc par la détermination des meilleurs poids applicables à chacune des connexions inter-neurales [ONO-2000] [BER-2010].

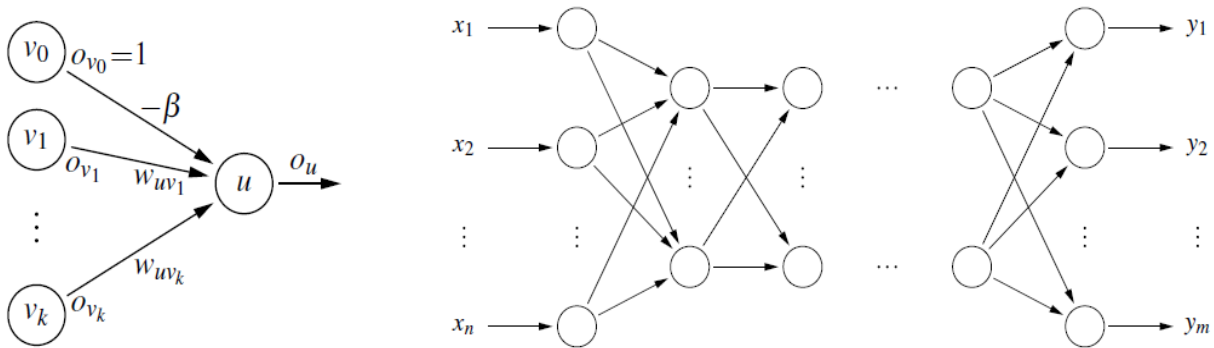


Figure 27 : Représentation d'un réseau de neurones [BER-2010]

Le nombre de couches, le nombre de neurones dans chaque couche, ainsi que d'autres paramètres, déterminent ce que l'on nomme la « structure » du réseau qui définissent nos paramètres de configuration :

- ⇒ **Le nombre de couches cachées**
- ⇒ **Le nombre de neurones par couche cachée**
- ⇒ **Le nombre maximum d'itérations pour l'apprentissage**

Au sein de cette structure, chaque neurone est connecté à l'ensemble des neurones des deux couches voisines (supérieure et inférieure). Ces connexions entre neurones sont dotées de poids qui sont déterminés lors de la phase d'entraînement, phase dans laquelle nous présentons au réseau des données d'entrée, telles que des statistiques sur la consommation électrique, des données météorologiques, météo, ainsi que la sortie désirée, telle qu'une valeur de prédiction horaire. Une fois la phase d'entraînement terminée et les poids déterminés, le réseau de neurones est prêt pour la phase de prédiction basée sur les données de test.

Dans la famille des réseaux de neurones, nous trouvons également le PNN (Perceptron Neural Network) ou encore le SVM (Support Vector Machine).

Concernant les PNN, cet algorithme génère des règles basées sur des données numériques. Chaque règle est définie comme une fonction gaussienne de grande dimension qui est ajustée par deux seuils, **theta moins et theta plus**, qui décrivent l'aire de conflit (Figure 28). Chaque fonction gaussienne est définie par un vecteur central (à partir de la première instance couverte) et un écart type qui est ajusté pendant la formation pour couvrir uniquement les instances non conflictuelles.

L'une des forces majeurs des réseaux de neurones est la technique de **back propagation** qui est fondamentale pour l'entraînement et est le secret du « deep learning » : depuis la couche de sortie, l'erreur d'estimation qui est faite durant l'entraînement est propagée en arrière dans chaque couche de neurone et corrige les poids des connections pour minimiser l'erreur.

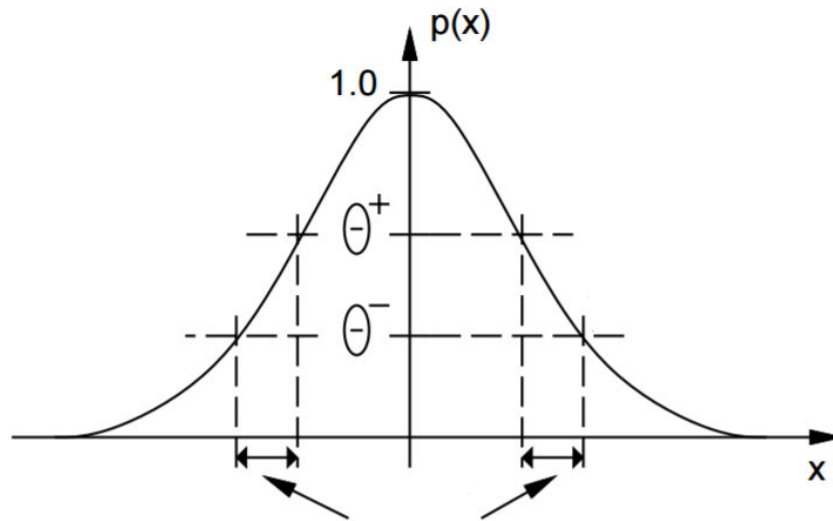


Figure 28: Paramètres de configuration de l'algorithme PNN qui correspondent à la délimitation de la zone de conflit [BER-2010]

Les Support Vector Machine nommé **SVM** sont des classificateurs qui permettent de traiter des problèmes de discrimination non linéaire [BER-2010]. Le cas simple est le cas d'une fonction discriminante linéaire, obtenue par combinaison linéaire du vecteur d'entrée $x = (x_1, \dots, x_n)^T$, avec un vecteur de poids $w = (w_1, \dots, w_n)^T$

$$h(x) = w^T x + w_0 \quad (2.2)$$

Il est introduit dans les SVM la notion de marge maximale qui est la distance entre la frontière de séparation et les échantillons les plus proches (Figure 29). Ces derniers sont appelés des vecteurs supports. La frontière de séparation est choisie comme celle qui maximise la marge.

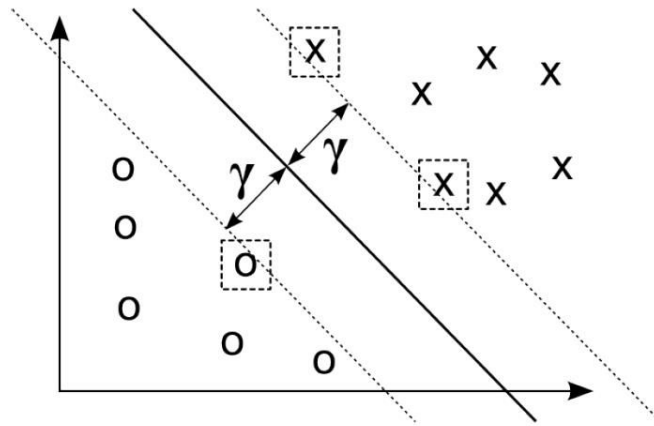


Figure 29 : Représentation de la notion de marge γ dans le cas de séparation de deux classes [BER-2010]

Pour résoudre des problèmes non linéaires, l'idée des SVM est de reconsidérer le problème dans un espace de dimension supérieure, éventuellement de dimension infinie. Il est appliqué aux vecteurs d'entrée une transformation non-linéaire appelé fonction noyau. Plusieurs **fonctions noyaux** sont intégrées dans des outils de prédictions comme les **RBF, Hyper tangentes ou polynomiales** décrites dans [BER-2010].

La fonction noyau RBF pour Radial Basic Fonction a été retenue pour l'ensemble de nos classifications qui nous a fourni nos meilleurs résultats en test.

2.4.2 Les arbres de décision

Les arbres de décision sont l'une des structures de données majeures de l'apprentissage statistique. Leur structure arborescente les rend **lisibles par un être humain**, contrairement à d'autres approches où le prédicteur construit est une « boîte noire » comme les réseaux de neurones. Cette faculté de lecture est possible puisqu'un arbre de décision fait une sélection explicite des paramètres qu'il utilise en exploitant un critère de quantité d'information [CAR-2010].

Il modélise une hiérarchie de tests sur les valeurs d'un ensemble de variables appelées attributs. À l'issue de ces tests, le prédicteur produit une valeur numérique ou choisit un élément dans un ensemble discret de conclusions. On parle de régression dans le premier cas et de classification dans le second [BRE-2001].

L'apprentissage d'un arbre de décision se fait sur un ensemble d'instances $T = \{(x, y)\}$ appelé « ensemble d'entraînement ». Un ensemble de valeurs pour les différents attributs est appelé une « instance ».

Si toutes ou la majorité des instances ont la même classe $c \in \{1, \dots, C\}$, c apparaît comme la meilleure prédiction possible par vote majoritaire. Un arbre de classification est d'autant meilleur que les instances de ses feuilles sont de classes homogènes. Une mesure de l'hétérogénéité d'une feuille est

calculée selon différentes méthodes comme le critère de Gini [BER-2010]. Il s'agit du taux d'erreur sur l'ensemble d'entraînement qui retourne la classe c avec une probabilité p_c (au lieu de toujours retourner la classe c^*). La mesure d'erreur est utilisée de même pour choisir le meilleur attribut de partage. De manière heuristique, l'attribut qui minimise cette erreur est considéré comme le meilleur choix possible.

Un arbre de décision est composé de nœuds de décision, qui permettent de tester les attributs. Un nœud de décision est étiqueté par un test qui peut être appliqué à chaque description d'un individu ou d'une population. Il est composé de branches qui représentent chacune une valeur de l'attribut testé. Il peut également être constitué de feuilles, c'est-à-dire les nœuds terminaux, qui indiquent la classe résultante. Les paramètres de configuration d'un arbre de décision sont :

- ⇒ **Critère de fractionnement (exemple : indice de Gini)**
- ⇒ **Nombre limite de niveaux (profondeur de l'arbre)**
- ⇒ **Taille minimale du nœud**

À toute description est associée une seule feuille de l'arbre de décision. Les arbres de décision tels qu'implémentés dans notre outil sont décrits dans [SHA-2010].

2.4.3 Extensions liées au boosting et bagging

Technique du « bagging »

Cela consiste à **sous-échantillonner** ou ré-échantillonner au hasard avec doublons les données d'entraînement et de faire générer à l'algorithme **un modèle pour chaque sous-échantillon**. Un ensemble de modèles est obtenu qu'il convient de **moyenner** lorsqu'il s'agit d'une régression ou de **faire voter** pour une classification.

Cette technique est utilisée particulièrement dans les arbres de décision tels que nous venons de les aborder qui présentent plusieurs limitations comme la sensibilité au bruit ou au sur-apprentissage. Proposées en 2002 par Leo BREIMAN et Adele CUTLER [BRE-2001], les forêts aléatoires utilisent cette technique pour tirer de manière aléatoire un ensemble de données dans le set d'entraînement et recopie des données sous représentés.

Un échantillon d'un ensemble nommé T est l'ensemble obtenu en tirant $|T|$ fois des éléments de T uniformément au hasard et avec remise. Généralement, cet échantillon représente en moyenne $1 - e^{-1} \approx 63\%$ instances uniques différentes de T quand $|T| \gg 1$.

Pour l'agrégation, plusieurs échantillons sont créés et chaque échantillon T_i est utilisé pour entraîner un arbre. Étant donnée une instance (\mathbf{x}, y) , une régression est réalisée au niveau de chaque arbre, ce qui nous donne un ensemble de valeurs y_1, \dots, y_m prédites. Celles-ci sont alors agrégées en calculant leur moyenne.

Technique du « boosting »

Le boosting est une technique d'ensemble qui consiste à agréger des modèles mathématiques élaborés de manière séquentielle sur les données d'apprentissage. Les poids des individus sont corrigés au fur et à mesure. Une pondération est réalisée au niveau des modèles selon leurs performances. Ces techniques sont décrites dans [FRE-1999] [BRE-1999] [FRE-2001].

Cette technique est utilisée dans des algorithmes encore peu utilisés dans le domaine de la prédiction énergétique comme l'algorithme nommé **AdaBoost**. C'est un algorithme de boosting qui s'appuie sur ce principe, avec un paramètre de mise à jour adaptatif permettant de donner plus d'importance aux valeurs difficiles à prédire, donc en boostant les classifieurs qui réussissent quand d'autres ont échoué.

Nous retrouvons cette notion également dans le **gradient boosting** ou descente du gradient qui est une technique itérative qui permet d'approcher la solution d'un problème d'optimisation. En apprentissage supervisé, la construction du modèle revient souvent à déterminer les paramètres (du modèle) qui permettent d'optimiser (max ou min) une fonction objective. L'idée est d'agréger plusieurs modèles ensembles mais en les créant itérativement. Elle est employée majoritairement avec des arbres de décision nommé alors **Gradient Tree Boosting** [FRE-1999].

Intégrés dans les arbres de décision, les paramètres de configuration des arbres tels que les **Random forest** et le **Gradient boosting tree** décrits par Leo Breiman et Adele Cutler sont :

- ⇒ Critère de fractionnement (indice de Gini)
- ⇒ Nombre limite de niveaux (profondeur de l'arbre)
- ⇒ Taille minimale du noeud
- ⇒ **Nombre de modèles**
- ⇒ Échantillonnage des données

2.5 Calcul de l'erreur

2.5.1 Courbe ROC dans le cas de problème binaire

La courbe ROC (de l'anglais Receiver Operating Characteristic) est une mesure de la performance d'un classificateur binaire. Graphiquement, on représente souvent la mesure ROC sous la forme d'une courbe qui donne le taux de vrais positifs (fraction des positifs qui sont effectivement détectés) en fonction du taux de faux positifs (fraction des négatifs qui sont détectés incorrectement). Nous étudions alors un ensemble de valeurs seuil possibles et, pour chacune, on calcule différentes statistiques dont les plus simples sont :

- ⇒ Vrais positifs (VP) : nombre d'heures classées positivement par le test et qui le sont effectivement.
- ⇒ Faux positifs (FP) : nombre d'heures classées positivement par le test mais qui sont en réalité négatives.

- ⇒ Vrais négatifs (VN) : nombre d'heures classées négativement par le test et qui le sont effectivement.
- ⇒ Faux négatifs (FN) : nombre d'heures classées négativement par le test mais qui sont en réalité positives.
- ⇒ Prévalence de l'évènement : fréquence de survenance de l'évènement dans l'échantillon total $(VP+FN)/N$.

Plusieurs indices synthétiques ont été mis au point afin d'évaluer la performance du test à une valeur seuil donnée :

La sensibilité (aussi appelée fraction de Vrais Positifs) est la proportion d'heures positives effectivement bien détectées par le test. Le test est parfait lorsque la sensibilité vaut 1, équivalent à un tirage au hasard lorsque la sensibilité vaut 0.5. S'il est inférieur à 0.5, le test est contre-performant et on aurait intérêt à inverser la règle pour qu'il soit supérieur à 0.5 (à condition que cela n'affecte pas la spécificité).

La définition mathématique est :

$$\text{Sensibilité} = \frac{VP}{(VP+FN)} \quad (2.3)$$

La spécificité (aussi appelée Fraction de Vrais Négatifs) est la proportion d'heures négatives effectivement bien détectées par le test. La définition mathématique est :

$$\text{Spécificité} = \frac{VN}{(VN+FP)} \quad (2.4)$$

On définit également **la fraction de faux positifs (FP)** qui est la proportion de négatifs détectés comme des positifs par le test (1-Spécificité) et **la fraction de faux négatifs (FN)** : proportion de positifs détectés comme des négatifs par le test (1-Sensibilité).

2.5.2 Méthodes statistiques

Nous utilisons également des outils statistiques qui permettent de déterminer la qualité de la prédiction. Nous en avons choisi quatre qui sont retenus dans la plupart des travaux de prédictions :

Le coefficient de détermination (R^2) est une mesure de la qualité de la prédiction d'une régression linéaire. Il est défini comme 1 moins le ratio entre l'erreur avec les valeurs prédites et la variance des données :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (2.5)$$

y_i les valeurs des mesures, \hat{y}_i les valeurs prédites et y_i Lorsqu'il est proche de 0, le pouvoir prédictif du modèle est faible et lorsqu'il est proche de 1, le pouvoir prédictif du modèle est fort.

L'erreur quadratique moyenne (Mean Square Error, MSE) est le carré moyen des erreurs ou erreur quadratique moyenne (MSE pour *Mean Square Error* ou MCE pour moyenne des carrés des erreurs): c'est la moyenne arithmétique des carrés des écarts entre les prévisions et les observations. L'erreur type **RMSE** (root mean square deviation) est la racine carrée de la MSE.

L'erreur moyenne absolue (Mean Absolute Error, MAE) est la moyenne arithmétique des valeurs absolues des écarts.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}_i| \quad (2.6)$$

L'écart quadratique moyen (MSD). Pour calculer l'erreur quadratique moyenne RMS (Root Mean Square), les erreurs individuelles sont tout d'abord élevées au carré, puis additionnées les unes aux autres. On divise ensuite le résultat obtenu par le nombre total d'erreurs individuelles, puis on en prend la racine carrée. Les valeurs aberrantes ont un effet plus important sur le MSD que sur le MAD. L'équation est la suivante :

$$MSD = \sqrt{\frac{\sum_{i=1}^n |y_i - \hat{y}_i|^2}{n}} \quad (2.7)$$

Enfin une erreur en termes de consommation énergétique sera également présentée. Elle est définie par :

$$Erreur = Réel - Prédiction \quad (2.8)$$

2.6 Conclusion

Les paramètres pouvant influencer la prédiction sont tellement nombreux, qu'il est difficile d'utiliser les résultats issus d'autres études. Par exemple, dans [ZHA-2016], la série temporelle de consommation est prédite à partir d'une méthode hybride qui utilise une régression suivie d'une méthode d'apprentissage avec l'algorithme SVM.

Au niveau de la prédiction de production solaire décentralisée, les différentes études concernent un large choix d'installations photovoltaïques et de type de données, (étude sur 12 cellules PV en laboratoire ou de 14 000 m² de panneaux) avec des modèles ARIMA ou des réseaux de neurones [ZAM-2014] [GRA-2016] [GUL-2016]. Basé sur un historique, les données de prédiction météorologiques sont parfois corrigées lors d'une première phase de test afin d'améliorer la prédiction de production des panneaux. Les données proviennent d'une seule station météo locale ou régionale.

Nous notons cependant que les réseaux de neurones sont aujourd'hui les modèles de prédictions les plus utilisés. Les études utilisent des PNN (Probabilistic Neural network), MLP (Multi Layer Perceptron, ANN (Artificial Neural Network), SVM (Support Vector Machine). Chaque algorithme a des paramètres spécifiques que l'on peut faire varier et représente à lui seul un domaine de spécialité à lui

seul. Il n'en reste pas moins que les réseaux de neurones sont considérés comme des boîtes noires difficilement interprétables.

L'arbre de décision est lui beaucoup plus simple et permet une lisibilité par le lecteur. Dans notre cas où plusieurs milliers de variables vont caractériser le vecteur d'entrée, seul, il est mal adapté. C'est pourquoi des techniques plus récentes ont été développées qui permettent de construire des modèles mathématiques adaptés à notre cas d'étude comme les ensembles d'arbres de décision. Les résultats statistiques issus de la construction des arbres permettent de recueillir une connaissance précieuse sur la pertinence d'une variable par rapport à une autre.

Tous les algorithmes cités ne sont pas toujours disponibles dans les différents outils de prédictions. Nous essayerons de choisir également un outil de traitement qui intégrera la majorité des algorithmes cités dans ce chapitre : les méthodes linéaires (régression linéaire et ARIMA) et des méthodes non linéaires (ANN, PNN, MLP, SVM, Random Forest, Adaboost et Gradient Boosted tree).

Cependant, l'intégration des ensembles d'arbres de décision nous semblent être un caractère indispensable pour la compréhension et la réduction de notre vecteur d'entrée.

Chapitre 3 : Mettre en place le système d'information

3.1 Introduction

L'objectif de chapitre est de mettre en place un **système d'information local** au niveau de chaque système qui nous permet de collecter, stocker, traiter et distribuer les connaissances liées aux prédictions à un niveau de pilotage supérieur (Figure 30).

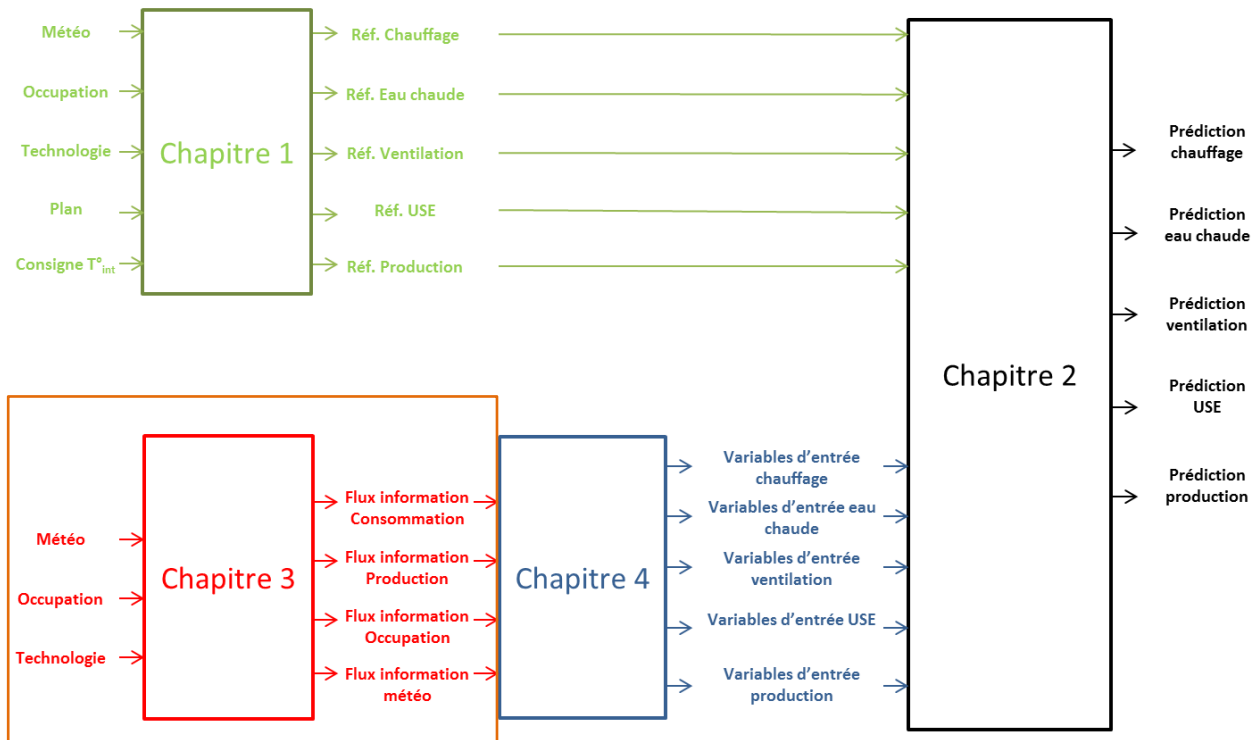


Figure 30 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision

L'étape de définition des modèles de référence permet de réaliser l'évaluation énergétique du système, en particulier au niveau du chauffage, du refroidissement et de la production solaire où des outils de simulations peuvent fournir des prévisions horaires. La partie « traiter » comprend la prédiction des différents flux énergétiques du système utilisant les modèles mathématiques décrits dans le chapitre 2 (Figure 31).

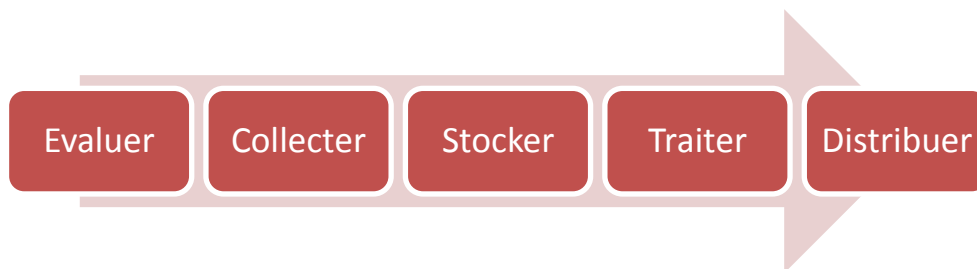


Figure 31 : Processus de mise en place du système d'information pour le pilotage énergétique pour nos systèmes d'étude

Les compteurs intelligents déployés aujourd'hui sont capables de recevoir et transmettre des informations entre le consommateur et le fournisseur (alertes sur la puissance soutirée, prix en temps réel, conseils automatiques concernant le report de consommation). Une couche de communication bidirectionnelle vient alors se superposer au réseau électrique, d'où la notion de réseau intelligent décrit dans le préambule.

3.2 Evaluer

3.2.1 Définition des acteurs

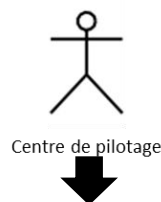
L'outil que nous créons correspond à une boîte noire qui doit fournir des services à son environnement. Par environnement, on entend les **acteurs** qui ont besoin de ce logiciel. Nous distinguons deux types d'acteurs : les acteurs principaux qui vont directement agir sur le système, dans notre cas le centre de pilotage et les acteurs secondaires qui n'ont pas d'accès direct à l'outil mais peuvent recevoir et échanger des informations avec celui-ci.

De plus, dans le langage UML (Unified Modeling Language) [BEN-2015], on appelle ce qu'on doit analyser, concevoir et réaliser : le système. Ici, le **système** est l'outil de prédiction des flux énergétiques qui permet de faire l'interface entre le centre de pilotage et les différentes charges du système.

Dans cette étude, notre outil fait le lien entre le système où il va tirer le maximum de connaissances et le centre de pilotage. La réactivité dans le pilotage de charges peut rendre des services aux systèmes de fournitures d'électricité. Les clients qui possèdent ou non une production décentralisée peuvent ainsi prendre part aux marchés de l'énergie et devenir de nouveaux acteurs côté demande par pilotage de leurs charges par un agrégateur [SUR-2004]. Des clients ou groupes de clients peuvent souscrire à des programmes orientés marché de l'énergie, dans lesquelles ils répondent à des appels tarifaires, en diminuant leur consommation.

De plus, ces programmes offrent la possibilité de participer activement aux échanges commerciaux d'électricité en proposant aux consommateurs de modifier leurs habitudes de consommation en retour de compensations financières. Ces dernières peuvent être un paiement direct de l'électricité non consommée, ou un paiement en fonction d'une garantie de disponibilité pour participer à un effacement de consommation à un moment opportun, faisant l'objet d'un accord [EU-2004]. Les clients concernés par ces programmes peuvent être des consommateurs individuels (gros clients industriels), ou des agrégations de petits consommateurs regroupés de manière à constituer une charge non négligeable au niveau du réseau (groupement de clients résidentiels).

Le centre de pilotage peut représenter un fournisseur et/ou producteur d'énergie ou un Virtual power plant [NIK-2012] qui aura agrégé des centaines de bâtiments/systèmes avec une production décentralisée photovoltaïque (Figure 32). Il peut être représenté par un logiciel ou par un être humain.



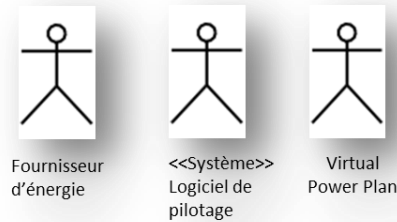


Figure 32 : Représentation des acteurs principaux dans notre outil

Les différents systèmes connectés sont définis par leurs secteurs d'activité. A l'intérieur de ces activités, nous proposons une différenciation des acteurs selon leurs contrats électriques qui les lient aux producteurs et fournisseurs d'énergie qui dépendent de leurs consommations énergétique. Nous présentons un exemple du système de facturation étudié en Suisse à Sierre dont les principaux paramètres tels que le raccordement ou le paiement du pic de puissance sont équivalents. Seul le prix va varier en fonction des pays Européens selon entre autre les différentes sources de production de l'électricité.

Détail du système de facturation dans le secteur tertiaire à Sierre, Suisse

Il existe deux grands types de facturation pour cette localité : la facturation de l'électricité chez les particuliers et les « petits consommateurs d'électricité », alimentés en moins de 100A, et celle pour les « grands consommateurs d'énergie », alimentés en plus de 100A.

Pour les compteurs de plus de 100A, les entreprises soumises à ce type de facturation bénéficient de tarifs préférentiels sur le kWh. En revanche, la pointe de puissance appelée mensuelle leur est facturée particulièrement cher (7.56CHF/kW). De plus, l'énergie réactive consommée peut leur être facturée si la quantité consommée dépasse 50% de l'énergie active consommée car cela déséquilibre le réseau électrique. L'énergie réactive correspond à l'énergie consommée par les appareils pour se mettre sur la même phase que l'énergie fournie par le réseau (Figure 33).

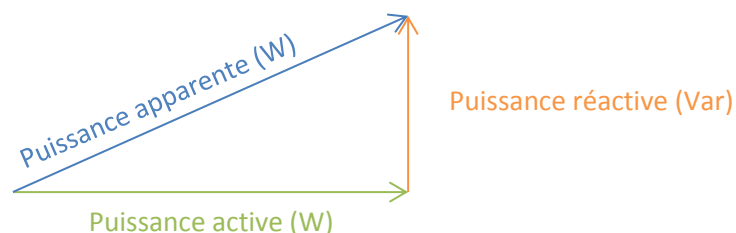


Figure 33 : Triangle des puissances indiquant la relation entre la puissance active, réactive et apparente

Selon le type de facturation, une prédiction précise du pic de puissance devient nécessaire à la fois pour la stabilisation du réseau mais également dans la diminution du prix dans la facture d'électricité

du consommateur pour les compteurs de plus de 100A. Par exemple, le pic de puissance représente 31% de la facture globale d'électricité sur l'année pour un restaurant d'entreprise (Figure 34).

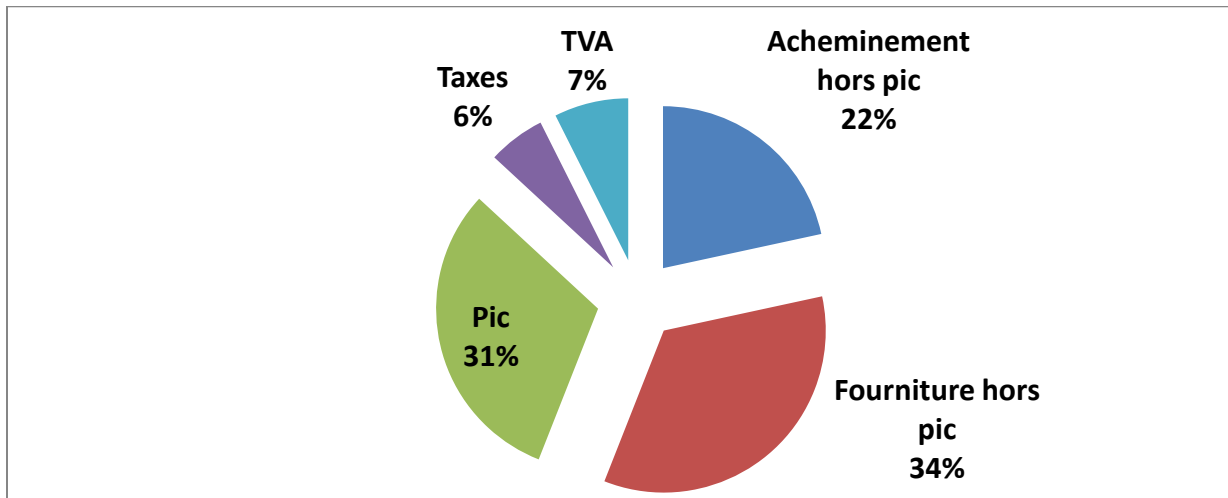


Figure 34 : Séparation de la facture d'électricité selon la fourniture et l'acheminement hors pic de puissance de l'électricité, la TVA, les différentes taxes et le pic de puissance qui comprend la part active et réactive pour le restaurant d'entreprise

A titre de comparaison, nous avons sélectionné trois entreprises avec trois secteurs d'activité différents (Studio de télévision, bureaux, restauration). Nous avons l'entreprise Canal 9 (Studio de télévision) qui consomme le plus parmi les trois entreprises, pourtant la pointe moyenne de consommation atteinte pendant l'année 2013 s'élève seulement à 38kW. En revanche, le Mikado (Restaurant d'entreprise ouvert seulement le midi), qui consomme l'équivalent de 36% de la consommation de Canal 9, a un pic moyen s'élevant à 52kW. Quant à l'entreprise Eticolle (Fabrication d'étiquettes), elle consomme à peu près la même quantité d'énergie que Canal 9 pourtant sa pointe de consommation s'élève à 99kW en moyenne sur l'année soit 2.6 fois plus que celle de Canal 9 (Figure 35). Ces trois entreprises, alimentées en plus de 100A, sont soumises à une facturation de la pointe de consommation mensuelle.

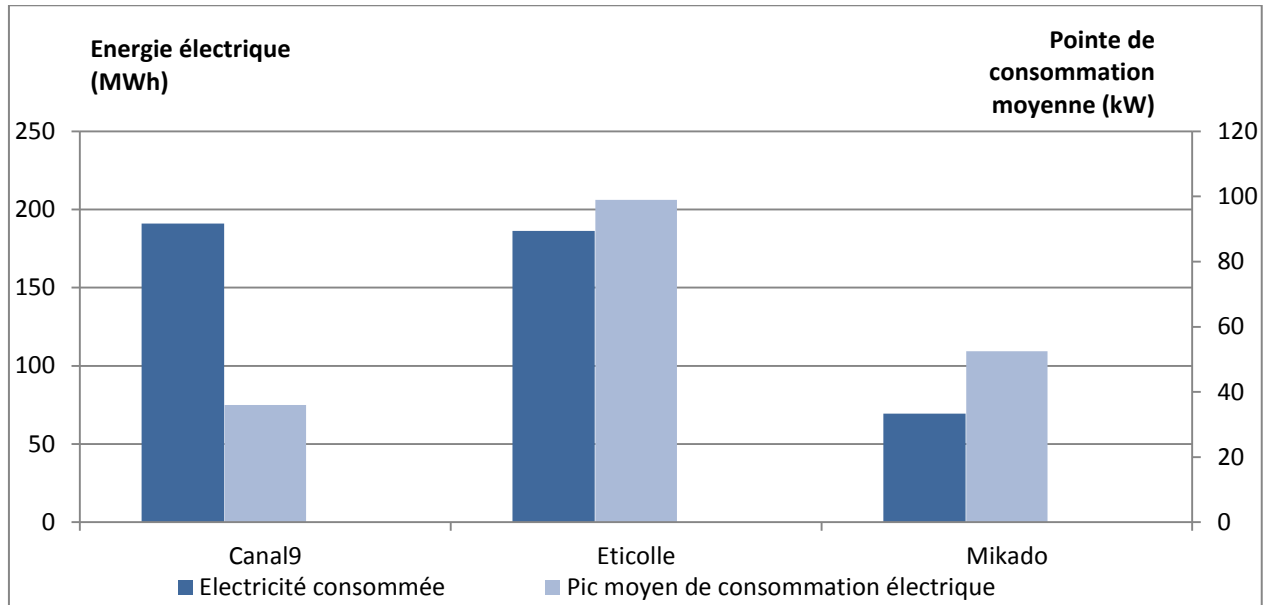


Figure 35 : Quantités d'énergies électriques consommées moyennes sur l'année 2013 et leurs pics de puissance associés pour trois consommateurs d'un des sites d'étude, Canal 9 (Studio de télévision), Eticolle (Fabrication d'étiquettes) et le Mikado (Restaurant d'entreprise)

Ces histogrammes sont expliqués par les courbes de charge journalières des entreprises. Le restaurant d'entreprise a une faible bande de base de 3kW, qui correspond à la consommation des groupes de froid. Cette entreprise ne consomme pas beaucoup d'électricité en soi, mais appelle beaucoup de puissance à certaines heures de la journée qui correspondent aux heures de préparation et de service des repas (Figure 36) avec la consommation des fours, des plaques électriques et des micro-ondes.

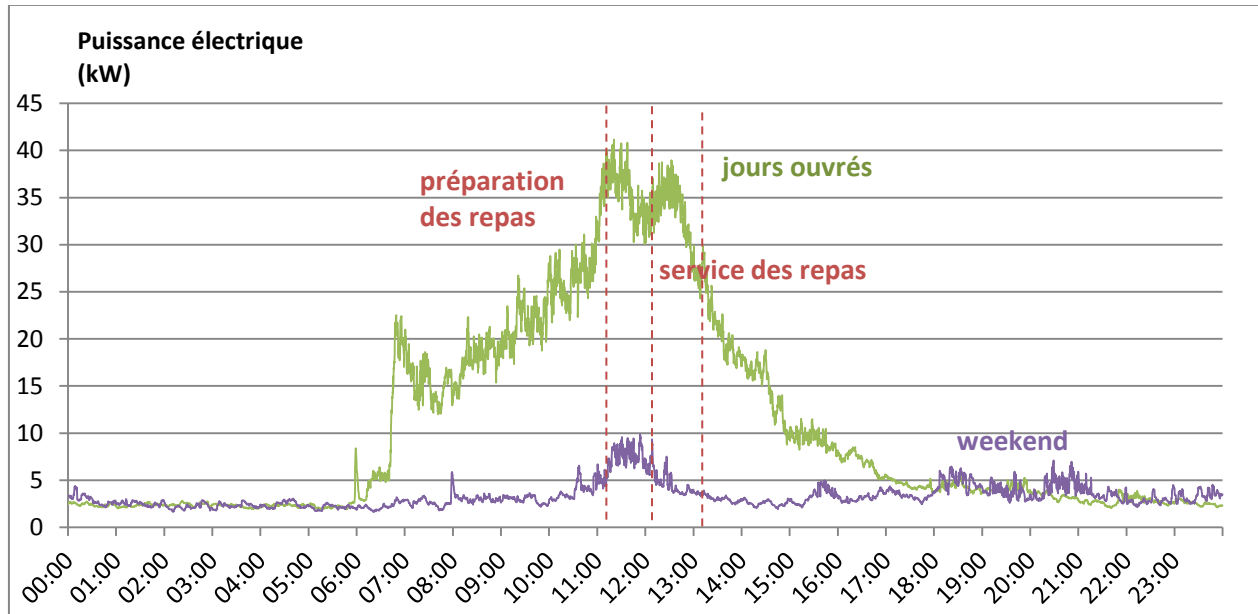


Figure 36 : Courbes de charges journalières moyennes en semaine et le weekend du restaurant d'entreprise moyennées sur le mois de Mars 2013

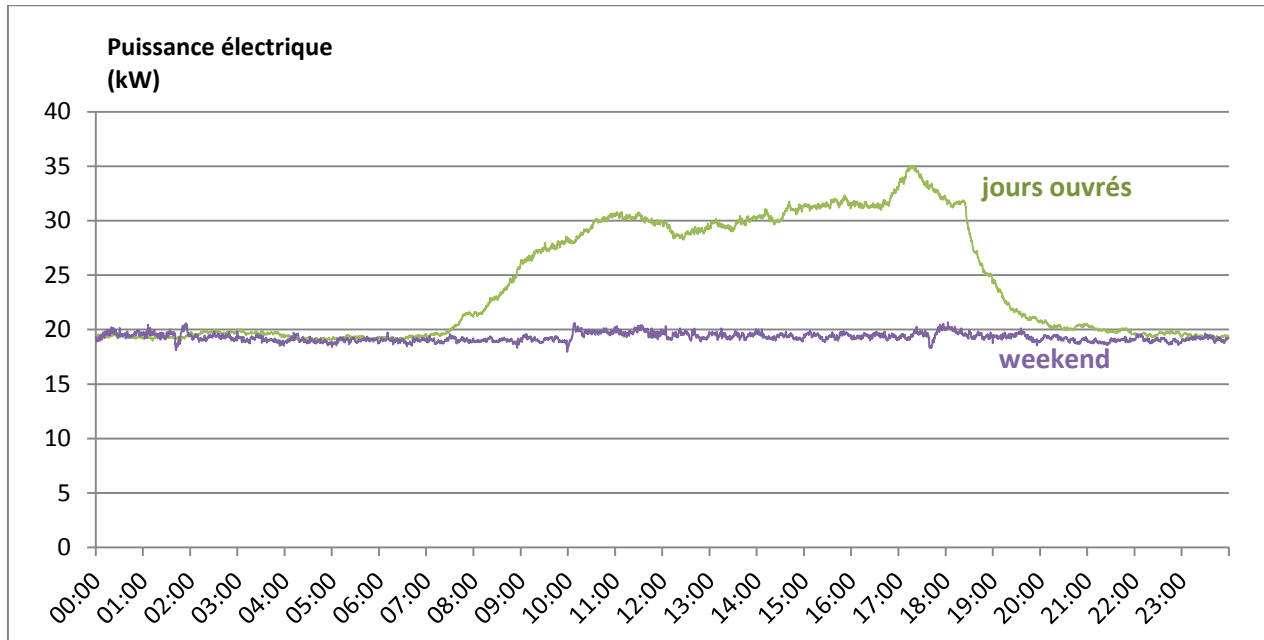


Figure 37 : Courbes de charges journalières moyennes en semaine et le weekend de l'entreprise de télévision Canal 9 moyennées sur le mois de Mars 2013

L'entreprise Canal9, qui diffuse des programmes de télévision, a une tendance de consommation très différente du restaurant d'entreprises, due à son secteur d'activité. En effet, les programmes et les supports pour leur diffusion devant être disponibles à toute heure du jour et de la nuit, 81% de la consommation de l'entreprise se situe en bande de base la journée (Figure 37).

Pour les compteurs de moins de 100A, les factures sont générées une fois par an. Si le prix du kWh est plus élevé, il n'y a pas de pointe de consommation ni d'énergie réactive facturée.

Ainsi, nous différencions les systèmes par leurs secteurs d'activité et leurs systèmes de facturation décrits dans la figure 38.

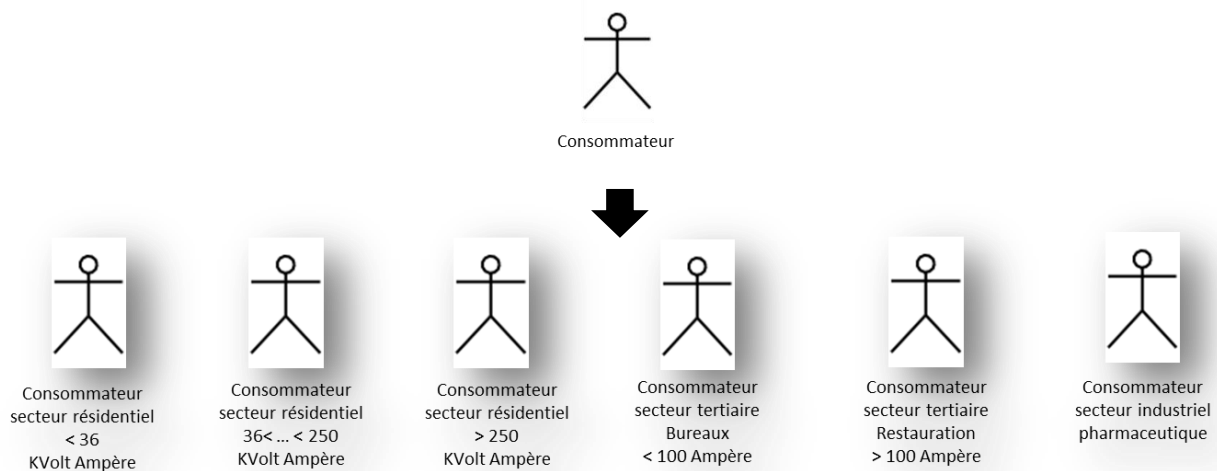


Figure 38 : Acteurs secondaires du système étudié qui est utilisé pour différencier les systèmes

3.2.2 Définition des fonctionnalités

Notre objectif est de réaliser un **bilan énergétique prédictif** pour l'heure suivante. Cela permettra d'avoir des possibilités de **scénarios de surproductions ou surconsommations** sur un quartier donné à **court terme au niveau de chaque flux énergétiques (t + 1 heure)**. Cela implique une prédiction fine des différents flux énergétiques d'un système, en particulier le chauffage et l'eau chaude, qui représentent le plus gros potentiel de flexibilité dans les bâtiments et permettra ainsi de **compléter l'offre de stockage**.

Il nous faut ainsi définir une « **Capacité d'effacement** » (Figure 39). Le potentiel d'effacement se situe avant tout au niveau des charges thermiques liées au besoin en eau chaude et en chauffage. Si le système possède une production solaire décentralisée, il devient utile de réaliser une prédiction fine de la consommation et de la production pour charger ou décharger la batterie suivant ces prédictions et des consignes de pilotage liés à l'autonomie du système ou à la vente d'énergie.

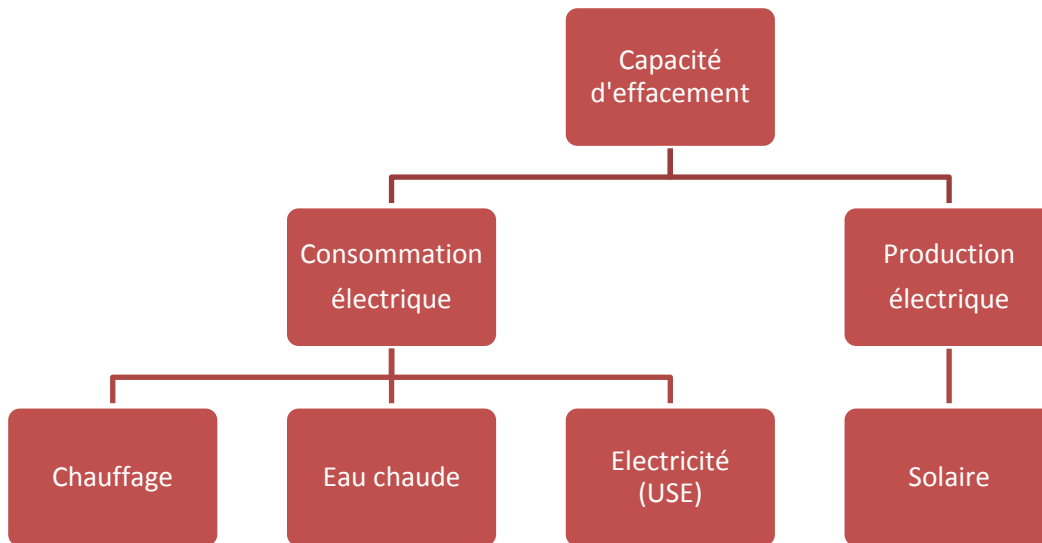


Figure 39 : Caractérisation de la capacité d'effacement qui nécessite la prédiction du chauffage, de l'eau chaude sanitaire, de la consommation électrique globale et de la production solaire

Une partie des données analysées peut-être envoyée aux consommateurs finaux qui sont les **acteurs secondaires**. L'**identification des appareils** regroupe les fonctionnalités qui permettent à partir du compteur global électrique d'identifier les différents usages. L'identification des appareils permet ainsi de fournir au consommateur final propriétaire ou non des connaissances sur la consommation globale du site et par usage qui diffèrent selon les secteurs d'activité. Nous nous attacherons à définir les appareils qui représentent la plus grande part dans la consommation énergétique du système. Dans le secteur résidentiel, le chauffage et l'eau chaude, en particulier sur les bâtiments les plus récents, représentent les plus gros consommateurs d'énergie.

L'usage de l'énergie électrique dans le secteur tertiaire est moins dominé par le chauffage que dans le résidentiel. A titre d'exemple, les usages d'eau chaude sanitaire et de cuisson sont très

développés dans la branche café-hôtel-restaurant alors que les usages spécifiques de l'électricité sont prédominants dans les commerces et les bureaux.

Le système avec seulement un appareil de mesure au niveau du compteur global, fournit **un outil de suivi de la consommation par usage** pour le consommateur final. Il permet également **d'augmenter la précision de la consommation électrique** en détectant les habitudes de consommation (par exemple, allumage de l'électroménager et du chauffage la nuit). De plus, si le système d'étude permet un effacement du chauffage et de l'eau chaude, cela permet de **limiter le nombre de données en transit** et stockées. Seules des informations de pilotage de type ON-OFF seraient envoyées.

En tant que	Je veux	Pour
Centre de pilotage	Prédire la consommation électrique globale	Adapter la consommation, la production et la capacité d'effacement
	Prédire la production globale	Adapter la consommation et la production
	Prédire le pic de puissance	Adapter la consommation et la production
	Définir une capacité d'effacement liée à l'eau chaude	Sécuriser le réseau
	Définir une capacité d'effacement liée au chauffage	Sécuriser le réseau
	Définir une capacité d'effacement liée au stockage électrique	Sécuriser le réseau
	Définir les données à collecter	Maintenir un haut niveau de prédiction
	Définir les données minimum à collecter	Limiter le nombre de capteurs à installer
	Définir les données minimum à collecter	Maintenir un haut niveau de prédiction
	Définir les modèles et méthodes de prédictions	Maintenir un haut niveau de prédiction
	Définir la fréquence de collecte	Limiter la capacité de stockage et de besoin en ressources informatiques
Consommateur	Connaître ma consommation totale et les principaux usages	Diminuer ma facture énergétique
	Un système non intrusif	Préserver ma vie privée
	Un système autonome de pilotage énergétique du chauffage, eau chaude sanitaire et du stockage	Me donner du temps
	Maximiser l'autoconsommation de mon système	Répondre aux enjeux climatiques à mon échelle
	Un système rentable	Gagner de l'argent

Tableau 4 : Tableau des fonctionnalités de l'outil

3.3 Collecter les données

3.3.1 Collecter les données au niveau des compteurs

Pour la collecte des données de consommation et de production dans le secteur résidentiel et tertiaire, nous nous sommes basés sur le projet i-BATS **[IBA-2013]** développé au sein de l'institut Informatique de Gestion à Sierre. Dans une optique de recherche, plusieurs niveaux de capteurs de la consommation ont été testés. Dans les locaux de la HES-SO Valais-Wallis, la température intérieure, le débit d'eau entrant et sortant du système de régulation de la température, l'intensité lumineuse, la détection de présence, la gestion des stores et la puissance active des appareils sont collectés. Dans ce cas, ces capteurs utilisent les protocoles de communication KNX et Zigbee décrit dans le projet européen IoT6 **[ANT-2014]**.

Un protocole de communication repose sur des règles qui définissent la transmission d'un message à travers un réseau. Un protocole de communication est donc nécessaire pour que deux équipements puissent échanger des données. Il existe des protocoles de communication standard comme le protocole **TCP/IP**. Des protocoles de communication sont spécialement adaptés au comptage électrique. Des protocoles de communication standardisés internationaux ont été définis. Les différents protocoles et caractéristiques liés au comptage électrique sont décrits dans **[COM-2010]**.

Malgré l'essor ces dernières années pour l'internet des objets, les capteurs testés qui permettent un relevé de la consommation d'énergie sans fil ne sont pas stables dans le temps (durée de vie supérieure à un an à comparer à celle d'un bâtiment). De plus, l'installation de capteurs filaires de type KNX **[ANT-2015]** par exemple dans les systèmes pour collecter la température ou le contrôle d'une vanne de chauffage demande des travaux complexes en particulier sur des logements existants.

Ceci nous conforte dans notre approche de limiter l'étude au niveau des compteurs globaux d'électricité concernant la consommation et la production si le système possède des capteurs solaires. Dans le cadre d'un outil qui se veut robuste et généralisable, nous nous sommes ainsi concentrés sur les communications filaires entre le capteur et les bases de données et sur les mesures globales de la consommation électrique par entreprise et de production solaire.

Dans le cadre du projet i-BATS, le choix des compteurs a été fait de façon à regrouper les communs et les entreprises consommant le plus pour obtenir 85% de la consommation électrique totale du site pour 20 compteurs installés. Les données relevées sont les nombres d'impulsions au niveau de chaque compteur d'électricité.

Chaque seconde, un capteur créé par la société Alro relève le nombre d'impulsions émises par le compteur auquel il est affecté. Une impulsion correspond à une quantité d'énergie consommée spécifique à chaque compteur. Le nombre d'impulsions est mesuré par seconde et est remonté à travers un bechhoff **[ANT-2014]** sur les différents serveurs soit par fibre optique soit par des câbles type RJ45 (Figure 40).

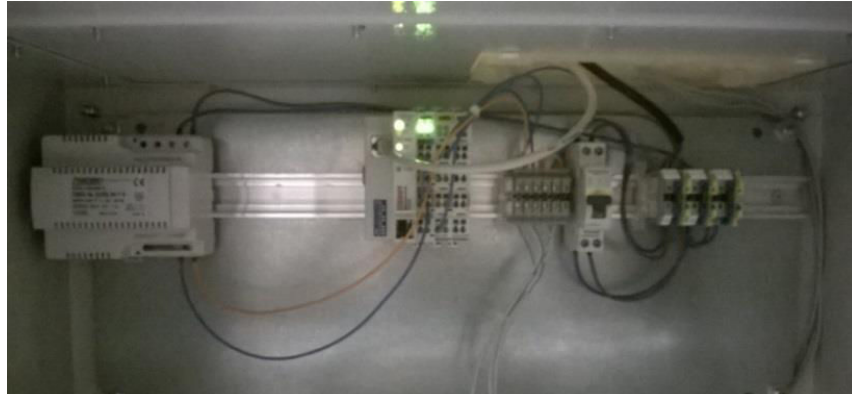


Figure 40 : Représentation d'un bekhoff qui fait le lien entre les compteurs d'électricité et les bases de données

Chaque compteur est identifié par une adresse IP qui lui permet de faire remonter les données acquises sur le serveur de l'entreprise. Il est affecté à ces données des coefficients (poids des impulsions et facteur compteur) qui sont stockées sous forme d'une puissance (Tableau 5). Ainsi la donnée en sortie du capteur est un nombre d'impulsions qui, après application d'un poids spécifique lié à cette impulsion, permet d'avoir une puissance pendant une durée déterminée, c'est-à-dire une énergie.

Cas d'étude	Données	Nom du compteur	Fréquence	Paramètres	Description Paramètres	Historique
Secteur tertiaire, entreprise connectée au système d'information	Consommation électrique globale	Beckhoff	1s	Sensorname	Identifiant unique par entreprise	18 mois
				gran	Granularité voulue	
				format	csv/json	
				fronts	timestamp	
				tots	timestamp	
				param	Puissance active (W)	

Tableau 5 : Données récoltées chaque seconde au niveau du transformateur basse tension pour mesurer la production photovoltaïque

Au niveau de la station transformatrice basse tension du quartier, un capteur Elco [ELC-2015] est placé au niveau de chaque entrée de bâtiment et au niveau de la production solaire. Ce capteur est branché sur chacune des phases du compteur et permet ainsi de renvoyer la puissance active et la puissance réactive par phase, à la fois pour la consommation en entrée des bâtiments, et pour la production à laquelle nous nous intéresserons dans cette étude (Tableau 6).

Cas d'étude	Données	Nom du compteur	Fréquence	Paramètres	Description Paramètres	Historique
Secteur tertiaire, panneaux photovoltaïques	Production électrique globale	ELCO	1s	Sensorname	Identifiant unique	18 mois
				gran	Granularité voulue	
				format	csv/json	
				fronts	timestamp	
				tots	timestamp	
				param	Puissance active (Watt) et puissance réactive (VAR) par phase	

Tableau 6 : Données récoltées chaque seconde au niveau de la transformatrice basse tension pour mesurer la production photovoltaïque

Au niveau du secteur résidentiel, un système permettant de mesurer la consommation électrique globale d'un foyer a été déployé sur 4 ménages tests. Ce système est installé à l'entrée du tableau électrique. Il comprend un appareil de mesures (Schneider Electric PM810) et un système de collecte de données (PC Engine) [ROD-2014]. Nous mesurons, à l'aide du PM810 de Schneider Electric les paramètres tels que la tension, le courant, la puissance active, la puissance réactive, chaque seconde directement sur chaque câble du compteur avec une précision de mesure de la puissance active de 1W (Tableau 7).

Cas d'étude	Données	Nom du compteur	Fréquence	Paramètres	Description Paramètres	Historique
Secteur résidentiel, maison	Consommation électrique globale	PM 810	1s	Sensorname	Identifiant unique par maison	18 mois
				gran	Granularité voulue	
				format	csv/json	
				fronts	timestamp	
				tots	timestamp	
				param	Puissance active (Watt) et puissance réactive (VAR) par phase	

Tableau 7 : Données récoltés chaque seconde au niveau du compteur global dans le secteur résidentiel

Pour la collecte des données dans le **secteur industriel**, nous avons travaillé à partir du projet Adaptive IES **[ADA-2015]**. Le matériel utilisé s'inscrit dans la stratégie globale de standardisation de l'automatisation du site de production de DEBIOPHARM RESEARCH & MANUFACTURING SA. En effet la CPU Siemens (S7-315 2PN/DP) choisie comporte 2 ports de communication Ethernet et suffisamment de ressources (Mémoire et Processeur) pour permettre de faire évoluer le système facilement. Les cartes d'entrées et de sorties sont tout à fait standards également. Le fait d'utiliser ce type de matériel permet une plus grande flexibilité et facilite la maintenance des installations.

3.3.2 Collecter les données météorologiques

Pour la majorité des installations photovoltaïques de faible puissance, aucune station météo n'est installée sur le site d'étude. Pour les autres installations, une station est installée et collecte les données de température et de radiation lumineuse. Il est possible que cette station collecte la température des cellules. C'est le cas pour notre station installée sur le site du technopôle. Cependant, nous avons voulu rester au plus proche de la réalité, c'est-à-dire dans le cas d'installations décentralisées de faible puissance avec aucune station météo présente sur le système.

Pour la collecte des données météorologiques, nous nous sommes basés sur le projet ELF4U **[ELF-2015]** développé au sein de l'institut Informatique de Gestion à Sierre. Nous nous sommes basés ainsi sur des services déjà existants et les solutions suivantes ont été analysées :

- ⇒ World weather online
- ⇒ Open Weather map
- ⇒ Meteoblue
- ⇒ Meteosuisse
- ⇒ Meteonews
- ⇒ Meteocentrale
- ⇒ Zygrib
- ⇒ Awekas

Les deux premières solutions proposent une version gratuite et fournissent les données sous forme API. Il s'agit de plateformes web automatiques qui fournissent des données mondiales. L'avantage de fournisseurs tels que Meteoblue, Meteosuisse, Meteonews et Meteocentrale est le contrôle des données par des météorologues qui les commentent et les ajustent en fonction de l'évolution du contexte météorologique **[ELF-2015]**. L'accès à ces services varie entre 300 et 2000 CHF/an.

Les meilleures solutions fournissant des données complètes et contrôlées sont Meteoblue, Meteonews, Meteosuisse. Nous avons utilisé les données de Météosuisse dans ce projet pour son meilleur rapport qualité / coût des données.

Notons que Meteonews ne fournit pas d'historique des prévisions météos et que l'historique des mesures météos provient des stations de Météosuisse. Les paramètres météorologiques récoltés sont : température, précipitations, irradiation, durée d'ensoleillement, hauteur de neige, vitesse et direction du vent, couverture nuageuse. Pour certaines des stations, les données réelles et prédites sont disponibles.

Sur la figure 41, nous avons un aperçu des stations météorologiques connectées à notre système d'information où nous distinguons 4 types de stations différentes :

- **Systèmes présentés dans ce rapport**
- **Stations avec des données météorologiques réelles et prédites**
- **Stations utilisées dans cette étude**
- **Stations avec uniquement des données hydrographiques.**

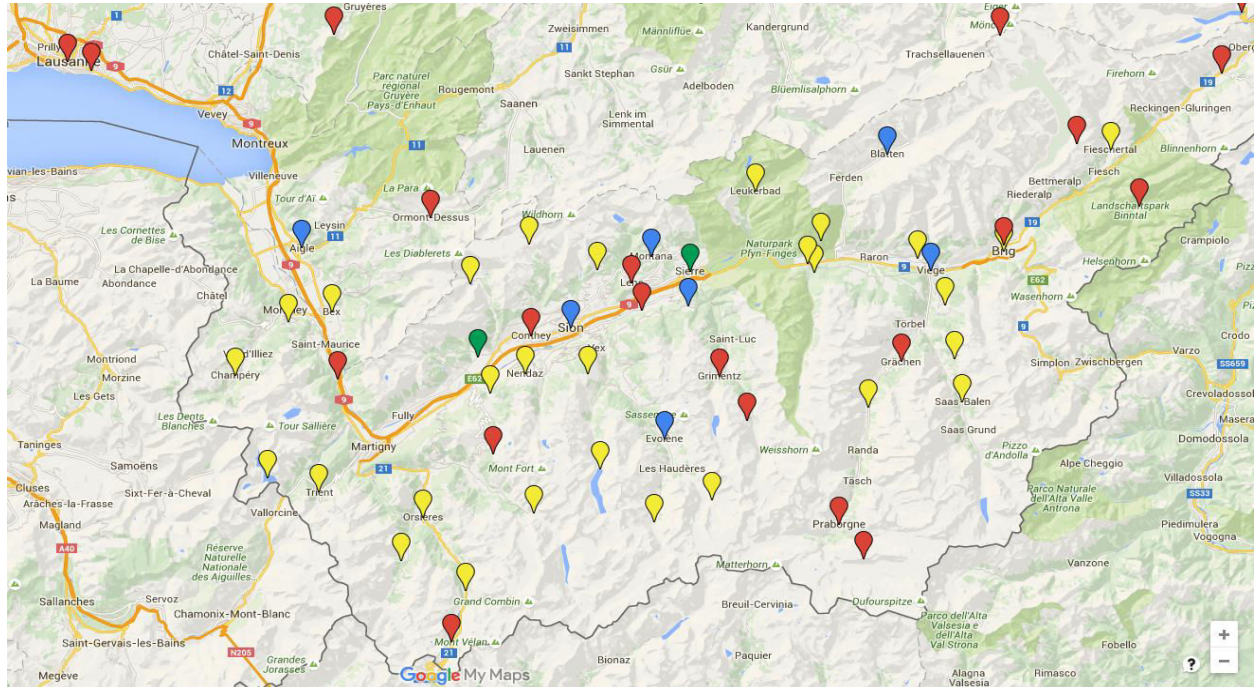


Figure 41 : Stations météorologiques connectées à notre système d'information

Pour caractériser à la fois la consommation énergétique et la production solaire décentralisée, nous nous sommes concentrés dans cette étude à récupérer chaque heure les données de température (°C) et de radiation lumineuse (W/m^2) réelles et prédites (Tableau 8).

Capteur	Nom du compteur	Fréquence	Paramètres	Description Paramètres	Historique
Station météo	Météo Suisse	1h	Sensorname	Identifiant unique par station	5 ans
			gran	Granularité voulue	
			format	json	
			fronts	timestamp	
			tots	timestamp	
			param	Température réelle et prédite	

				(°C), radiation lumineuse réelle et prédite (W/m ²)	
--	--	--	--	---	--

Tableau 8 : Données météorologiques utilisées dans notre outil de prédiction

3.3.3 Identifier la présence humaine

L'identification de la présence humaine est un des points fondamentaux dans la prédiction de la consommation électrique et dans la création de scénarios liés au pilotage de charges. Lorsque l'on veut surveiller l'intérieur d'une zone considérée, pour nous une pièce intérieure, on parle de détection intérieure [BEL-2010]. Plusieurs détecteurs existent : des détecteurs hyperfréquence, des détecteurs ultrason ou encore des détecteurs infrarouges passifs. Une étude comparative des différents systèmes est disponible [BEL-2010] et résumée dans le tableau 9. Une méthode d'identification consiste à calculer un taux d'occupation basé sur la mesure de la consommation au niveau des prises pour caractériser les usages spécifiques électriques utiles. Nous pouvons également citer les travaux qui consistent à placer un bracelet connecté sur les personnes pour connaître précisément le nombre de personnes [VIC-2013].

Type de détecteurs	Précision	Information				Coût	Intrusive
		Détection de mouvement	Nombre d'occupants	Localisation	Activité Physique		
Barrières lumineuses	Basse	+/-	+	-		-	-
Détecteur micro-ondes	Basse	+	-	-	-	+/-	-
Ultrasons	Basse	+	-	-	+/-	+/-	-
Détecteurs de chocs	Basse	+	-	-	+/-	+/-	-
Caméra infrarouge	Très haute	+	+	+	+	++	+
Pyrodétecteurs	Très haute	+	+	+	+	+/-	-
Transpondeurs	Basse	+	-	-	-	+/-	-
Caméra Vidéo	Très haute	+	+	+	+	+	+
Systemes biométriques	Basse	-	+	-	-	+	-

Tableau 9 : Techniques d'identification de la présence humaine

Il est toujours possible comme première approche et définition des modèles de référence d'utiliser les normes SIA 380/4 et SIA 2024 mais les valeurs restent des moyennes, qui ne répondent pas à un pilotage réactif.

En suivant nos contraintes pour la création d'un outil non intrusif, nous pouvons remarquer que les méthodes d'identification nécessitent la mise en place de capteurs à l'intérieur de notre système et ne sont pas adaptés à notre outil.

Nous préférons ajouter le planning de production comme information a priori sur nos différents systèmes pour à la fois améliorer nos données d'entraînement et identifier des zones de pilotage en fonction de la présence humaine (Tableau 10).

Cas d'étude	Données	Fréquence	Paramètres	Description Paramètres	Historique
Secteur tertiaire et industriel	Planning de production journalier	1 an	Id	Identifiant spécifique à un système	3 ans
			Heure_in	Heure d'arrivée des employés	
			Heure_out	Heure de sortie des employés	
	Planning de production par semaine		Id	Identifiant spécifique à un système	
			Jour_in	Premier jour de la semaine travaillé	
			Jour_out	Dernier jour de la semaine travaillé	

Tableau 10 : Données de présence utilisées dans notre outil de prédiction pour identifier les périodes de production

Dans le secteur tertiaire et industriel, les heures d'arrivée et de départ du personnel et les jours travaillés comme informations d'entrée permettront d'affiner la qualité du set d'entraînement en enlevant des données que l'on peut considérer comme du bruit (Figure 42). Dans le cas du secteur résidentiel, cela passe par la connaissance sur les consignes données aux systèmes de chauffage et d'eau chaude.

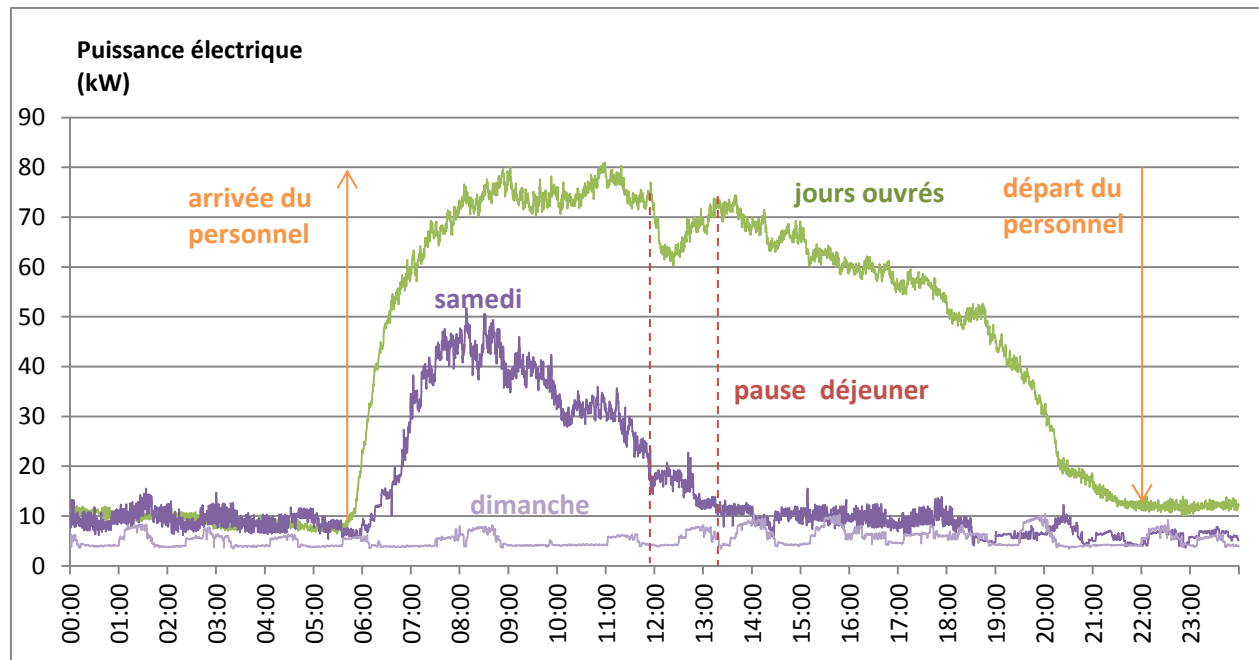


Figure 42 : Evolution des courbes de charges journalières moyennées sur le mois de Mars 2013 caractérisant les jours de la semaine, le samedi et le dimanche pour une industrie de fabrication d'étiquettes

3.4 Stocker les données

Selon l'objectif voulu et la taille du système étudié (maison, quartier) en question, le système de stockage mis en place diffère. Lors de nos premiers essais en 2012, lors du lancement du projet I-BATS, nous avons commencé le plus simplement possible avec des bases de données MySQL et nous avons très vite migré vers des bases de données NoSql [NOS-2015]. Cependant, notre objectif n'est pas seulement de stocker des données mais d'y faire appel sur des temps très courts (la seconde ou la minute par exemple pour le suivi de contraintes électriques) pour accéder aux informations des capteurs. La solution finale a été de créer plusieurs bases de données en parallèle distinctes MONGODB pour les données concernant la consommation et la production.

Pour les données météorologiques, nous nous sommes basés sur le projet ELF4U. Un centre de stockage et d'importations de données énergétiques a été créé. Après analyse, le choix de l'outil s'est porté sur une base de données 'Open Source' spécialisée dans le stockage et le traitement de séries temporelles : Influx DB [ELF-2015].

Cette base de données a été choisie pour ses fonctionnalités avancées en terme de traitement de séries temporelles notamment avec l'agrégation de données, l'interfaçage avec Grafana (un outil de visualisation) ainsi que pour son aspect 'Open source' et la communauté active qui l'entoure. Aujourd'hui le centre e-Energy center à travers le centre Icare nous met à disposition à travers une API spécifiques les données stockées (Figure 43).

A partir d'un logiciel d'analyse décrit dans le paragraphe suivant, nous pouvons accéder à tout moment aux données voulues. Pour communiquer avec les autres composants du système d'information, une solution à base de services web de type REST utilisant le format JSON a été utilisée. Le format JSON est un standard d'échange de données et de nombreuses bibliothèques permettant de le traiter sont présentes sur toutes les plateformes du marché. De tels services permettent la construction d'applications lisibles et facilement utilisables dans le cadre de développement d'applications clientes [CAT-2011] [BAK-2012].

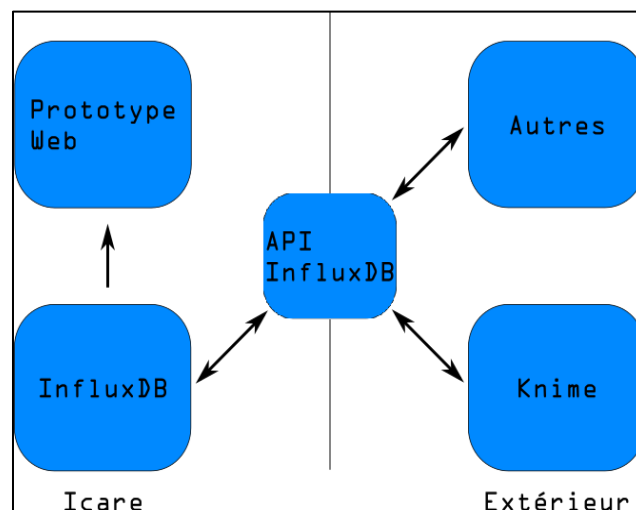


Figure 43 : Echanges d'information mises en place dans le cadre du projet ELF4U, [ELF-2015]

A ce stade, nous utilisons les bases de données MONGODB pour les données relatives à la consommation et la production de nos différents systèmes et les bases de données Influx pour les données météorologiques (Tableau 11). L'objectif final sera de basculer l'ensemble du système vers des bases de données Influx DB.

Cas d'étude	Données	Nom du compteur	Fréquence	Historique	Type base de données	Accès	Format
Secteur tertiaire	Consommation électrique globale	Bechkoff	1s	18 mois	MONGO DB	API REST	JSON
	Données météorologiques	Météo Suisse	3h	5 ans	Influx DB		

Tableau 11 : Accès aux données collectées par les capteurs dans le secteur tertiaire

3.5 Traiter et distribuer l'information

Par rapport à notre contrainte imposant de limiter au maximum nos coûts d'installation, nous sommes intéressés aux logiciels libres. Il existe des logiciels open source (gratuits), à l'instar de RapidMiner ou encore Weka, R, Knime résumé dans le tableau 12 et open source :

- ⇒ **KNIME** acronyme de Konstanz Information Miner, est un logiciel libre édité par un laboratoire de l'université de Constance dénommé Nycomed Chair for Bioinformatics and Information Mining. Il intègre tous les modules d'analyse de Weka et permet de créer des scripts en langage R.
- ⇒ **R** est un langage et un environnement permettant d'effectuer des calculs statistiques et de créer leurs graphiques. R fournit un large éventail de techniques statistiques et graphiques telles que la modélisation linéaire et non linéaire, les tests statistiques classiques, l'analyse des séries chronologiques, la classification. R est publié selon les termes de la licence GNU sous forme de code source.
- ⇒ **Orange** est un logiciel libre créé à l'université de Ljubljana en Slovénie. Ce logiciel est doté d'une interface homme-machine conviviale. Il est développé en C++ et en Python.
- ⇒ **Weka** est un logiciel libre de fouille de données créé par l'université de Waikato (Nouvelle-Zélande).

- ⇒ **RapidMiner** est un logiciel libre distribué par la société Rapid-I, basée à Dortmund en Allemagne. Il intègre le *Business Intelligence* dont les principales fonctionnalités sont l'ETL, l'OLAP, la production d'états et l'exploration de données et les techniques classiques comme les SVM, l'ACP, les arbres de décision et les réseaux neuronaux.
- ⇒ **Tanagra** est un logiciel libre d'exploration de données développé du laboratoire ERIC de l'Université Lumière Lyon 2.

Volume de données	Produit	Spécialité	Editeur
Faible (dizaines de milliers d'enregistrements)	NeuralWorks Predict	Réseau de neurones	NeuralWare
	NeuroOne	Réseau de neurones	Netral
	Wizwhy		Wizsoft
	WEKA		"open source" Université de Waikato, Nouvelle-Zélande)
	R		"open source"
	RapidMiner		Rapid-I
	DataLab	Prétraitement des données	Complex Systems
	Alice	Arbre de décision	Isoft
Moyen (centaines de milliers d'enregistrements)	KnowledgeSEEKER	Arbre de décision	Angoss
	KnowledgeSTUDIO		Angoss
	C 5.0	Arbre de décision	RuleQuest Research
	CART	Arbre de décision	Salford Systems
	MARS	Arbre de décision	Salford Systems
	TreeNET	Gradient boosting	Salford Systems
	Predictive Modeling		Salford Systems
	Polyanalyst	Data mining + Text mining	Megaputer
	TANAGRA		Université de Lyon
	JMP		SAS Institute
	SPAD		Coheris
	S-PLUS		TIBCO Software
	Elevé (millions d'enregistrements)	InfiniteInsight	
KNIME		Data mining + Text mining	Knime AG incorporated in Switzerland
Oracle Data Mining			Oracle
IBM SPSS Statistics			IBM
IBM SPSS Modeler		Data mining + Text mining	IBM
Statistica Data Miner		Data mining + Text mining	Statsoft
Insightful Data Miner			TIBCO Software
SAS/STAT			SAS Institute
Enterprise Miner		Data mining + Text mining	SAS Institute

Tableau 12 : Tableau récapitulatif des logiciels d'analyse de données [DAT-2015]

Avec nos différentes bases de données qui représentent des centaines de millions d'enregistrements, nous nous sommes concentrés sur l'utilisation du logiciel KNIME. Nous avons ainsi pu créer autour de cet outil un système d'information complet qui part des données fournies par les capteurs, les stocke, est connecté avec KNIME qui nous permet le traitement des données et d'envoyer les informations relatives aux différentes prédictions à un niveau supérieur de pilotage. Plusieurs modèles mathématiques peuvent être testés. Nous pouvons ainsi comparer plusieurs méthodes et algorithmes de prédictions décrites dans le chapitre 2 :

- ⇒ Méthode linéaire : Modèle ARIMA et une régression linéaire
- ⇒ Méthode non linéaire : MLP, PNN, SVM, Random Forest et le gradient Boosted Tree

Une fois les résultats obtenus, il est possible de les réécrire dans la base de données, d'envoyer l'information vers un web service ou directement à des automates programmables (Figure 39).

L'ensemble du processus de prédiction a été testé et industrialisé dans le cadre de deux projets. Le premier nommé ELF4U [ELF-2014] consistait à déployer un système de prédiction pour les installations solaires pour le fournisseur et producteur d'énergie local ESR (Energie Sion Région). Le deuxième nommé APEAS avait pour objectif de fournir un outil de désagrégation de la courbe de charges de 200 maisons [APE-2015]. Dans ces projets, nous avons utilisé le logiciel KNIME qui met à disposition des interfaces qui permettent la connexion directement avec les bases de données ou à travers d'API (Figure 44). Nous avons utilisé cette dernière à travers des requêtes de type GET qui peuvent être envoyées avec une fenêtre de temps choisie. Après analyse et prédiction du ou des flux considérés, les données sont reconverties en format JSON puis envoyées sur la base de données par des requêtes POST [KNI-2017].

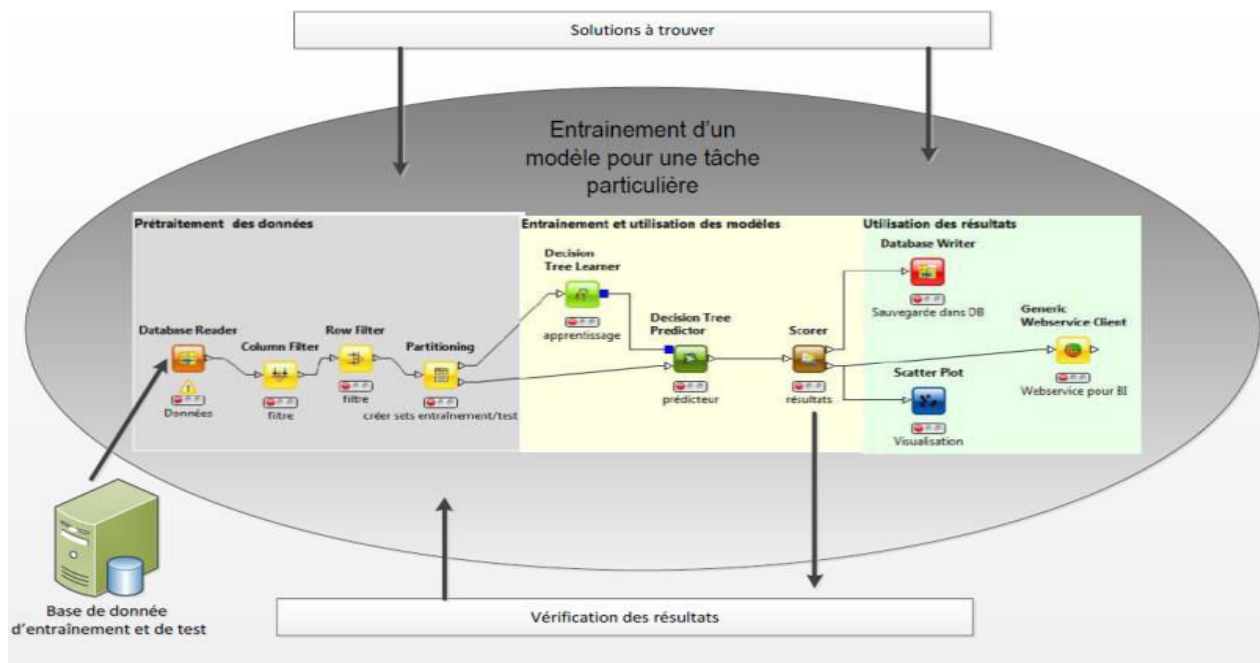


Figure 44 : Représentation du traitement des données à l'intérieur du logiciel d'analyse KNIME [HES-2012]

3.6 Conclusion

Nous avons décrit le processus qui nous a permis de mettre en place notre système d'information. Au vu de la bibliographie en particulier sur l'identification des appareils électriques, nous avons choisi une collecte des données de consommation à la seconde. Nous récoltons la puissance active et réactive au niveau du secteur résidentiel pour identifier les plus gros consommateurs et ceux ayant le plus gros potentiel de flexibilité : le chauffage et l'eau chaude. La production solaire est également récoltée sur chacun de nos sites.

Puisque nous voulons limiter le nombre de capteurs présents au sein de nos systèmes, nous apportons des informations a priori sur la technologie et l'énergie associées au chauffage et à l'eau chaude. Enfin le planning de production est également ajouté dans le secteur tertiaire et industriel pour limiter le bruit dans l'apprentissage de nos modèles mathématiques.

Chapitre 4 : Traitement des données

4.1 Introduction

Nous avons développé dans les chapitres précédents les différents flux énergétiques et les différents flux d'informations. Ces données sont stockées dans des bases de données et accessibles à tout instant. L'objectif de ce chapitre est de présenter le vecteur d'entrée pour chaque flux énergétique avant l'entraînement des différents modèles mathématiques décrits dans le chapitre 2.

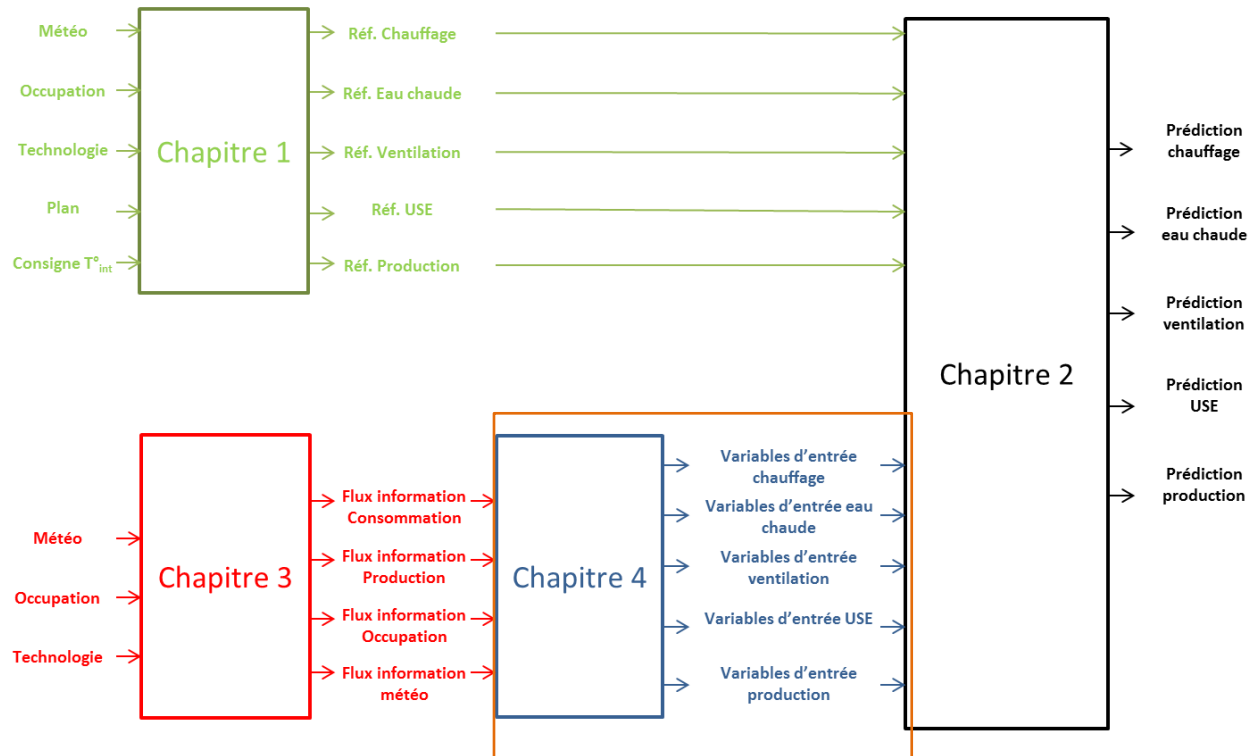


Figure 45 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision

4.2 Extraction des caractéristiques depuis le compteur global

4.2.1 Préparer et extraire les statistiques de puissance

Le prétraitement des données dépend du choix du capteur sélectionné pour mesurer la consommation d'énergie. Dans cette partie, nous présentons les connaissances tirées depuis le compteur global d'électricité, en particulier au niveau du secteur résidentiel où la puissance active et la

puissance réactive sont mesurées chaque seconde. Les connaissances tirées de cette partie sont généralisables peu importe la fréquence de mesure, seule la fenêtre d'échantillonnage évoluera.

Dans le cas de l'identification de l'ensemble des appareils, une étude à la seconde est nécessaire pour les appareils avec des durées de cycle court comme les convecteurs électriques ou le tambour du lave-linge. Cela permet de soutirer des connaissances liées aux appareils à cycle court : nous parlons de densité d'évènements [GUE-2011].

Ainsi, la fenêtre de temps d'identification des variations de puissance doit être différente selon le problème posé. Dans cette thèse, nous nous intéressons à l'identification du chauffage et de l'eau sanitaire fournis par des pompes à chaleur. Ces appareils ont des cycles lents supérieurs à la minute. Un filtre médian permettra d'enlever du bruit sur le signal électrique et d'enlever les appareils à cycle court qui pourrait se confondre avec des pompes à chaleur. Un filtre médian de taille 10 est appliqué sur chacune des phases nommé p au niveau du compteur global :

$$E_{\text{Consommation_électrique_totale}} t(i) = [Date ;$$

$$P_{\text{active}} t(i) = \text{médiane} \sum_{p=1}^3 \sum_{i=1}^{10} P_{\text{active_norm}} t(i) + P_{\text{active_norm}} t(i) + P_{\text{active_norm}} t(i) ;$$

$$P_{\text{réactive}} t(i) = \text{médiane} \sum_{p=1}^3 \sum_{i=1}^{10} P_{\text{réactive_norm}} t(i) + P_{\text{réactive_norm}} t(i) + P_{\text{réactive_norm}} t(i)]; \quad (4.1)$$

Le lissage permet d'enlever du bruit mais il a comme conséquence de perdre également de l'information (Figures 46 et 47). Comme énoncé ci-dessus, tout dépend de l'identification des appareils électriques recherchés.

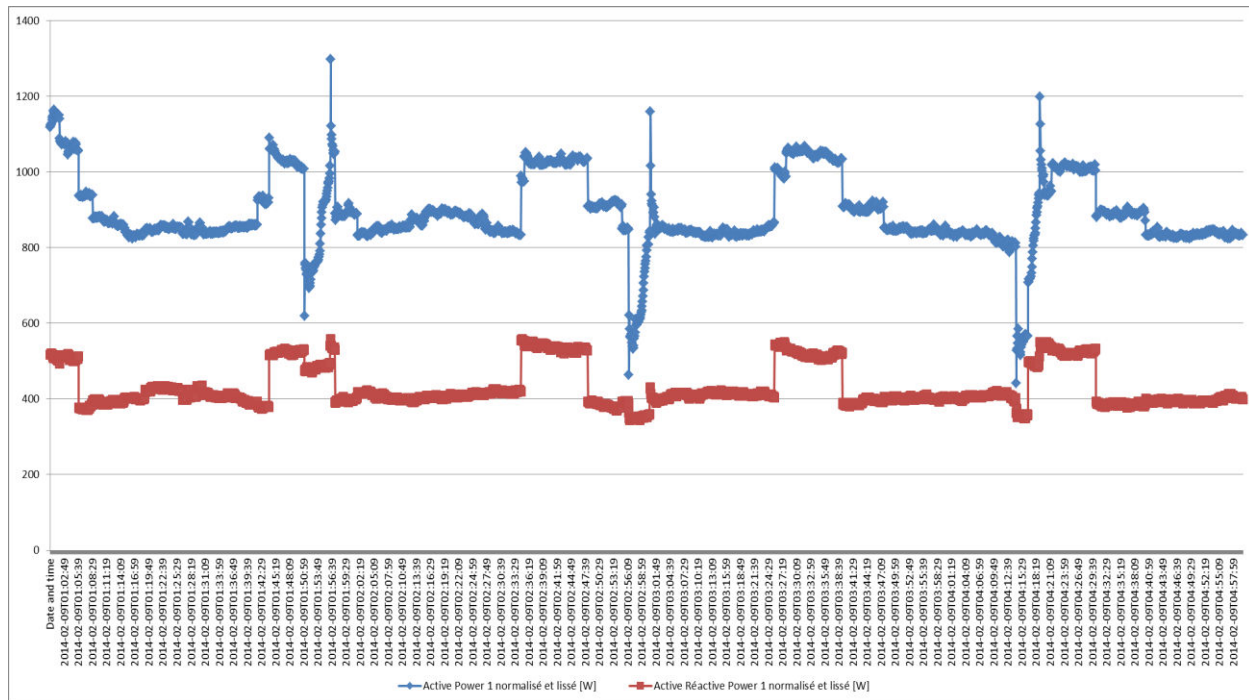


Figure 46 : Puissance active et réactive à la seconde en sortie du compteur d'électricité dans le secteur résidentiel sur une phase

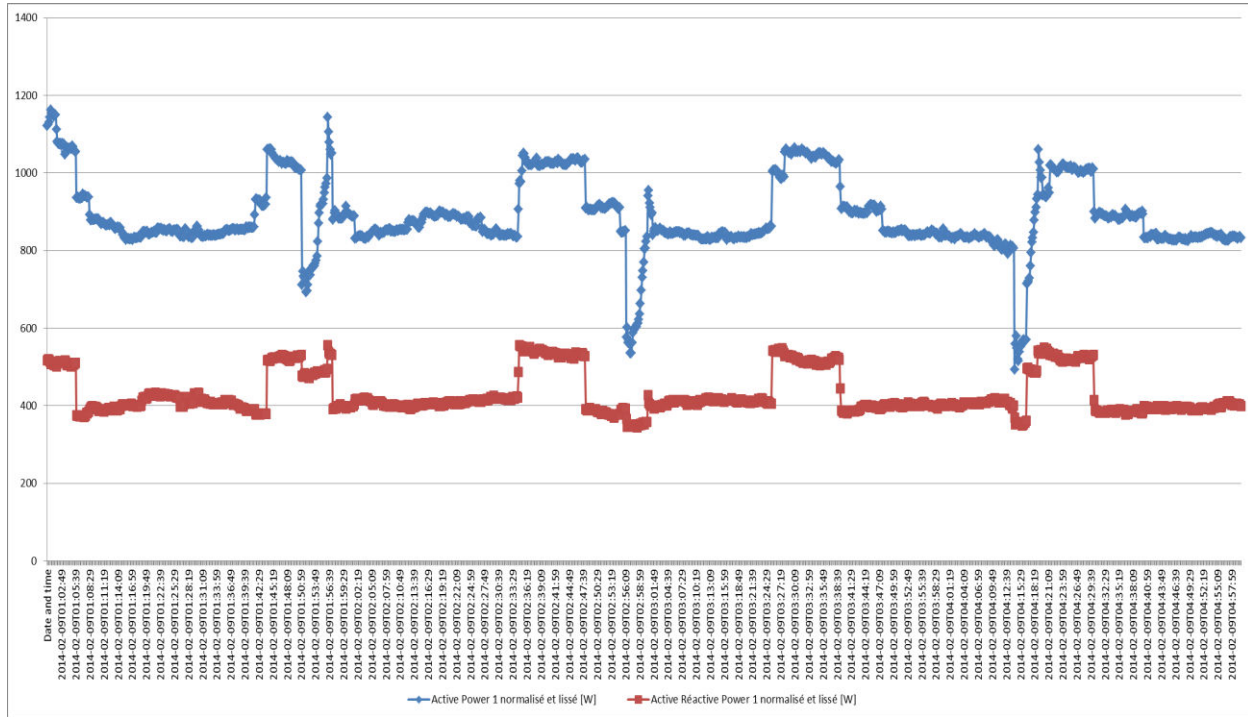


Figure 47 : Puissance active et réactive sans filtre au-dessus et après un filtre médian de taille 10

D'un point de vue électrique, en régime permanent, une variation est caractérisée par la transition entre un état électrique « stable » caractéristique du fonctionnement d'un appareil et un état instable. Dans la littérature, un évènement est une variation d'état au niveau de la seule puissance active et/ou réactive si le capteur récolte cette information [HART-1992].

Les seuils de puissance utilisés sont compris entre **40 Watts et 50 Watts** pour une identification complète des charges individuelles et de **15 VAR pour la puissance réactive** [HART-1992]. Cela sera notre point de départ comme seuil pour l'identification d'évènements sur la courbe de charge globale par phase. Cette variation peut se faire à partir d'une moyenne glissante sur une fenêtre de temps à définir selon le problème et les charges à identifier.

$$\text{Evènement } t(i) = [\text{Date}, \Delta P t(i), \Delta Q t(i)] \quad (4.2)$$

Un évènement est détecté si la puissance active et la puissance réactive dépasse un seuil sur une fenêtre de temps prédéfinie :

$$\begin{aligned} [\Delta P t(i), \Delta Q t(i)] &= [\text{Date}, \\ \Delta P t(i) &= \frac{P_{\text{active}} t(i-1) * (n-1) + P_{\text{active}} t(i)}{n} > \text{Seuil } P_{\text{active}} (40 \text{ Watt}), \\ \Delta Q t(i) &= \frac{P_{\text{réactive}} t(i-1) * (n-1) + P_{\text{réactive}} t(i)}{n} > \text{Seuil } P_{\text{réactive}} (15 \text{ VAR}) \end{aligned} \quad (4.3)$$

Nous avons maintenant en sortie, pour chaque heure, une table avec l'énergie globale normalisée et les différentes variations correspondante à une date donnée :

$$\begin{aligned}
 E_{\text{Consommation_électrique_totale}} t(i) &= [Date ; \\
 P_{\text{active_normalisé_phase1}} t(i) &= \left(\frac{230}{V t(i)}\right)^2 P_{\text{active}} t(i), \\
 P_{\text{réactive_normalisé_phase1}} t(i) &= \left(\frac{230}{V t(i)}\right)^2 P_{\text{réactive}} t(i) ; \\
 P_{\text{active_normalisé_phase2}} t(i) &= \left(\frac{230}{V t(i)}\right)^2 P_{\text{active}} t(i), \\
 P_{\text{réactive_normalisé_phase2}} t(i) &= \left(\frac{230}{V t(i)}\right)^2 P_{\text{réactive}} t(i) ; \\
 P_{\text{active_normalisé_phase3}} t(i) &= \left(\frac{230}{V t(i)}\right)^2 P_{\text{active}} t(i), \\
 P_{\text{réactive_normalisé_phase3}} t(i) &= \left(\frac{230}{V t(i)}\right)^2 P_{\text{réactive}} t(i) ; \\
 \Delta P t(i), \Delta Q t(i), Date_Evènement] &
 \end{aligned} \tag{4.4}$$

Statistiques sur les puissances et sur les variations

Les calculs statistiques permettent de caractériser un flux énergétique. Ils sont réalisés sur les puissances et sur les évènements. Ils sont les points d'entrée de la caractérisation passée de chaque flux énergétique.

La moyenne glissante : est une notion statistique, où la moyenne au lieu d'être calculée sur n valeurs fixes, est calculée sur n valeurs consécutives « glissantes ».

$$\bar{x} = \frac{\bar{x}_{n-1} * (n-1) + x_n}{n} \tag{4.5}$$

La médiane est la valeur à laquelle 50 % des valeurs observées sont inférieures. En supposant que l'on ait, au préalable, rangé les valeurs observées de sorte qu'elles se trouvent indexées suivant l'ordre des valeurs croissantes ($x_1 + x_2 + x_i + x_{i+1}$). Pour un nombre pair $2n$ de valeurs, la médiane est la *moyenne* des deux valeurs centrales, soit

$$\frac{x_n + x_{n+1}}{2} \tag{4.6}$$

ou toute autre valeur strictement comprise entre x_n et x_{n+1}

pour un nombre impair $2n+1$ de valeurs, la médiane est unique et égale à x_{n+1}

La **variance** est une mesure servant à caractériser la dispersion d'un échantillon ou d'une distribution. Si la série statistique est de moyenne m et prend les valeurs x_1, x_2, \dots, x_n , sa variance est

$$V(X) = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 \tag{4.7}$$

La déviation moyenne, la moyenne géométrique, le minimum et le maximum sont également ajoutés [BER-2010].

Identification des habitudes de consommations

Nous allons nous baser sur la variable de temps pour extraire les différentes connaissances qui permettront à nos modèles mathématiques de **définir des habitudes de consommations**. Nous ajoutons, tirés de planning public, des jours particuliers comme les jours fériés ou les vacances scolaires des enfants caractérisés par des variables booléennes :

- ⇒ des **mois** m allant de 1 à 12
- ⇒ des **jours de l'année** allant de 1 à 365
- ⇒ des **jours de la semaine** allant de 1 à 7 (lundi, mardi, mercredi, jeudi, vendredi, samedi et dimanche)
- ⇒ **l'heure** h allant de 0 à 23
- ⇒ des jours correspondant aux **vacances scolaires** cantonales caractérisés par une variable booléenne 0 ou 1.
- ⇒ des jours correspondant aux **jours fériés** cantonaux caractérisés par une variable booléenne 0 ou 1.

Comme décrit dans le chapitre 3, nous ajoutons comme information le **planning de production** pour chacun des flux énergétiques. Dans le cas du secteur résidentiel, cela passe par la connaissance des **consignes données aux pompes à chaleur** où une distinction est réalisée entre les jours de la semaine et les jours du week-end.

Concernant le secteur résidentiel, la consommation électrique globale, le besoin en chauffage et le besoin en eau chaude, nous **distinguons les jours de la semaine et le week-end**. **Concernant le secteur tertiaire et industriel**, les jours et les heures en dehors du planning de production sont **enlevées** de notre set de données d'entraînement et de test.

Si nous reprenons notre vecteur d'entrée, nous avons maintenant :

$$E_{\text{Consommation_électrique_totale}} t(i) = [\text{Date},$$

$$\begin{aligned} & \text{Statistiques Puissance phase 1 } t(i), \\ & \text{Statistiques Variations phase 1 } t(i) \\ & \text{Statistiques Puissance phase 2 } t(i), \\ & \text{Statistiques Variations phase 2 } t(i) \\ & \text{Statistiques Puissance phase 3 } t(i), \\ & \text{Statistiques Variations phase 3 } t(i) \end{aligned}$$

$$\text{Année}_j, \text{Mois}_m, \text{Jour}_j, \text{Semaine}_{0,1}, \text{Heure}_h, \text{Jour fériés}_{0,1}, \text{Vacances scolaires}_{0,1},$$

$$\Delta P t(i), \Delta Q t(i), \text{Date_Variations}] \quad (4.8)$$

Un modèle d'entraînement sera créé pour les jours k allant de lundi au vendredi (de 1 à 5) et un autre modèle pour le samedi et le dimanche (6 et 7) :

$$E_{\text{Consommation_électrique_totale}} t(i) = \sum_{j=1, k=1, w=0, h=1}^{j=365, k=5, w=1, h=24} E_{\text{Consommation_électrique_totale}} t(i) \quad (4.9)$$

$$E_{\text{Consommation_électrique_totale}} t(i) = \sum_{j=1, k=6, w=0, h=1}^{j=365, k=7, w=1, h=24} E_{\text{Consommation_électrique_totale}} t(i) \quad (4.10)$$

4.2.2 Identification du chauffage et de l'eau chaude

La plus grande difficulté dans l'identification des appareils électriques de manière non intrusive est sa généralisation. Dans les études qui ont porté sur des centaines de logements [ENE-2010] [ADE-2010], les données collectées ont souvent portées que sur la puissance active et sur la consommation d'énergie mais les analyses présentent des variations extrêmes entre les logements. Il devient ainsi difficile avec une fréquence à la seconde de séparer les évènements, en particulier ceux liés à des faibles puissances (généralement en dessous de la puissance du réfrigérateur-congélateur de 250 Watt en pic) mais qui représentent la plus grande part dans la répartition parmi l'ensemble des évènements détectés (Figure 48) [DUF-14]. Leurs nombres sont conséquents et correspondent à des appareils à un état (lampes, congélateurs) ou un sous-état spécifique (compresseur d'une pompe à chaleur) (Figure 43).

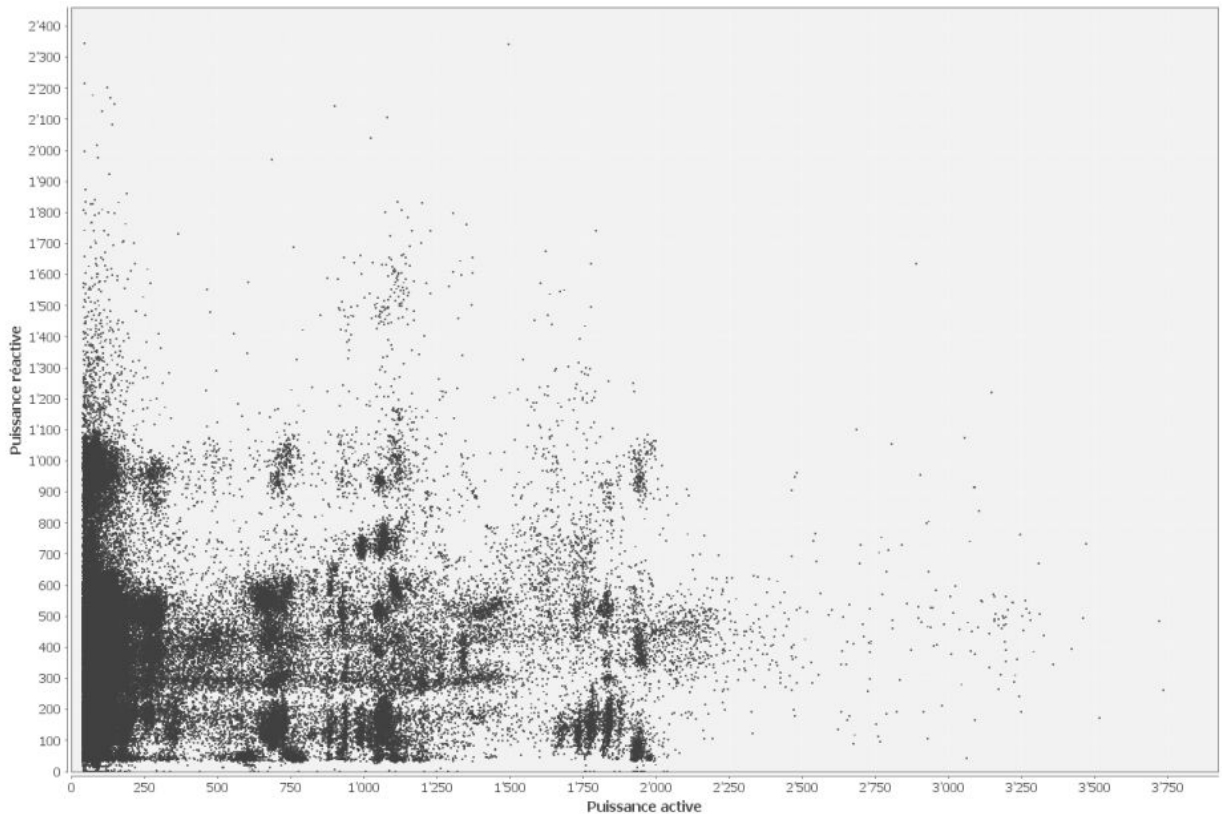


Figure 48 : Evènements détectés sur chacune des phases en sortie du compteur global entre le 1 Janvier et le 18 Juin 2014

Pour l'électroménager, une étude à la seconde est nécessaire pour tirer parti du nombre d'évènements liés aux appareils électriques à forte densité. Cette notion est définie dans [GUE-2009]. Elle permet par exemple d'extraire les variations dites rapides d'un convecteur électrique ou d'un lave-linge. Ainsi, le type de traitement avant l'identification des appareils de manière individuelle dépend de l'appareil visé. Nous nous sommes basés sur le secteur résidentiel car nous avons accès aux bases de contenant la puissance active et réactive d'appareils électriques.

Nous nous sommes concentrés sur les plus gros consommateurs et avec le plus gros potentiel de pilotage : le chauffage et l'eau chaude. Ces usages énergétiques sont fournis par une pompe à chaleur, appareil électrique qui a un cycle de fonctionnement lent. Les variations entre une variation positive et négative sont supérieures à la minute dans nos cas d'étude. Notre méthode peut se résumer en six étapes (Figure 49) :

- ⇒ Création d'une base de données d'appareils électriques avec la puissance active et réactive
- ⇒ Décomposition des appareils multi-états en sous états
- ⇒ Définition des différents évènements à partir de la puissance active et réactive sur chacune des phases (3 points de mesure) à partir d'un filtre médian à la minute
- ⇒ Recherche spécifique par classe : Pompe à chaleur/Autre
- ⇒ Ajouts de seuils spécifiques liés à la puissance active et réactive
- ⇒ Identification des évènements dans le plan actif/réactif à partir d'une base de données d'apprentissage et de méthodes d'identification non supervisé

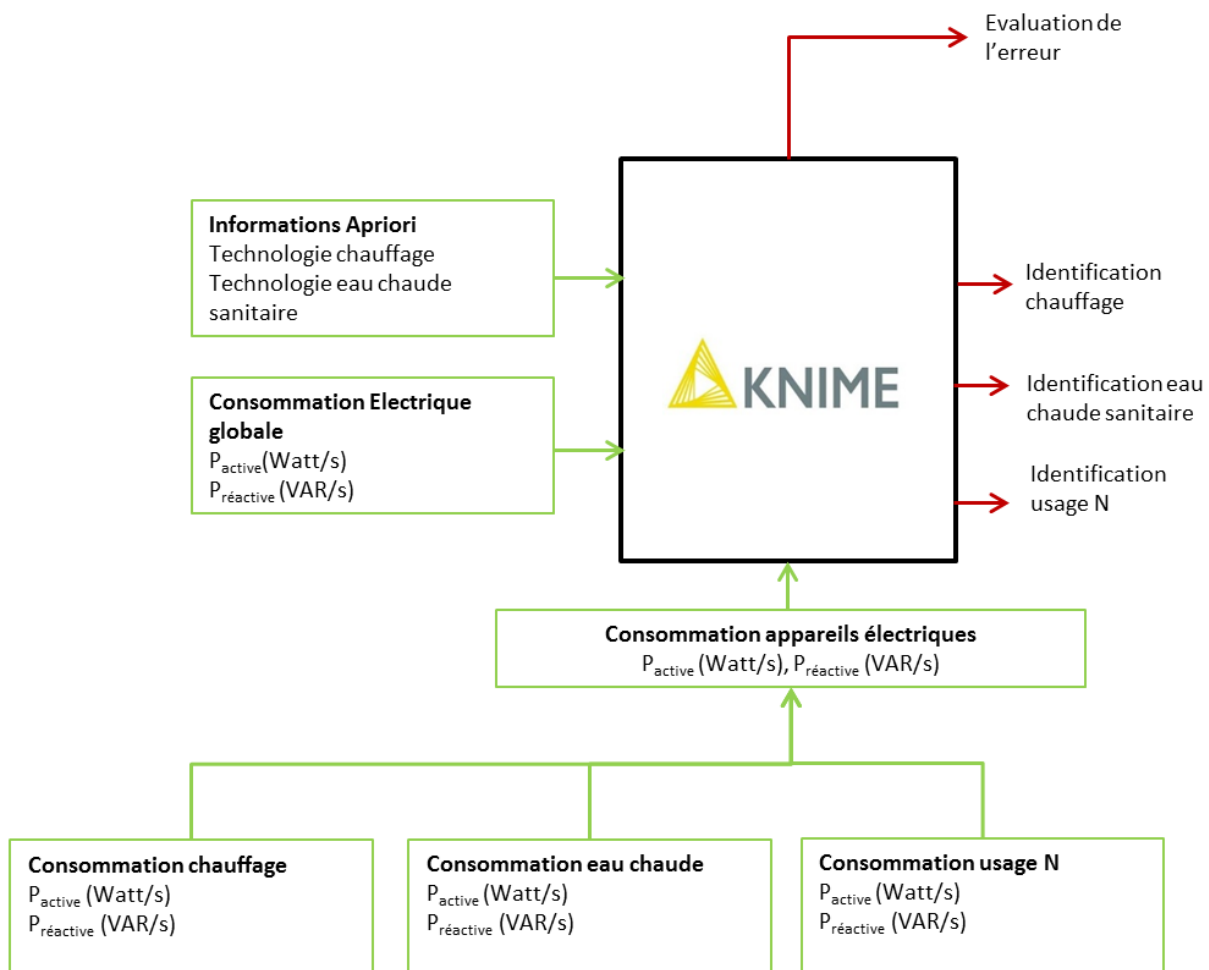


Figure 49 : Identification des appareils électriques par usages à partir de la consommation électrique globale en sortie du compteur et d'une base de données d'apprentissage

L'étude commence hors ligne par une meilleure décomposition des appareils électriques dans le but d'améliorer l'identification. Nous allons décomposer les appareils multi-états par une classification non supervisée. L'identification doit porter sur les appareils qui représentent soit le plus grand potentiel de flexibilité, soit la plus grande part dans la consommation d'énergie du système, soit les deux.

Notre objectif est de collecter la puissance active et réactive d'un maximum d'appareils utilisés dans le secteur résidentiel et tertiaire. Cette base de données représentera notre set d'entraînement de notre modèle d'identification. Concernant la campagne de mesure réalisée par la HES-SO Fribourg [RID-2015], deux sessions d'acquisition d'une heure avec une fréquence de collecte de 10 secondes sur 100 appareils domestiques divisés dans 10 catégories ont été réalisées (Tableau 13).

Appareils électriques	Paramètres	Unité	Fréquence de collecte	Historique
10 congélateurs	[Puissance active, Puissance réactive]	[Watt, VAR]	Seconde	1 mois
10 réfrigérateurs				
10 systèmes Hi-fi (lecteurs de CD)				
10 lampes				
10 machines à café				
10 Micro-ondes				
10 Imprimantes				
10 Ordinateurs fixes				

Tableau 13 : Base de données d'appareils électriques [RID-2015]

Appareils électriques	Paramètres	Unité	Fréquence de collecte	Historique
Pompe à chaleur Eau chaude	[Puissance active, Puissance réactive]	[Watt, VAR]	Seconde	1 mois
Pompe à chaleur Chauffage				
Lave-Vaisselle				
Lave-linge				

Tableau 14 : Base de données d'appareils électriques dans la maison test [DUF-2015]

Chaque appareil à plusieurs états présent dans nos bases de données est décomposé en sous états par un apprentissage supervisé.

Chaque pompe à chaleur de notre base de données (une pompe à chaleur liée à l'eau chaude et une autre liée au chauffage) est décomposée en plusieurs états. Nous avons choisi une décomposition en deux états pour se focaliser sur les variations à haute puissance en relation avec le condenseur.

Ces variations sont les données d'entrée d'un apprentissage supervisé développé dans le chapitre 4.4. Cela nous permet de créer des classes de manière supervisée ou non supervisée. Dans le cas de la pompe à chaleur, nous utilisons un algorithme, le 2-means [BER-2010] pour réaliser la classification qui va créer deux classes de manière itérative centrées chacune d'elles autour d'une moyenne.

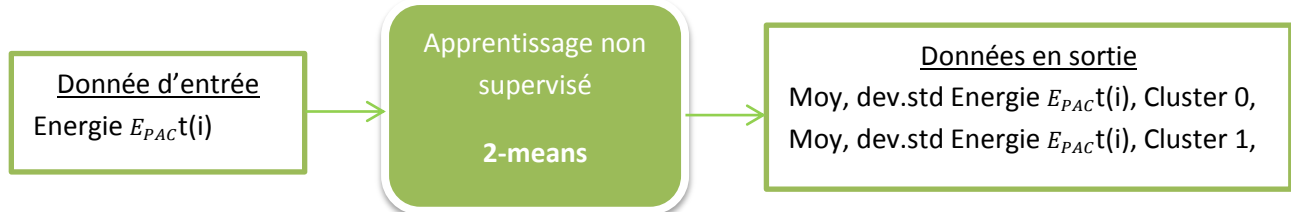


Figure 50 : Donnée d'entrée et de sortie de l'algorithme de décomposition de la signature électrique d'une pompe à chaleur en 2 états

Nous utilisons les évènements détectés en sortie des compteurs de la PAC liée au chauffage et celle liée à l'eau chaude comme données d'entrée pour identifier les classes de fonctionnement. La valeur absolue est calculée pour chaque évènement identifié. La connaissance liée à l'enclenchement ON ou déclenchement OFF de l'évènement est gardée.

Nous avons en sortie différents capteurs liés à nos pompes à chaleur les différentes variations liées à la puissance active et réactive :

$$\text{Variations } t(i) = [|\Delta P t(i), \Delta Q t(i)|, ON/OFF]. \quad (4.11)$$

Une identification à partir de l'algorithme SVM [DUF-2015] est réalisée sur la fenêtre de temps choisie par le centre de pilotage. Le détail de la méthodologie de prédiction du chauffage et de l'eau chaude par pompe à chaleur est décrit dans l'annexe 3.

En sortie nous avons l'énergie de chauffage et d'eau chaude consommée pendant l'heure considérée :

$$E_{\text{Consommation_électrique_totale}} t(i) = [\text{Date},$$

$$\text{Statistiques Puissance phase 1 } t(i),$$

$$\text{Statistiques Variations phase 1 } t(i),$$

$$\text{Statistiques Puissance phase 2 } t(i),$$

$$\text{Statistiques Variations phase 2 } t(i),$$

$$\text{Statistiques Puissance phase 3 } t(i),$$

$$\text{Statistiques Variations phase 3 } t(i),$$

$$E_{\text{Chauffage}} t(i),$$

$$E_{\text{ECS}} t(i),$$

$$\text{Année}_j, \text{Mois}_m, \text{Jour}_j, \text{Semaine}_{0,1}, \text{Heure}_h, \text{Jour fériés}_{0,1}, \text{Vacances scolaires}_{0,1}] \quad (4.12)$$

4.3 Créer son propre modèle de prédictions météorologiques

Les données météorologiques en particulier pour la production solaire décentralisée et pour le chauffage sont essentielles particulièrement les données prédites pour l'heure suivante.

Nous avons remarqué que les valeurs de prédiction fournies par MétéoSuisse comprenaient des erreurs en luminosité et en température extrêmes correspondant à des phénomènes locaux difficilement prévisibles. Nous définissons les erreurs liées aux prédictions fournies de température et de luminosité qui correspondent chaque heure à la différence entre la réalité et la prédiction (Equations 4.16 et 4.17). Notre objectif sera d'apprendre l'erreur sur des années d'historique pour l'appliquer sur nos années utilisées en test.

$$Erreur_{Luminosité} = Luminosité_{réelle} - Luminosité_{prédite} \quad (4.13)$$

$$Erreur_{Température} = Température_{réelle} - Température_{prédite} \quad (4.14)$$

Par exemple pour la luminosité, l'erreur est égale à la luminosité réelle moins la luminosité prédite. La dispersion de chaque erreur est gaussienne avec des écarts de plus de 1000 W/m2 pour la luminosité et de 24 degrés pour la température (Figures 51 et 52).

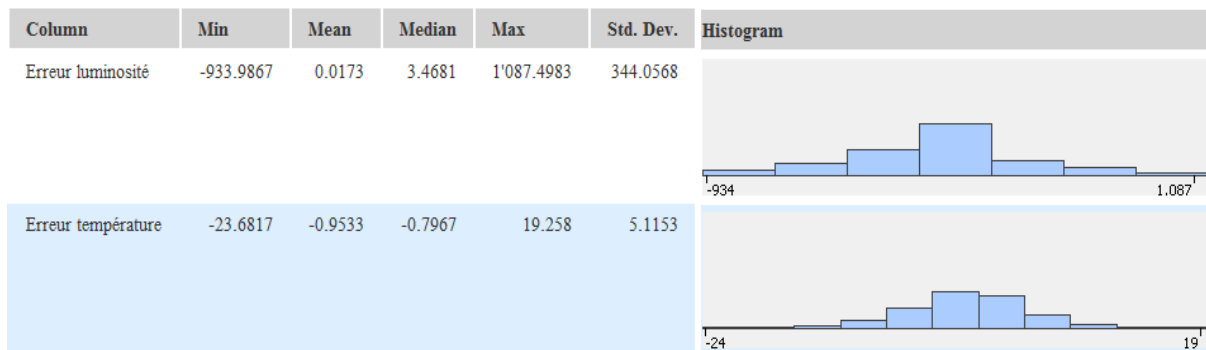


Figure 51 : Statistique et distribution de l'erreur de prédiction en luminosité et en température, station Aigle

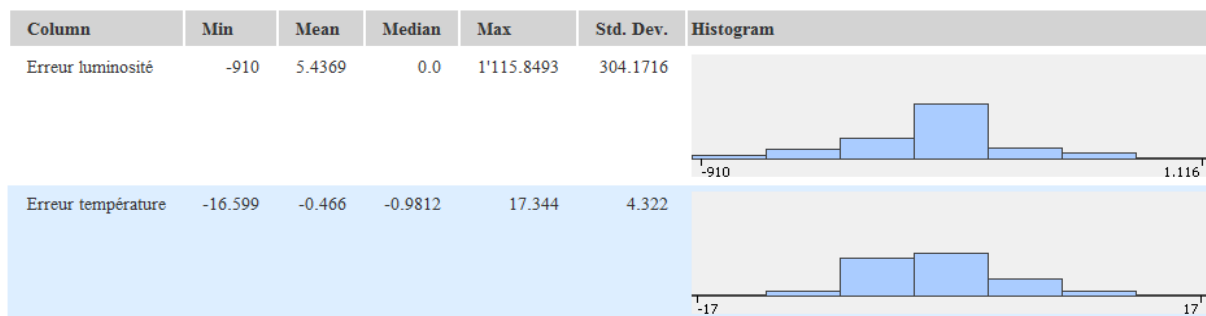


Figure 52 : Statistique et distribution de l'erreur de prédiction en luminosité et en température, station Sion

Puisque nous avons plusieurs années de données réelles et prédites, nous allons créer notre propre modèle de prédiction de la température et de la radiation lumineuse. Ainsi, nous avons créé un modèle mathématique de l'erreur par station météo et par donnée d'entrée (température, luminosité) (Figure 53).

Cette erreur est apprise par une méthode non paramétrique détaillée dans le chapitre 2. Pour cette étape, nous utilisons un ensemble d'arbres de décision. Les nouvelles prédictions de luminosité et de température corrigées seront nos nouvelles données d'entrée.

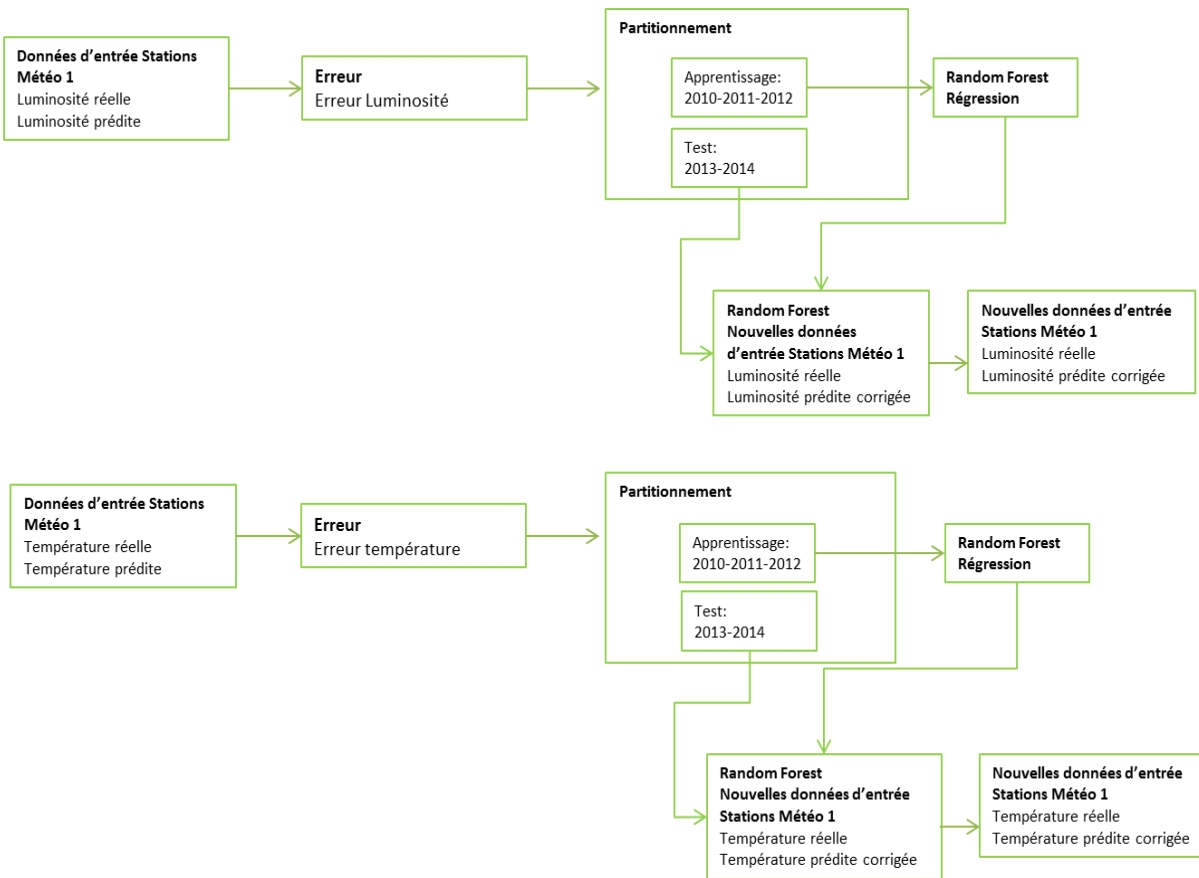


Figure 53: Méthodologie de réduction de l'erreur prédite météorologique provenant de MétéoSuisse

Pour caractériser l'erreur, un historique basé sur les données réelles et prédites est construit. Cet historique représente les données d'entrée de notre modèle mathématique. Les données d'apprentissage correspondent aux mesures du 1 Janvier 2010 au 31 Août 2012. Les données de test vont du 1 septembre 2012 au 31 Mars 2014 (Figure 54). Cette répartition a été réalisée pour avoir les données corrigées qui correspondent à nos périodes de test pour la prédiction des flux énergétiques.

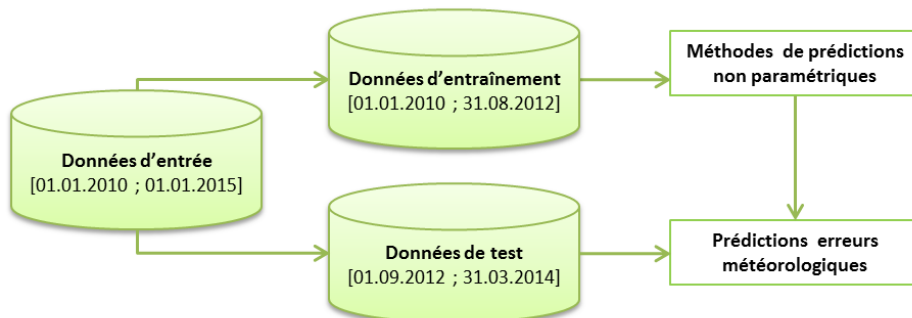


Figure 54 : Répartition des données d'entrée pour la correction des erreurs météorologiques

En sortie du modèle mathématique, une erreur est prédite pour chaque heure correspondant à notre set de test représenté par les figures 49 et 50.

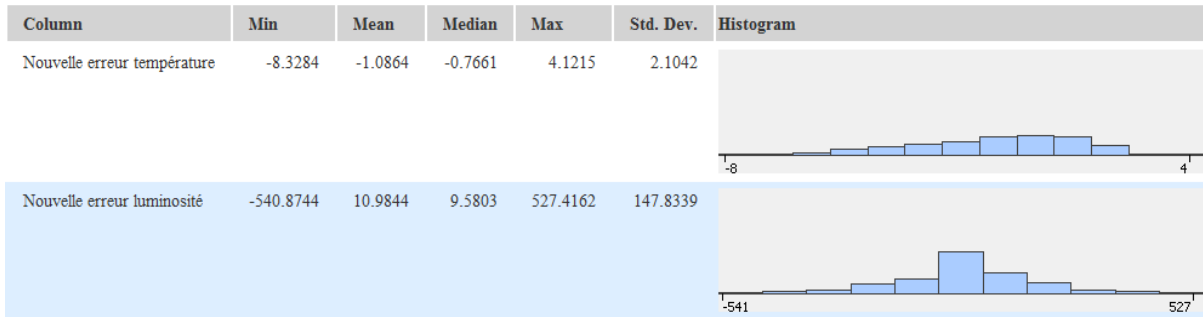


Figure 55: Statistiques et distribution de la nouvelle erreur de prédiction en luminosité et en température, station Aigle

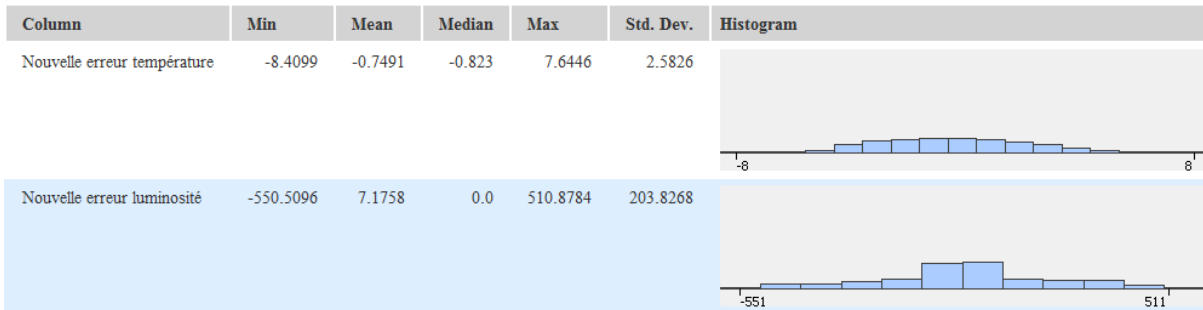


Figure 56: Statistiques et distribution de la nouvelle erreur de prédiction en luminosité et en température, station Sion

Les données météorologiques corrigées sont intégrées à notre vecteur d'entrée :

$$E_{\text{Consommation_électrique_totale}} t(i) = [\text{Date},$$

$$\text{Statistiques Puissance phase 1 } t(i),$$

$$\text{Statistiques Variations phase 1 } t(i),$$

$$\text{Statistiques Puissance phase 2 } t(i),$$

$$\text{Statistiques Variations phase 2 } t(i),$$

$$\text{Statistiques Puissance phase 3 } t(i),$$

$$\text{Statistiques Variations phase 3 } t(i),$$

$$E_{\text{Chauffage}} t(i),$$

$$E_{\text{ECS}} t(i),$$

$$\text{Données Réelles_Météo}_{\text{Station}_1} t(i),$$

$$\text{Données Prédites_Météo}_{\text{Station}_1} t(i),$$

$$\text{Données Réelles_Météo}_{\text{Station}_2} t(i),$$

$$\text{Données Prédites_Météo}_{\text{Station}_2} t(i),$$

$$\text{Données Réelles_Météo}_{\text{Station}_3} t(i),$$

$$\text{Données Prédites_Météo}_{\text{Station}_3} t(i),$$

$$\text{Données Réelles_Météo}_{\text{Station}_4} t(i),$$

$$\text{Données Prédites_Météo}_{\text{Station}_4} t(i),$$

$$\text{Données Réelles_Météo}_{\text{Station}_5} t(i),$$

$$\text{Données Prédites_Météo}_{\text{Station}_5} t(i),$$

$$\text{Année}_j, \text{Mois}_m, \text{Jour}_j, \text{Semaine}_{0,1}, \text{Heure}_h, \text{Jour fériés}_{0,1}, \text{Vacances scolaires}_{0,1}] \quad (4.15)$$

4.4 Classification

4.4.1 Apprentissage supervisé

Le processus se passe en deux phases. Lors de la première phase (hors ligne, dite d'apprentissage), il s'agit de déterminer un modèle des données étiquetées. La seconde phase (en ligne, dite de test) consiste à prédire l'étiquette d'une nouvelle donnée, connaissant le modèle préalablement appris. Cette méthodologie est utilisée dans le cas de la labellisation de l'eau chaude et du chauffage. L'objectif est de définir deux classes ON et OFF qui correspondent respectivement à l'allumage et à l'arrêt du système de production de chauffage et d'eau chaude.

A partir du besoin en chauffage, un apprentissage supervisé est réalisé sur les données d'entraînement et appliqué sur les données de test. Si l'énergie de chauffage est en dessus d'un seuil, elle est labellisée ON, sinon OFF. Une classification binaire est réalisée à partir de l'énergie de chauffage pour définir les quarts d'heure et les heures où le chauffage est allumé ou éteint (Figure 51). Si le besoin en chauffage ou en eau chaude à l'instant $t(i)$ est en dessous d'un seuil prédéfini, nous considérons le besoin comme OFF. Le même apprentissage est réalisé pour l'eau chaude sanitaire à partir de l'énergie par heure.

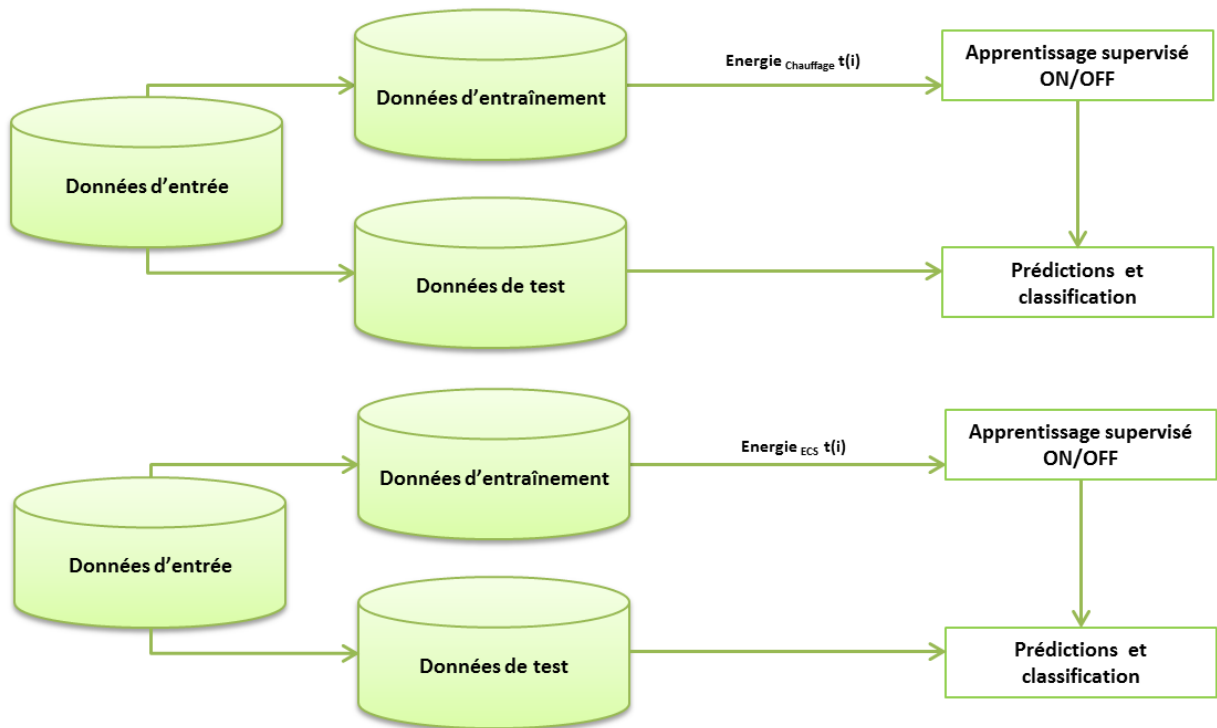


Figure 57 : Méthodes de classification des données pour le besoin en eau chaude sanitaire et en chauffage

Un historique des consommations et des données météorologiques est créé par heure pour chacune des variables d'entrée caractérisées. Cet historique correspond au besoin en chauffage heure

par heure sur les deux jours précédents la prédiction à calculer. Par exemple, pour une prédiction à l'heure, nous avons :

$$E_{\text{Chauffage}} t(i+1) = [\text{Date ;} \\ \text{Statistiques Puissance Phase 1 } t(i), \dots, \text{Statistiques Puissance Phase 1 } t(i-48), \\ \text{Statistiques Variations Phase 1 } t(i), \dots, \text{Statistiques Variations Phase 1 } t(i-48) \\ \text{Statistiques Puissance Phase 2 } t(i), \dots, \text{Statistiques Puissance Phase 2 } t(i-48), \\ \text{Statistiques Variations Phase 2 } t(i), \dots, \text{Statistiques Variations Phase 2 } t(i-48), \\ \text{Statistiques Puissance Phase 3 } t(i), \dots, \text{Statistiques Puissance Phase 3 } t(i-48), \\ \text{Statistiques Variations Phase 3 } t(i), \dots, \text{Statistiques Variations Phase 3 } t(i-48), \\ \text{Données Réelles_MétéoStation } t(i), \dots, \text{Données Réelles_MétéoStation } t(i-48), \\ \text{Données Prédites_MétéoStation } t(i+1), \dots, \text{Données Prédites_MétéoStation } t(i-48) \\ \text{Année}_j, \text{Mois}_m, \text{Jour}_j, \text{Semaine}_{0,1}, \text{Heure}_h, \text{Jour fériés}_{0,1}, \text{Vacances scolaires}_{0,1}, \\ \text{Classes}_{\text{ON/OFF}}] \quad (4.16)$$

Un historique des consommations et des données météorologiques est également créé par heure pour chacune des variables d'entrée caractérisées pour l'eau chaude sanitaire:

$$E_{\text{ECS}} t(i+1) = [\text{Date ;} \\ \text{Statistiques Puissance Phase 1 } t(i), \dots, \text{Statistiques Puissance Phase 1 } t(i-48), \\ \text{Statistiques Variations Phase 1 } t(i), \dots, \text{Statistiques Variations Phase 1 } t(i-48) \\ \text{Données Réelles_MétéoStation } t(i), \dots, \text{Données Réelles_MétéoStation } t(i-48), \\ \text{Données Prédites_MétéoStation } t(i+1), \dots, \text{Données Prédites_MétéoStation } t(i-48) \\ \text{Année}_j, \text{Mois}_m, \text{Jour}_j, \text{Semaine}_{0,1}, \text{Heure}_h, \text{Jour fériés}_{0,1}, \text{Vacances scolaires}_{0,1}, \\ \text{Classes}_{\text{ON/OFF}}] \quad (4.17)$$

4.4.2 Apprentissage non supervisé

Quand le système ne dispose que d'exemples non labellisés et que le nombre de classes n'a pas été prédéterminé, on parle d'apprentissage non supervisé. L'algorithme doit découvrir par lui-même la structure plus ou moins cachée des données.

Dans cette thèse, nous utilisons l'algorithme EM (espérance-maximisation) proposé par [VIC-2014]. C'est un algorithme itératif qui permet de trouver les paramètres de vraisemblance maximum d'un modèle probabiliste lorsque ce dernier dépend de variables latentes non observables. L'algorithme d'espérance-maximisation comporte :

- ⇒ une étape d'évaluation de l'espérance (E), où l'on calcule l'espérance de la vraisemblance en tenant compte des dernières variables observées,
- ⇒ une étape de maximisation (M), où l'on estime le maximum de vraisemblance des paramètres en maximisant la vraisemblance trouvée à l'étape E.

On utilise ensuite les paramètres trouvés en M comme point de départ d'une nouvelle phase d'évaluation de l'espérance, et l'on itère ainsi. On dispose de n observations, considérées comme les réalisations de n variables aléatoires indépendantes et identiquement distribuées (X_1, \dots, X_n) .

On note $L_\theta(x_i)$, $L(x_i; \theta)$, ou plus souvent $L(x_i | \theta)$, dans la littérature anglophone, la densité de X_i . On appelle vraisemblance de l'échantillon, la densité jointe de (X_1, \dots, X_n) :

$$L_n(x_1, \dots, x_n) = \prod_{i=1}^n L(x_i; \theta) \quad (4.18)$$

Dans le cas particulier d'une loi discrète, cela se ramène à :

$$L_n(x_1, \dots, x_n) = P_\theta \{(X_1 = x_1, \dots, X_n = x_n)\} = \prod_{i=1}^n P_\theta \{X_i = x_i\} \quad (4.19)$$

L'estimateur de θ est :

$$\theta = \arg \max L_n(x_1, \dots, x_n; \theta) \quad (4.20)$$

Cette méthodologie est utilisée dans le cas de la labellisation de la consommation électrique globale dans les secteurs résidentiel, tertiaire et industriel et de la production décentralisée solaire. En sortie, chaque classe est représentée par une moyenne, une déviation standard et sa répartition dans le set de données d'entraînement et de test.

La consommation électrique globale

A partir de la consommation électrique globale, un apprentissage non supervisé est réalisé. Cet apprentissage est réalisé pour chacun des systèmes étudiés, peu importe le secteur d'activité. Il est réalisé à partir de l'énergie électrique en sortie du compteur. Pour cet apprentissage, la donnée d'entrée utilisée est la consommation électrique globale et le modèle de classification est ensuite appliqué sur les données de test (Figure 52).

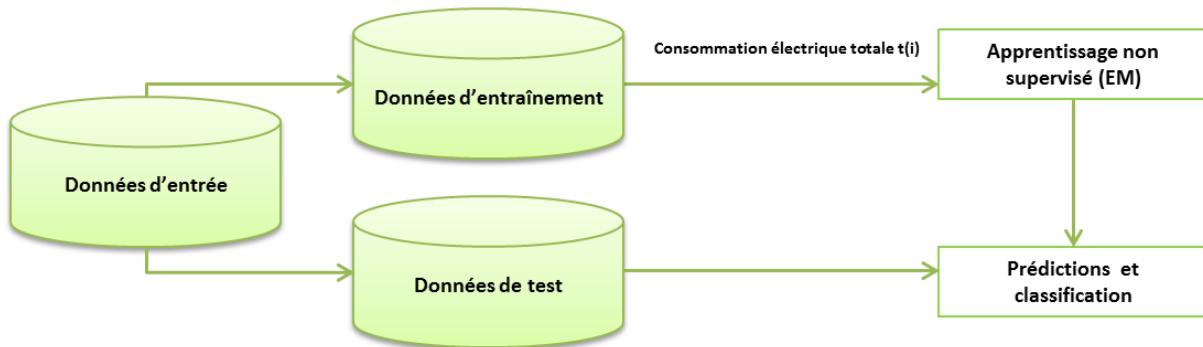


Figure 58 : Méthodes de classification des données pour la consommation électrique totale

En sortie, chaque ligne de notre vecteur d'entrée est labellisé par une classe représentée par une gaussienne avec une moyenne et une déviation standard comme décrit en exemple dans le tableau 15.

Par exemple, nous avons pour le secteur résidentiel 6 classes en sortie de l'apprentissage décrites dans le tableau 15. L'objectif de notre outil sera également de prédire dans quelle classe se

trouvera la consommation électrique l'heure suivante. Un historique des consommations et des données météorologiques est créé par heure pour chacune des variables d'entrée caractérisées :

$$E_{\text{Consommation_électrique_totale}} t(i) = [\text{Date} ;$$

$$\text{Statistiques Puissance Phase 1 } t(i), \dots, \text{Statistiques Puissance Phase 1 } t(i - 48),$$

$$\text{Statistiques Variations Phase 1 } t(i), \dots, \text{Statistiques Variations Phase 1 } t(i - 48)$$

$$\text{DonnéesRéelles_MétéoStation } t(i), \dots, \text{DonnéesRéelles_MétéoStation } t(i - 48),$$

$$\text{DonnéesPrédites_MétéoStation } t(i + 1), \dots, \text{DonnéesPrédites_MétéoStation } t(i - 48)$$

$$\text{Année}_j, \text{Mois}_m, \text{Jour}_j, \text{Jour_Semaine}_s, \text{Heure}_h, \text{Jour fériés}_{0,1}, \text{Vacances scolaires}_{0,1},$$

$$\text{Classes}_c] \quad (4.21)$$

Energie à prédire	Sortie apprentissage	Moyenne	Déviati on standard	Répartition %
Classification de l'énergie électrique, Secteur résidentiel	0	0.5817	0.0504	30
	1	0.821	0.1371	11
	2	1.2571	0.2616	9
	3	2.0145	0.3985	11
	4	3.0272	0.1062	5
	5	3.3657	0.5458	35

Tableau 15 : Répartition des différentes familles caractérisant la consommation électrique dans nos différents cas d'étude

La production-décentralisée-solaire

A partir de l'énergie produite, de la luminosité locale à Sierre et de la luminosité régionale à Sion distante de 15 km du site de test, nous réalisons un apprentissage non supervisé (Figure 53). L'algorithme converge vers 11 classes décrites dans le tableau 16.

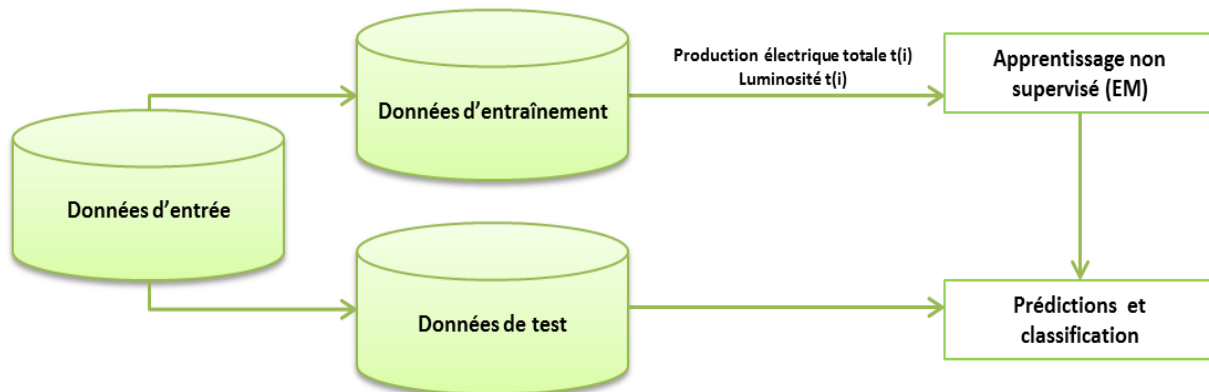


Figure 59 : Méthodes de classification des données pour la production décentralisée solaire

Energie à prédire	Sortie apprentissage	Moyenne	Déviati on standard	Répartiti on %
Production décentralisée solaire	0	1.0798	0.0794	3
	1	2.9921	1.3607	6
	2	7.8636	3.8702	9
	3	15.5819	6.9769	5
	4	21.4828	9.7723	5
	5	46.0692	14.7446	35
	6	46.7811	15.2881	16
	7	84.1565	12.5396	7
	8	101.8406	24.4283	17
	9	140.2383	13.679	9
	10	151.414	9.0872	4

Tableau 16 : Classes décrites par une moyenne, une déviati on standard et leurs répartiti ons dans le set d'entraîn ement caractérisant la production solaire en kWh

Nous regroupons les familles les plus proches pour faciliter la classification. Nous regroupons les 11 familles en 5 familles que nous avons décrites dans le tableau 15 et schématisé par la figure 54.

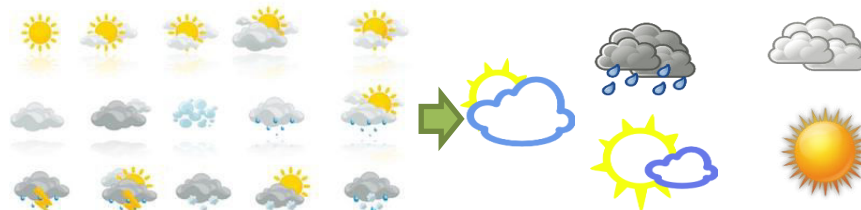


Figure 60 : Représentation du regroupement des classes dans la classification de la production décentralisée solaire

Energie à prédire	Sortie apprentissage	Moyenne	Déviati on standard	Répartiti on %
Production décentralisée solaire	0	1.0798	0.0794	18
		2.9921	1.3607	
		7.8636	3.8702	
	1	15.5819	6.9769	10
		21.4828	9.7723	
	2	46.0692	14.7446	24
		46.7811	15.2881	
	3	84.1565	12.5396	24
		101.8406	24.4283	
	4	140.2383	13.679	13
		151.414	9.0872	

Tableau 17 : Nouvelles classes décrites par une moyenne, une déviati on standard et leurs répartiti ons dans le set d'entraîn ement

Pour la classification, nous devons nous assurer d'une réparti ti on équitabl e entre les différentes classes en particulier celle qui requiert un intérêt particulier comme par exemple le pic de puissance de

consommation ou de production. Dans le cas de la production solaire, nous remarquons par exemple que les heures de la classe 10 ne représentent que 4% des données d'entraînement mais 10 % après regroupement. Sans ce regroupement, une étape nommée « bootstrap » est nécessaire. Elle permet de recopier une partie des données sous représentés pour permettre à nos algorithmes itératifs d'avoir une part plus équitable dans les données d'entraînement. Nous comprenons que ces étapes sont particulièrement adaptées à notre problématique de prédictions de variations brutales de la consommation ou de la production et que des algorithmes comme le AdaBoost ou le Gradient Boosting Tree décrit dans le chapitre 2 qui intègre ces mécanismes automatiquement seraient particulièrement adaptés.

En sortie de la classification, chaque heure de notre vecteur d'entrée est caractérisée par une classe et un historique des données de productions et météorologiques sont également créés :

$$\begin{aligned}
 E_{ENRE} \ t(i + 1) = [& \text{Date ;} \\
 & \text{Statistiques Puissance}_{\text{Phase 1}} \ t(i), \dots, \text{Statistiques Puissance}_{\text{Phase 1}} \ t(i - 48), \\
 & \text{Statistiques Variations}_{\text{Phase 1}} \ t(i), \dots, \text{Statistiques Variations}_{\text{Phase 1}} \ t(i - 48), \\
 & \text{Données}_{\text{Réelles_MétéoStation-Local}} \ t(i), \dots, \text{Données}_{\text{Réelles_MétéoStation-Local}} \ t(i - 48), \\
 & \text{Données}_{\text{Réelles_MétéoStation}} \ t(i), \dots, \text{Données}_{\text{Réelles_MétéoStation}} \ t(i - 48), \\
 & \text{Données}_{\text{Prédites_MétéoStation}} \ t(i + 1), \dots, \text{Données}_{\text{Prédites_MétéoStation}} \ t(i - 48), \\
 & \text{Année}_j, \text{Mois}_m, \text{Jour}_j, \text{Semaine}_{0,1}, \text{Heure}_h, \\
 & \text{Classes}_c]
 \end{aligned}
 \tag{4.22}$$

4.5 Réduction du nombre de variables

Comme énoncé dans le préambule, l'une des questions liée à la prédiction énergétique de flux est le nombre de données et ainsi le nombre de capteurs dont nous avons besoin pour maintenir un niveau acceptable de prédictions.

En sortie des différentes implémentations des arbres de décision, il est possible de sortir les variables qui sont utilisées pour la construction des arbres. Ainsi, nous calculons pour chaque flux énergétique le nombre de fois qu'une variable a été utilisée pour la création des ensembles d'arbres de décision au niveau 1, au niveau 2 et au niveau 3 :

$$\text{Evaluation} = \text{Nb_Utilisation}_{\text{Niveau1}} + \text{Nb_Utilisation}_{\text{Niveau2}} + \text{Nb_Utilisation}_{\text{Niveau3}}
 \tag{4.23}$$

Un classement par ordre décroissant est ensuite réalisé et le calcul prédictif par flux sera réalisé en faisant diminuer le nombre de variables basées selon leurs importances dans la construction des arbres.

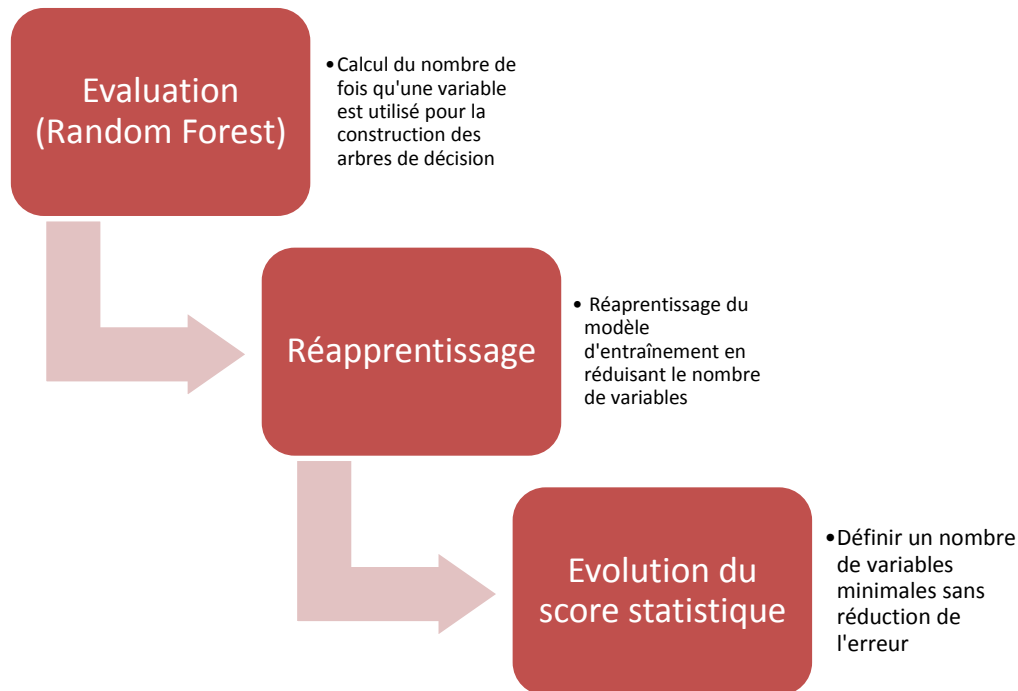


Figure 61 : Méthode d'analyse de l'impact du nombre de variables d'entrée sur la précision de la prédiction

4.6 Conclusion

Nous avons défini le traitement en sortie des différents capteurs et l'agrégation réalisée des caractéristiques tirées des capteurs pour prédire nos différents flux énergétiques.

Nous pouvons repartir de **la problématique globale pour la prédiction de n'importe quel flux énergétique** :

- ⇒ **Quelles données ?** Nous avons défini les données d'entrée par flux énergétique avec l'étude des variations de puissance à la seconde et les statistiques liés directement à la puissance. Pour le cas de la consommation électrique dans le secteur résidentiel, une pré-étape d'identification non intrusive des appareils est réalisée. Cette identification permet la prédiction de l'eau chaude et du chauffage de manière non intrusive également. Enfin, les données météorologiques par le biais de 4 stations cantonales sont corrigées et utilisés également comme données d'entrée.
- ⇒ **À quelles fréquences de collecte ?** L'impact de la collecte des données à la minute ou à l'heure pour une prédiction pour l'heure suivante est mesuré.
- ⇒ **Avec quel système d'information (appareil, stockage...)?** Nous utilisons différents connecteurs avec du modbus ou du Zigbee comme protocole de communication pour collecter les données à la seconde. Elles sont ensuite stockées dans des bases de données MongoDB. Les données météorologiques sont accessibles par API grâce à l'application de MétéoSuisse. Elles sont stockées dans une autre base de données, influx DB.

- ⇒ **Avec quelles méthodes de prédictions ?** Des méthodes linéaires et non linéaires par apprentissage sont testées.
- ⇒ **Avec quels modèles mathématiques ?** Des algorithmes linéaires (Régression linéaire, ARIMA) et non linéaires (MLP, PNN, Random Forest et Gradient Boosted Tree) sont testés.
- ⇒ **Pour quels niveaux de prédictions ?** Les résultats sont présentés dans le chapitre 5.
- ⇒ **Sur quel horizon ?** Nous avons choisi comme horizon de prédiction l'heure suivante pour réaliser notre bilan énergétique prédictif.

Chapitre 5 : Présentation des résultats

5.1 Introduction

Pour permettre une généralisation massive des microgrids, il nous faut limiter le besoin d'information a priori sur le système. Puisque la présence humaine représente une des composantes essentielles dans la prédiction [VIC-2014], nous avons identifié des plannings de production en particulier dans les secteurs tertiaire et industriel.

Les chapitres précédents nous ont permis d'obtenir des bases de données conséquentes pour caractériser le besoin en énergie de différents systèmes dans des secteurs d'activité différents. Ce besoin en énergie a été collecté au niveau du compteur global d'électricité qui caractérise différents flux d'énergie selon les systèmes étudiés.

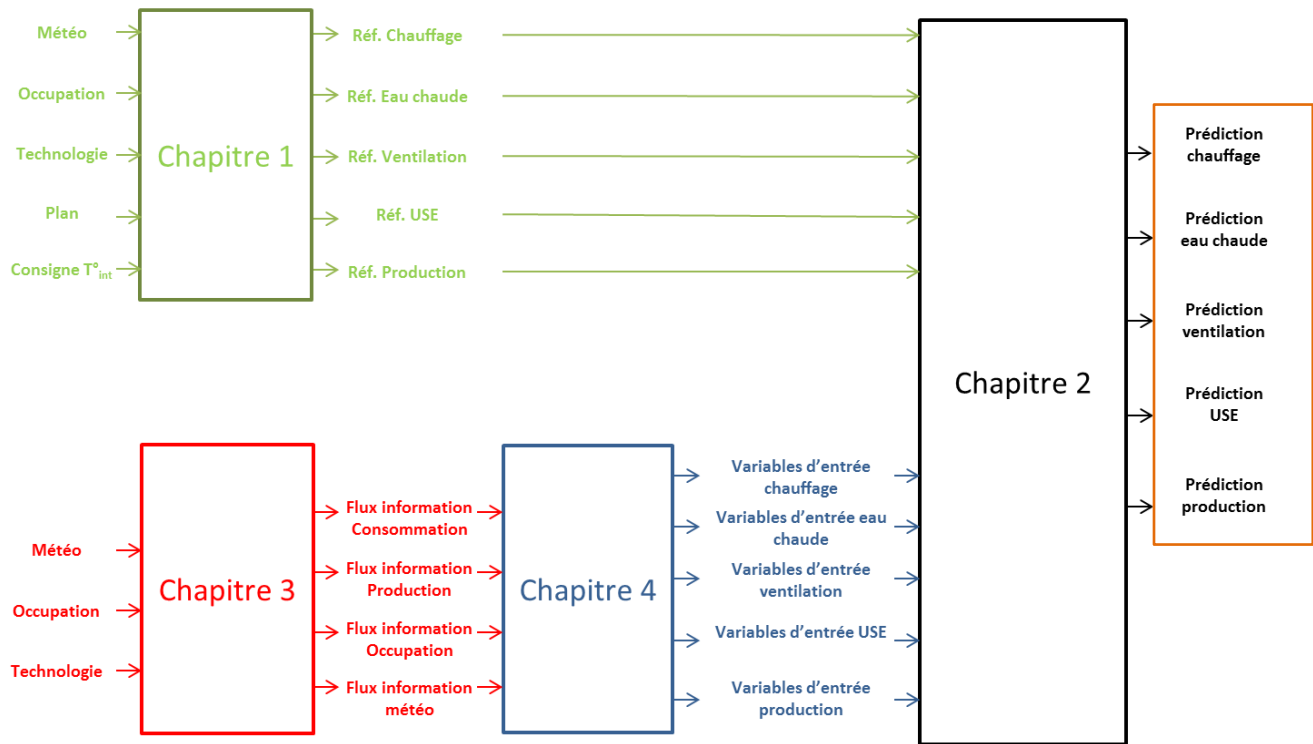


Figure 62 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision

5.2 Prédire les besoins en énergie dans le secteur résidentiel

5.2.1 Présentation du cas d'étude

Il s'agit d'une maison construite en 2004 d'une surface de 250m² comprenant deux étages de 125m² chacun représentée à la figure 63. Elle se situe à Chamoson en Suisse dans le canton du Valais à une altitude de 730 mètres. Elle est construite en béton armé et briques et est isolée thermiquement par l'extérieur, aussi bien au niveau des murs que du plancher bas et de la toiture.



Figure 63 : Schématisation de la maison test sous Pleiades COMFIE

La maison est chauffée par un système de plancher chauffant dont le fluide caloporteur est chauffé par une pompe-à-chaleur (PAC) de type air-eau de puissance nominale 3kW. Si la PAC ne suffit pas à subvenir aux besoins de chauffage, un poêle à bois permet de compléter les apports de chaleur. Le rendement de la PAC (COP) fourni est de 3.5 en données constructeur. Les pompes à chaleur sont préprogrammées par l'utilisateur de la maison. Au niveau de la pompe à chaleur liée au chauffage, une consigne de température de +2 degrés est insérée entre minuit et 6h du matin du lundi au vendredi. Pour le week-end, aucune consigne n'est donnée et la pompe à chaleur a un fonctionnement normal suivant une température de consigne minimale de 21 °C.

L'eau chaude sanitaire est produite par une PAC qui chauffe de l'eau ensuite stockée dans un ballon de 300 litres. Son rendement est de 3.7. L'air chaud extrait du local cellier, très bien isolé car contenant plusieurs appareils générant beaucoup de chaleur, permet de chauffer le fluide caloporteur.

La maison possède en toiture une centrale solaire photovoltaïque (PV) de 68m² de panneaux de puissance nominale 9.5kWc. Enfin, la famille dispose des usages spécifiques électriques suivants :

- ⇒ un four électrique
- ⇒ une cuisinière électrique
- ⇒ deux congélateurs
- ⇒ un réfrigérateur
- ⇒ une machine à café électrique et une bouilloire
- ⇒ un lave-vaisselle
- ⇒ un lave-linge
- ⇒ un sèche-linge
- ⇒ un serveur informatique
- ⇒ un rétroprojecteur
- ⇒ deux ordinateurs fixes
- ⇒ des enceintes
- ⇒ deux écrans

La consommation électrique globale pour notre maison test regroupe la consommation en chauffage, en eau chaude, en lumineaire et en usage spécifique électrique (Figure 57).

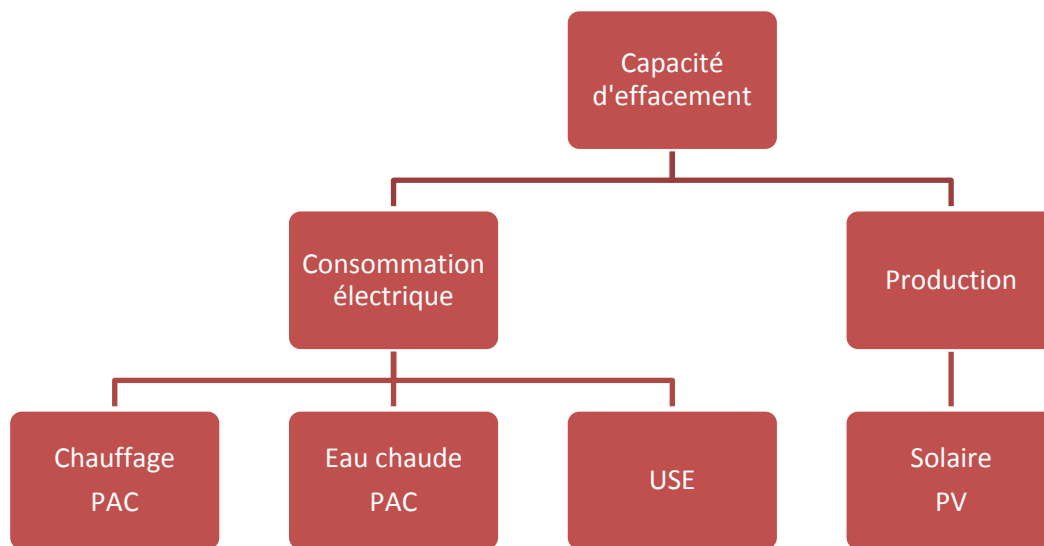


Figure 64 : Caractérisation de la capacité d'effacement qui nécessite la prédiction du chauffage, de l'eau chaude sanitaire, de la consommation électrique globale et de la production solaire

5.2.2 Prédire la consommation électrique globale

Dans [ZHA-2016], la série temporelle de consommation est prédite à partir d'une méthode hybride qui utilise une régression suivie d'une méthode d'apprentissage avec l'algorithme SVM pour Support Vector Machine. Des désagrégations de courbes de charges globales sont également réalisées

pour identifier les appareils électriques pour une meilleure prédiction de la consommation électrique totale [KOL-2012] [PAR-2011].

Les modèles de prédiction de consommation sont très importants, les échelles de temps sont multiples et les prédictions sont réalisées à l'heure, à la journée, au mois, à l'année, etc. [REB-2015] [MOU-2016] [CHE-2016]. Nous avons choisi une prédiction à l'heure avec différentes méthodes et algorithmes différents avec aucune information a priori dans le cas du secteur résidentiel.

Données d'entrée

Dans cette étude, la consommation électrique globale pour notre maison test regroupe la consommation en chauffage, en eau chaude, en luminaires et les usages spécifiques électriques. Puisque nous sommes sur du tout électrique avec une pompe à chaleur pour le chauffage et une autre pour l'eau chaude sanitaire, nous pouvons convertir les énergies en puissance. Puisque la ventilation est manuelle et en estimant l'énergie lumineuse comme nulle, nous avons à chaque instant $t(i)$:

$$E t(i) = (P_{\text{Chauffage}} + P_{\text{Eau chaude}} + P_{\text{USE}}) t(i)$$

Notre vecteur d'entrée qui caractérise la consommation électrique globale pour notre maison test regroupe les statistiques liées à la puissance et aux différentes variations sur chacune des phases du compteur.

Les données météorologiques sont données par les valeurs réelles et prédites horaires des stations entourant notre cas d'étude.

La temporalité de notre signal est gardée et délimitée par les variables relatives au mois, au jour, au jour de la semaine et à l'heure.

Pour les résultats de la classification, chaque heure est labélisée par une classe c où il s'agira à notre outil de définir l'heure suivante dans quelle classe se situe la consommation électrique :

$$E_{\text{Consommation_électrique_totale}} t(i) = [\text{Date}, \\ \text{Statistiques Puissance}_{\text{Phase 1}} t(i), \dots, \text{Statistiques Puissance}_{\text{Phase 1}} t(i - 48), \\ \text{Statistiques Variations}_{\text{Phase 1}} t(i), \dots, \text{Statistiques Variations}_{\text{Phase 1}} t(i - 48), \\ \text{Données Réelles_Météo}_{\text{Station}} t(i), \dots, \text{Données Réelles_Météo}_{\text{Station}} t(i - 48), \\ \text{Données Prédites_Météo}_{\text{Station}} t(i + 1), \dots, \text{Données Prédites_Météo}_{\text{Station}} t(i - 48), \\ \text{Année}_j, \text{Mois}_m, \text{Jour}_j, \text{Jour_Semaine}_j, \text{Heure}_h, \text{Jour fériés}_{0,1}, \text{Vacances scolaires}_{0,1}, \\ \text{Classes}_c]$$

L'année nous permet de délimiter les heures utilisées pour l'entraînement et le test. L'ensemble de notre set de données utilisé pour présenter nos résultats de prédictions représente les heures du 1 Janvier 2013 jusqu'au 31 Mars 2014. Cela représente 13471 heures.

Les données depuis le mois de Janvier 2013 à Août 2014 pour entraîner nos modèles et les mois de Février et de Mars 2014 sont les deux mois de test (Figure 65). Nous avons sélectionné ces deux mois

de test pour inclure le chauffage de la maison qui s'étale sur une période dépendant de la température extérieure mais comprise généralement entre les mois de Septembre à Mars.

Nous utilisons les méthodes décrites dans le chapitre 2 pour modéliser la consommation électrique à partir des différentes variables tirées de nos bases de données.

Pour caractériser la présence humaine, deux modèles différents ont été créés à partir des jours de la semaine.

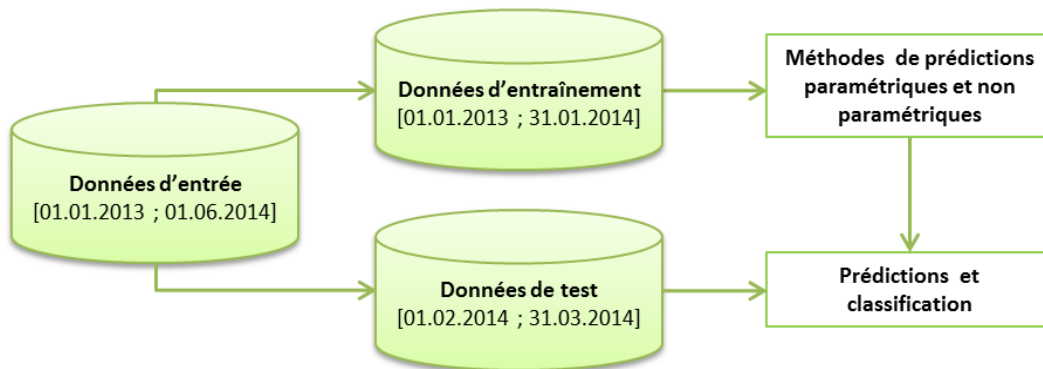


Figure 65 : Répartition des données pour la prédiction de la consommation électrique globale, secteur résidentiel

Résultats de la prédiction

Les résultats au niveau de la prédiction de consommation électrique globale sont très encourageants avec une MAE de 0.123 kWh ce qui est faible puisque la consommation électrique globale comprend une PAC liée au chauffage et une autre PAC liée à l'eau chaude de plus de 3 kW de puissance nominale. Les ensembles d'arbres de décision avec le RF (Random Forest) et le GBT (Gradient Boosted Tree) fournissent les meilleurs résultats. Parmi ces deux algorithmes, le GBT nous fournit la meilleure précision pour les deux mois de test (Tableau 18).

Modèles	MAE	MSE	RMSE	MSD	R ²
MLP	0.25 +/- 0.01	0.065	0.25 +/- 0.01	0.134	0.291
PNN	0.26 +/- 0.01	0.074	0.30 +/- 0.01	0.14	0.264
Régression linéaire	0.35 +/- 0.01	0.083	0.35 +/- 0.01	0.157	0.216
Random Forest regression	0.16 +/- 0.01	0.048	0.22 +/- 0.01	0.05	0.338
Gradient Boosted Tree	0.12 +/- 0.01	0.042	0.21 +/- 0.01	0.011	0.419

Tableau 18 : Résultats statistiques en sortie des différents algorithmes pour la prédiction de la consommation électrique globale pour le secteur résidentiel

Résultats de la classification

Pour le modèle jour de la semaine, 8602 heures sont utilisées pour entraîner le modèle par un tirage aléatoire. 1029 heures représentent notre set de test que nous devons prédire le plus précisément possible. En sortie de l'apprentissage non supervisé, la consommation électrique globale dans le secteur

résidentiel est représentée par six classes pour caractériser la consommation électrique globale durant les jours de la semaine (Tableau 19).

Pour le modèle caractérisant le week-end, 3456 heures sont utilisées pour entraîner le modèle par un tirage aléatoire. 384 heures représentent notre set de test que nous devons prédire le plus précisément possible. En sortie de l'apprentissage non supervisé, la consommation électrique globale dans le secteur résidentiel est représentée par quatre classes pour les jours de la semaine (Tableau 20).

Energie à prédire	Sortie apprentissage	Moyenne	Déviations standard	Répartition %
Classification de l'énergie électrique, Secteur résidentiel, Semaine	0	0.56	0.06	35
	1	0.81	0.14	34
	2	1.25	0.25	10
	3	1.97	0.39	8
	4	3.03	0.11	4
	5	3.33	0.57	10

Tableau 19 : Classes pour caractériser la consommation électrique globale pour les jours de la semaine pour le secteur résidentiel en kWh

Energie à prédire	Sortie apprentissage	Moyenne	Déviations standard	Répartition %
Classification de l'énergie électrique, Secteur résidentiel, Week-end	0	0.57	0.06	31
	1	0.86	0.16	36
	2	1.47	0.39	16
	3	3.10	0.68	18

Tableau 20 : Classes pour caractériser la consommation électrique globale dans le secteur résidentiel pour le week-end en kWh

Comme pour les résultats de prédictions, le RF et le GBT nous fournissent les meilleurs scores de classification avec plus de 98% de classification réussie pour le GBT. Les SVM nous fournissent également de très bons résultats et une analyse combinatoire de ces trois algorithmes pourraient augmenter encore notre score de classification. Nous ne voyons pas de différence entre le modèle caractérisant les jours de la semaine et celui pour le samedi dimanche dans nos résultats de classification (Tableau 21).

Modèles	Taux de classification semaine	Taux de classification Week-end
MLP	85.35 +/- 2.4	84.51 +/- 2.3
PNN	88.84 +/- 2.6	86.84 +/- 2.4
SVM	92.57 +/- 1.8	91.42 +/- 1.8
Random Forest	96.09 +/- 1.4	95.72 +/- 1.6
Gradient Boosted Tree	98.06 +/- 1.2	97.13 +/- 1.5

Tableau 21 : Résultats de la classification de la consommation électrique globale dans le secteur résidentiel

Impact de la collecte de données à la minute ou à l'heure

Nous testons ici la différence dans nos résultats entre **les données à la seconde agrégées à la minute et une donnée par heure** pour aller prédire l'heure suivante. Contrairement au test précédent où nous avons accès aux statistiques sur la puissance et sur les variations, nous utilisons dans ce test seulement l'historique de la consommation électrique totale par heure.

L'erreur y est 5 fois supérieure entre une collecte des données à la minute et à l'heure pour la prédiction de la consommation électrique globale. Pour un pilotage local des flux énergétiques et un besoin d'autoconsommation, une étude plus fine des puissances est ainsi à préconiser si cette erreur moyenne de 0.526 kWh n'est pas compensée par un stockage électrique (Tableau 22). L'étude des variations de la puissance caractérise des appareils ou des états d'appareils et permet d'apporter des connaissances perdues dans une consommation moyennée à l'heure. De plus, il nous devient impossible d'identifier des appareils électriques de manière non intrusive avec des données à la minute, au quart d'heure ou à l'heure de manière précise.

Modèles	Collecte	MAE	MSE	RMSE	MSD	R ²
Gradient Boosted Tree	1 min	0.123 +/- 0.01	0.042	0.205	0.011	0.419
Gradient Boosted Tree	1 heure	0.526 +/- 0.01	0.533	0.73	-0.003	0.184

Tableau 22 : Résultats statistiques en sortie du GBT en fonction de deux fréquences de mesures pour la prédiction de la consommation électrique globale pour le secteur résidentiel

Impact du nombre de variables d'entrée

Dans ce test, notre objectif est de déterminer l'impact du nombre de variables sur la qualité de nos résultats. Pour cela, nous utilisons les statistiques en sortie de l'algorithme RF qui nous donne pour chacune des variables d'entrée le nombre de fois qu'elle a été utilisée pour construire les trois premiers niveaux des arbres de décision. Notre vœux est de déterminer le nombre minimal de variables à niveau de prédiction équivalent et nous remarquons une dégradation significative des résultats entre l'utilisation de 10 et de 5 variables (Tableau 23).

Nombre de variables	MAE	MSE	RMSE	MSD	R ²
1135	0.043	0.008	0.092	0.013	0.819
800	0.044	0.009	0.095	0.012	0.807
400	0.046	0.009	0.097	0.018	0.8
200	0.047	0.01	0.101	0.012	0.782
100	0.051	0.011	0.105	0.015	0.763
50	0.046	0.09	0.097	0.018	0.8
10	0.045	0.014	0.117	0.001	0.706

5	0.053	0.016	0.125	0.01	0.666
---	-------	-------	-------	------	-------

Tableau 23 : Résultats statistiques en sortie du GBT en fonction du nombre de variables pour la prédiction de la consommation électrique globale pour le secteur résidentiel

5.2.3 Prédire le besoin en chauffage

Les premières études pour comprendre le comportement du bâtiment dans un environnement donné datent des années 1970. A cette période, la communauté scientifique et les énergéticiens ont été confrontés à un manque d'informations détaillé sur la constitution des bâtiments, une incertitude sur les usages et le comportement des occupants et des capacités de calcul limitées.

Aujourd'hui, l'information sur les bâtiments peut-être connue, principalement sur les logements neufs. Concernant les logements en cours de rénovation, des approximations peuvent être réalisées. Les capacités dédiées pour le calcul ne sont plus un problème. Des modèles d'identification ont en effet fait l'objet de nombreuses recherches toujours en cours actuellement. Nous pouvons citer l'un des plus connus qui consiste à modéliser le bâtiment par un circuit électrique équivalent simple [ARB-2013]. Par exemple, pour le modèle R6C2, les données d'entrée du système sont :

- ⇒ le profil d'occupation
- ⇒ la puissance maximale
- ⇒ les apports internes dus au soleil
- ⇒ les apports solaires sur les surfaces opaques
- ⇒ la température extérieure

De plus, pour l'identification des paramètres lors de l'apprentissage, il est nécessaire de disposer de la température intérieure moyenne et de la puissance de chauffage injectée dans le bâtiment, en plus des données précédentes. La mise en place du modèle R6C2 sur un site nécessite plusieurs étapes :

- ⇒ Une instrumentation fine pour la mesure des températures intérieures dans chaque zone thermique du bâtiment
- ⇒ La mesure de la puissance de climatisation au niveau du réseau secondaire ou des émetteurs (mesure des besoins thermiques)
- ⇒ Une connaissance des systèmes d'émission pour les intégrer au modèle R6C2
- ⇒ La mesure de la consommation électrique des prises électriques pour le calcul du profil d'occupation
- ⇒ L'obtention de toutes les entrées nécessaires au modèle de calcul de gain solaire (nébulosité, facteur de trouble de Linke)
- ⇒ Les prévisions de température extérieure et de nébulosité

Dans cette thèse, nous avons voulu, dans le cas de la prédiction du chauffage dans le secteur résidentiel, créer un outil totalement non intrusif où nous n'avons pas accès à des données météorologiques locales comme la température intérieure ou extérieure ou encore la luminosité locale.

Données d'entrée

Notre vecteur d'entrée qui caractérise le besoin en chauffage pour notre maison test regroupe les statistiques liées à la puissance et aux différentes variations sur chacune des phases du compteur.

Les données météorologiques sont données par les valeurs réelles et prédites horaires des stations entourant notre cas d'étude.

La temporalité de notre signal est gardée et délimitée par les variables m , j , et h . Pour les résultats de la classification, chaque heure est labélisée par une classe qui correspond à l'allumage ou à l'arrêt de la PAC :

$$E_{\text{Chauffage}} t(i + 1) = [\text{Date} ;$$

Statistiques Puissance Phase 1 $t(i), \dots, \text{Statistiques Puissance phase 1 } t(i - 48)$,
 Statistiques Variations Phase 1 $t(i), \dots, \text{Statistiques Variations phase 1 } t(i - 48)$
 Statistiques Puissance Phase 2 $t(i), \dots, \text{Statistiques Puissance phase 2 } t(i - 48)$,
 Statistiques Variations Phase 2 $t(i), \dots, \text{Statistiques Variations phase 2 } t(i - 48)$,
 Statistiques Puissance Phase 3 $t(i), \dots, \text{Statistiques Puissance phase 3 } t(i - 48)$,
 Statistiques Variations Phase 3 $t(i), \dots, \text{Statistiques Variations phase 3 } t(i - 48)$,
 Données Réelles _MétéoStation $t(i), \dots, \text{Données Réelles _MétéoStation } t(i - 48)$,
 Données Prédites _MétéoStation $t(i + 1), \dots, \text{Données Prédites _MétéoStation } t(i - 48)$

Année $_j$, Mois $_m$, Jour $_j$, Semaine $_{0,1}$, Heure $_h$, Jour fériés $_{0,1}$, Vacances scolaires $_{0,1}$,

Classes $_{\text{ON/OFF}}$]

Les données d'entrée représentent les heures depuis le mois de janvier 2013 à mars 2014 pour entraîner nos modèles. Les mois de février et de mars 2014 sont les deux mois de test (Figure 66). Les données d'entraînement représentent 5687 heures. Les données représentent 1487 à prédire et classifier le plus précisément possible.

La saison d'été est enlevée de notre set d'entraînement et correspond du 21 Juin 2013 au 21 septembre 2013. Un modèle est créé pour les jours de la semaine et un autre caractérisant le week-end.

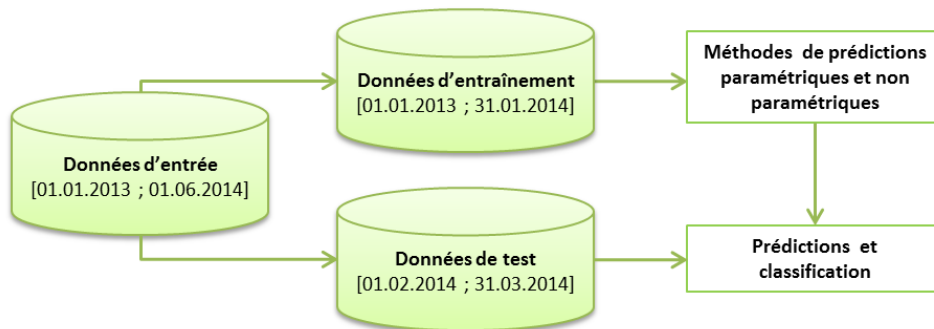


Figure 66 : Répartition des données pour la prédiction du besoin en chauffage

Résultats de la prédiction

Pour les jours de la semaine, le GBT nous fournit les meilleurs résultats avec une moyenne des écarts en valeur absolue y est de 0.228 kWh (tableau 24).

Ce résultat est encourageant mais il devrait être meilleur pour un pilotage réactif sûr en particulier au niveau des pics de chauffage où le modèle mathématique a du mal à les appréhender (Figure 67).

Modèles	MAE	MSE	RMSE	MSD	R ²
MLP	0.349 +/- 0.02	0.545	0.738 +/- 0.02	-0.294	0.533
PNN	0.331 +/- 0.02	0.524	0.716 +/- 0.02	-0.282	0.552
Régression linéaire	0.457 +/- 0.01	0.612	0.785 +/- 0.01	-0.321	0.416
ARIMA	0.421 +/- 0.01	0.415	0.816 +/- 0.01	-0.342	0.452
Random Forest regression	0.262 +/- 0.01	0.216	0.464 +/- 0.01	-0.014	0.736
Gradient Boosted Tree	0.228 +/- 0.01	0.212	0.46 +/- 0.01	-0.029	0.74

Tableau 24 : Résultats numériques par modèle mathématique pour une prédiction du besoin en chauffage à $t + 1$ heure pour les jours de la semaine

Modèles	Taux de classification	Déviations standard
MLP	96.6 +/- 0.5	0.7
PNN	94.5 +/- 0.4	0.9
SVM	95.3 +/- 0.6	1.1
Random Forest	97.4 +/- 0.4	0.7
Gradient Boosted Tree	98.6 +/- 0.6	0.9

Tableau 25 : Résultats de la classification représentant le chauffage pour les jours de la semaine

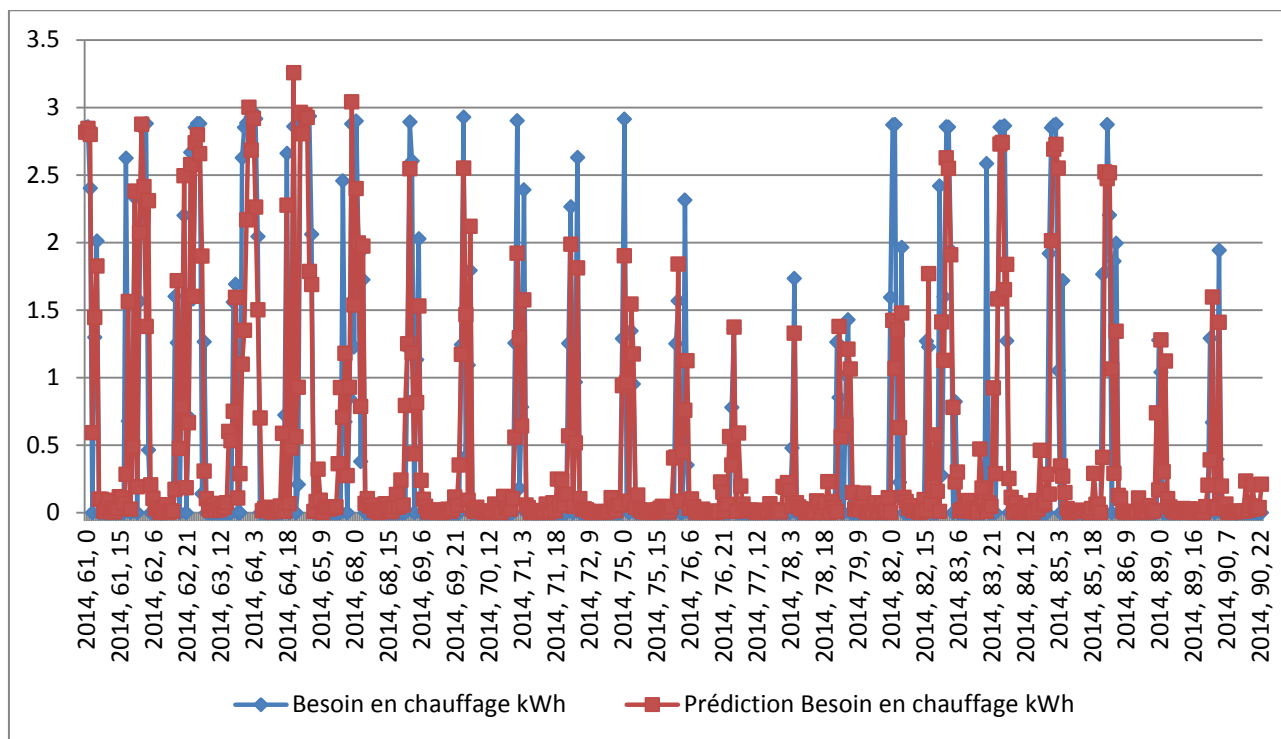


Figure 67 : Besoin et prédiction en chauffage en sortie du GBT pour les jours de la semaine (Année, jour de l'année, heure)

Pour les jours du week-end, l'erreur de précision de la prédiction diminue alors que le système de chauffage ne suit plus de planning de production décidé par l'utilisateur. La moyenne des écarts en valeur absolue y est de 0.16 kWh (Tableau 25). Nous remarquons que l'outil appréhende plus facilement les pics d'énergie liés au chauffage mais les faibles variations d'énergie sont mal représentées. Les hautes variations d'énergie correspondant aux différents pics de puissances sont à quant à eux correctement prédits (Figure 68).

Algorithme	Modèles	MAE	MSE	RMSE	MSD	R ²
Gradient Boosted Tree	Semaine	0.228 +/- 0.01	0.212	0.46	-0.029	0.74
	Week-end	0.16 +/- 0.01	0.101	0.317	-0.006	0.841

Tableau 26 : Résultats statistiques de la prédiction du besoin en chauffage pour le modèle semaine et week-end

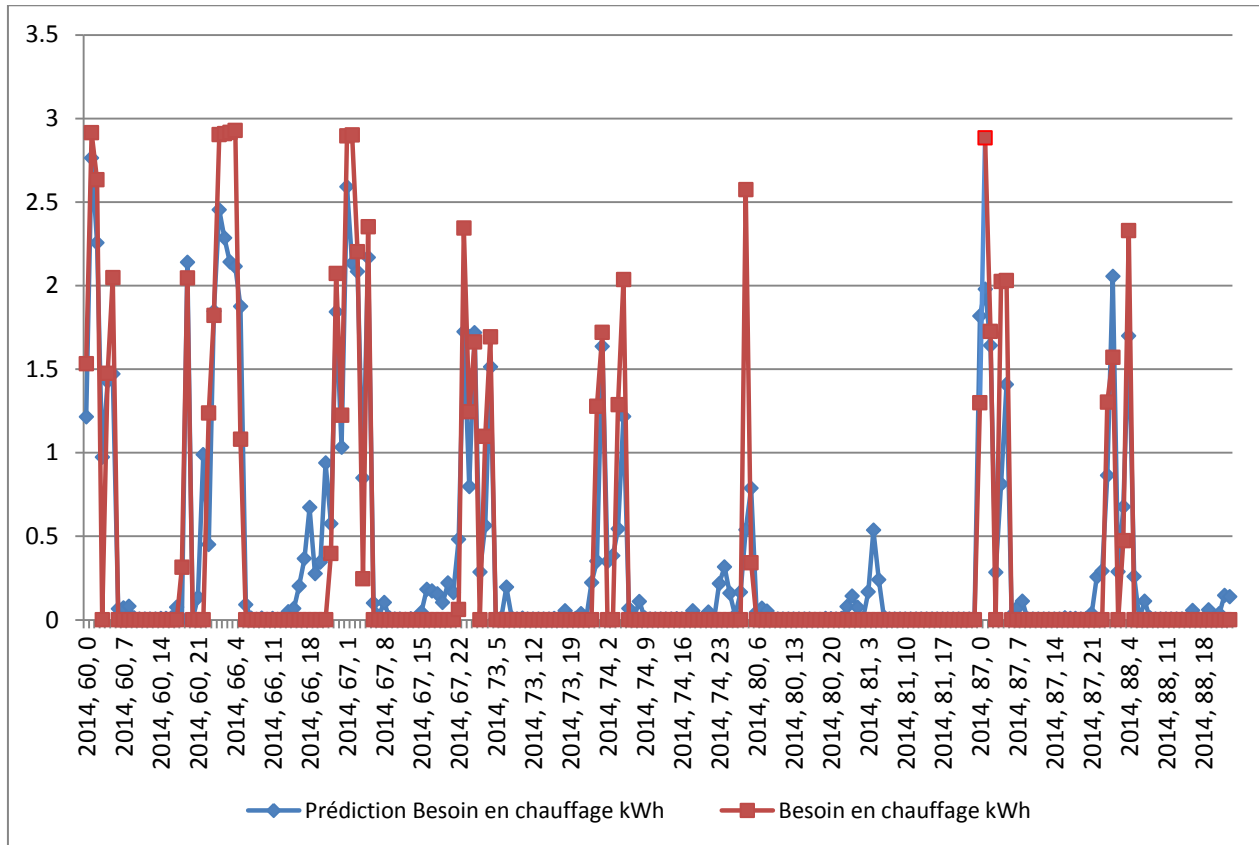


Figure 68 : Besoin en chauffage en kWh et la prédiction en sortie du GBT sur nos données de test pour les jours du week-end

Pour la classification binaire du besoin en chauffage, 358 heures sont labellisées OFF et 242 sont labellisées ON pour le set de test.

98,6 % de ces heures ont été correctement classifiées par le GBT (Tableau 27).

Modèles	Taux de classification	Deviation standard
MLP	97.6 +/- 0.3	1.1
PNN	96.5 +/- 0.4	1.4
SVM	97.3 +/- 0.6	0.9
Random Forest	98.2 +/- 0.4	1.1
Gradient Boosted Tree	98.6 +/- 0.3	1.1

Tableau 27 : Résultats de la classification représentant le chauffage pour les jours du week-end

Impact de la collecte de données à la minute ou à l'heure

Nous testons également la différence dans nos résultats entre des données collectées à la minute et une donnée moyennée par heure pour aller prédire l'heure suivante. Nous avons un facteur deux, c'est-à-dire que l'erreur moyenne des écarts augmente de 0.2 kWh (Tableau 28).

Algorithmes	Fréquence	MAE	MSE	RMSE	R ²
Gradient Boosted Tree	1 min	0.228 +/- 0.01	0.239	0.488 +/- 0.01	0.719
	1h	0.468 +/- 0.02	0.481	0.693 +/- 0.01	0.626

Tableau 28 : Résultats numériques par modèle mathématique pour une prédiction du besoin en chauffage à t + 1 heure pour le week-end

Impact du nombre de variables d'entrée

Nous faisons évoluer notre nombre de variables pour y évaluer l'impact sur nos résultats et nous remarquons une dégradation significative de l'erreur à partir de 50 variables pour les jours de la semaine (Tableau 29) tandis que 10 variables sont suffisantes pour le samedi et dimanche (Tableau 30).

Nombre de variables	MAE	MSE	RMSE	MSD	R ²
800	0.287	0.264	0.514	-0.031	0.795
400	0.288	0.263	0.513	-0.038	0.796
200	0.29	0.264	0.514	-0.031	0.795
100	0.301	0.302	0.549	-0.039	0.765
50	0.305	0.319	0.565	-0.035	0.752
10	0.467	0.643	0.802	-0.046	0.5

Tableau 29 : Résultats statistiques en sortie du GBT en fonction du nombre de variables pour la prédiction du besoin en chauffage pour le secteur résidentiel

Nombre de variables	MAE	MSE	RMSE	MSD	R ²
400	0.146	0.092	0.308	0.002	0.825
200	0.145	0.092	0.308	0.002	0.825
100	0.149	0.097	0.312	0.004	0.846
50	0.157	0.114	0.338	0.043	0.82
10	0.173	0.129	0.36	0.075	0.796

Tableau 30 : Résultats statistiques en sortie du GBT en fonction du nombre de variables pour la prédiction du besoin en chauffage pour le secteur résidentiel, modèle week-end

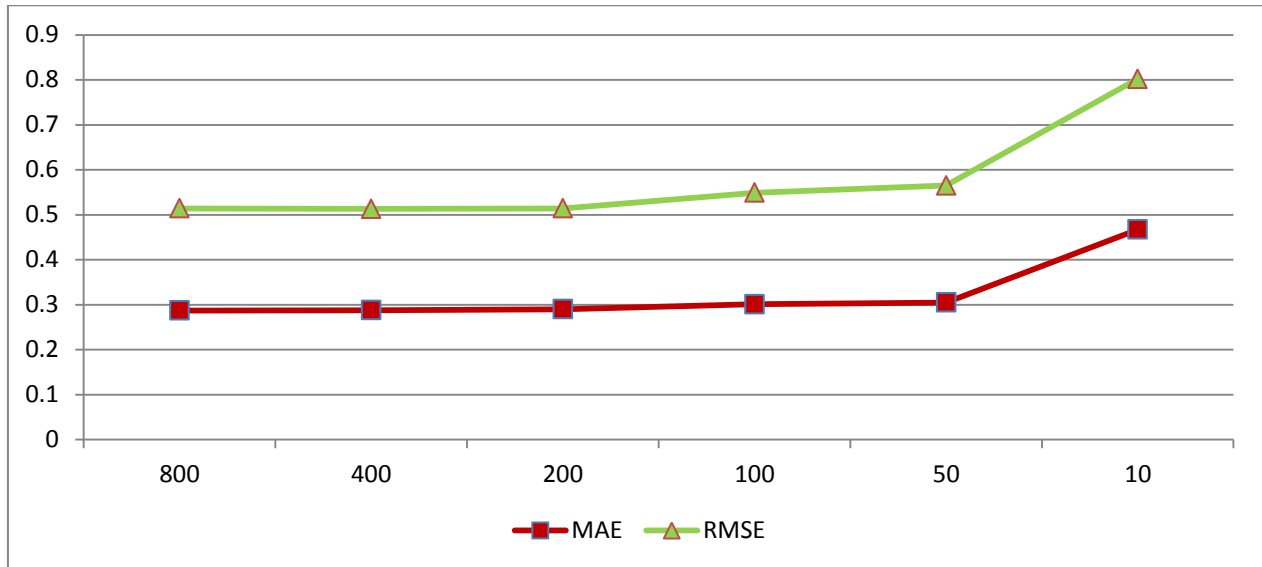


Figure 69 : Evolution de la précision de la prédiction en fonction du nombre de variables d'entrée pour la prédiction du besoin en chauffage

Les 10 variables les plus utilisées pour la construction des arbres de décision pour la prédiction du besoin en chauffage pour les jours de la semaine :

- ⇒ Consommation électrique totale t (-1)
- ⇒ Moyenne Puissance électrique totale t (-1)
- ⇒ Moyenne Puissance électrique totale t (-24)
- ⇒ Déviation standard Besoin en chauffage t (-24)
- ⇒ Déviation standard Puissance électrique totale t (-24)
- ⇒ Variance Puissance électrique totale t (-24)
- ⇒ Puissance électrique totale t (-24)
- ⇒ Besoin en chauffage t (-1)
- ⇒ Besoin en chauffage t (-24)
- ⇒ Maximum Variations puissance électrique Phase 2 t (-1)

Les 10 variables les plus utilisées pour la prédiction du besoin en chauffage le samedi et le dimanche sont :

- ⇒ Consommation électrique totale t (-1)
- ⇒ Heure
- ⇒ Déviation standard Puissance électrique totale t (-24)
- ⇒ Variance Puissance électrique totale t (-24)
- ⇒ Somme Puissance électrique totale t (-48)
- ⇒ Moyenne Variations Puissance électrique totale t (-1) Phase 1
- ⇒ Somme Variations Puissance électrique totale t (-1) Phase 1
- ⇒ Maximum Variations Puissance électrique totale t (-1) Phase 1
- ⇒ Maximum Variations Puissance électrique totale t (-1) Phase 2

⇒ Maximum Variations Puissance électrique totale t (-1) Phase 3

Nous pouvons remarquer que la température n'intervient pas dans les 10 premières variables énoncées. Ceci pourrait s'expliquer par la distribution de l'énergie de chauffage qui se fait par un plancher chauffant qui est un système très lent aux contraires des radiateurs électriques classiques [ARB-2013].

5.2.4 Prédire le besoin en eau chaude sanitaire

La consommation d'énergie pour l'eau chaude sanitaire est calculée à partir des besoins en énergie nécessaires pour fournir l'eau chaude au point de puisage divisés par le rendement du système. Les variables qui influencent le besoin en eau chaude sanitaire sont la fréquence d'utilisation en particulier dans l'habitat résidentiel et la température de consigne (généralement de 40°C pour un lavabo et de 60°C pour une machine à laver). Pour simplifier, certaines méthodes utilisent des valeurs basées sur des études statistiques qui dépendent du nombre d'occupants et de la surface de la zone chauffée [SAVE-2000].

Données d'entrée

Notre vecteur d'entrée qui caractérise le besoin en eau chaude sanitaire pour notre maison test regroupe les statistiques liées à la puissance et aux différentes variations de puissance. Les données météorologiques sont représentées par les valeurs réelles et prédites horaires des stations entourant notre cas d'étude. La temporalité de notre signal est gardée et délimitée par les variables m, j, et l'heure. Pour les résultats de la classification, chaque heure est labélisée par une classe qui correspond à l'allumage ou à l'arrêt de la PAC :

$$E_{ECS} t(i + 1) = [Date ;$$

Statistiques Puissance Phase 1 t(i),..., Statistiques Puissance Phase 1 t(i - 48),
 Statistiques Variations Phase 1 t(i),..., Statistiques Variations Phase 1 t(i - 48)
 DonnéesRéelles_MétéoStation t(i),..., DonnéesRéelles_MétéoStation t(i - 48),
 DonnéesPrédites_MétéoStation t(i + 1),..., DonnéesPrédites_MétéoStation t(i - 48)

Année_j, Mois_m, Jour_j, Semaine_{0,1}, Heure_h, Jour fériés_{0,1}, Vacances scolaires_{0,1},

Classes_{ON/OFF}]

L'ensemble du set de données représentent 552 heures après suppression des valeurs manquantes. Pour l'entraînement, 310 heures sont utilisées où la pompe à chaleur est éteinte et 313 heures où elle est allumée. Pour le test, 86 heures représentent l'arrêt de la pompe à chaleur, 25 heures représentent le fonctionnement de la pompe à chaleur pour l'eau chaude sanitaire. Nous utilisons les données depuis le mois de Février 2013 jusqu'à Août 2013. Les données sont tirées aléatoirement avec 80 % des données utilisées pour l'entraînement et 20% pour le test à partir duquel seront basés nos résultats (Figure 70).

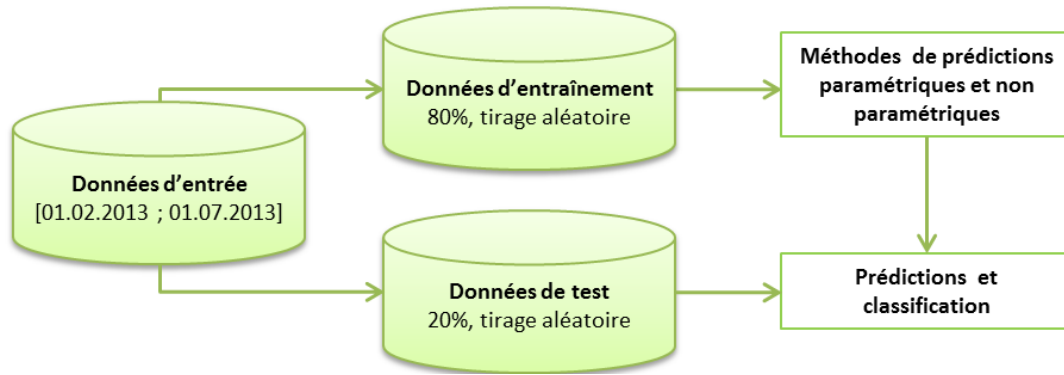


Figure 70 : Répartition des données pour la prédiction du besoin en eau chaude sanitaire

L'erreur y est beaucoup plus faible que pour le chauffage avec une MAE de 0.038 kWh fournit par le GBT (Tableau 31) avec un maximum de l'erreur en énergie qui ne dépasse pas 0.51 kWh (Figure 71). Cependant, nous avons eu des problèmes de connexions avec le capteur Zigbee placé sur la pompe à chaleur. Après suppression des données, ils ne restent que 86 heures à analyser ou près des deux tiers de ces heures représentent l'arrêt de la pompe à chaleur.

Modèles	MAE	MSE	RMSE	MSD	R ²
Random Forest regression	0.089 +/- 0.01	0.016	0.126 +/- 0.01	0.024	0.661
Gradient Boosted Tree	0.038 +/- 0.01	0.012	0.109 +/- 0.01	0.006	0.744

Tableau 31 : Résultats de la prédiction par modèle mathématique pour une prédiction à t + 1 heure pour les jours de la semaine

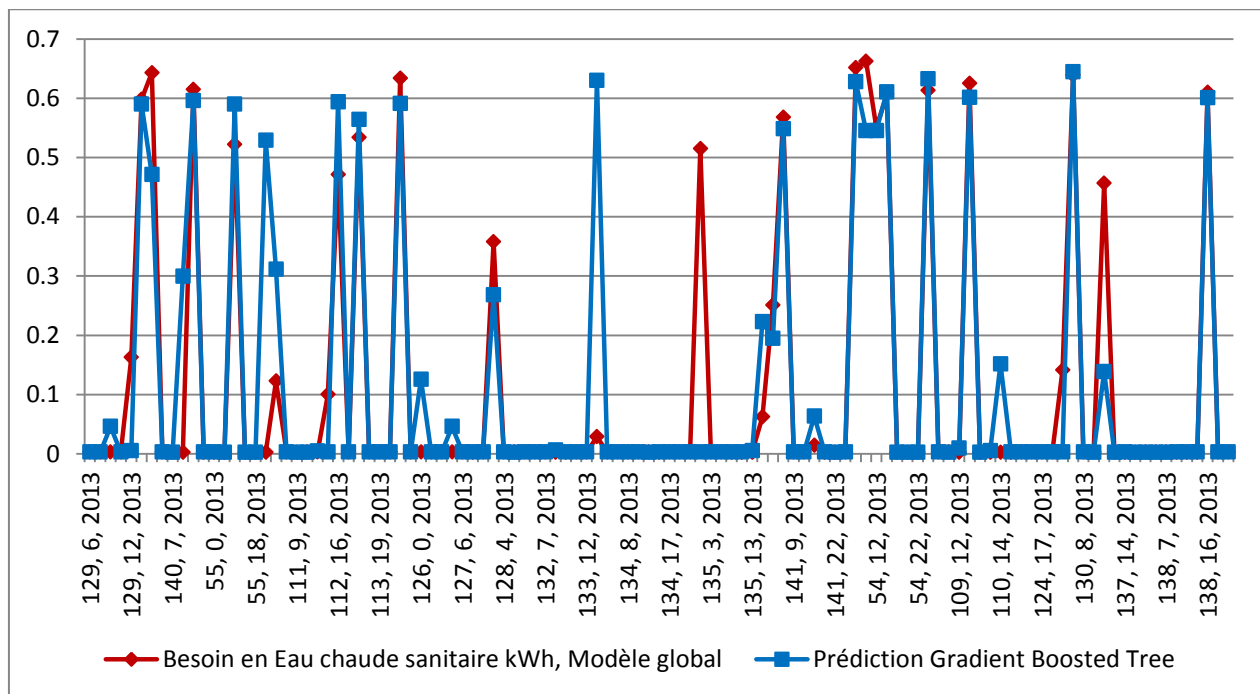


Figure 71 : Besoin en eau chaude sanitaire et la prédiction associée sur les heures représentant nos données de test

Résultats de la classification

L'objectif est de prédire l'enclenchement ou l'arrêt de la pompe à chaleur pour l'heure suivante : nous avons deux classes à prédire. Avec les intervalles de confiance, nous ne pouvons pas dégager un algorithme parmi les autres et une combinaison des différents algorithmes permettrait d'augmenter notre taux de classification qui est autour de 92% +/- 1.5 en séparant le week-end et la semaine (Tableau 32). Si nous prenons le Random Forest, 102 heures ont été classifiées correctement dont 19 heures « ON » et 83 heures « OFF » décrits dans les modèles ROC par les figures 72 et 73.

Modèles	Taux de classification Global	Taux de classification Semaine	Taux de classification Week-end
SVM	88.288 +/- 2.3	92.208 +/- 2.1	91.176 +/- 2.2
RF	91.892 +/- 2.4	93.61 +/- 1.9	88.235 +/- 2.4
GBT	90.09 +/- 2.1	92.45 +/- 2.2	91.23 +/- 1.7

Tableau 32 : Résultats de la classification ON-OFF par modèle mathématique pour une prédiction à t + 1 heure pour les jours de la semaine

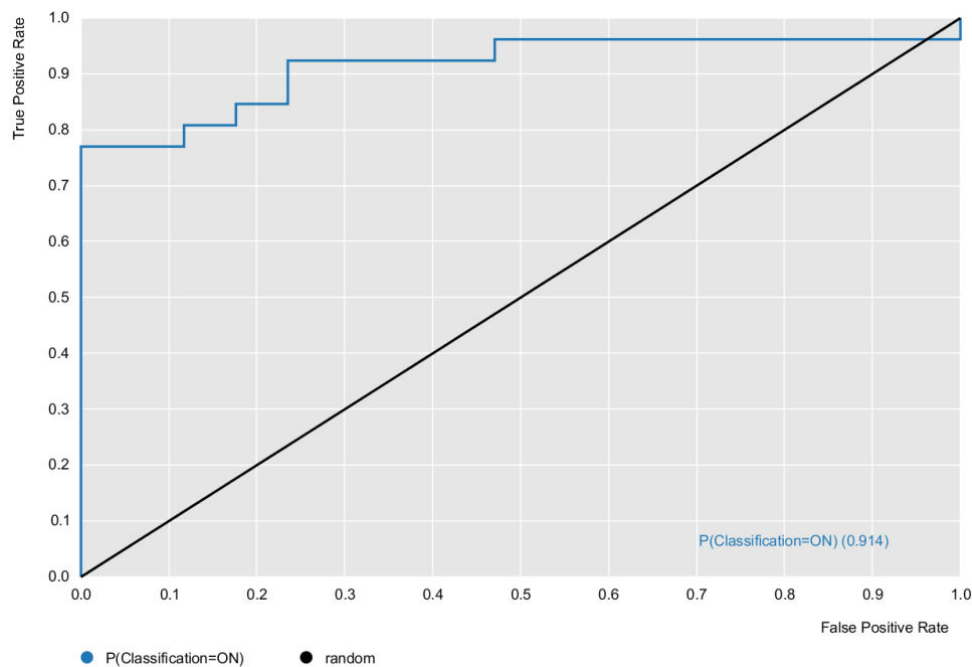


Figure 72 : Courbe ROC qui montre la performance du modèle à classer correctement l'allumage de l'eau chaude sanitaire pour les jours de la semaine

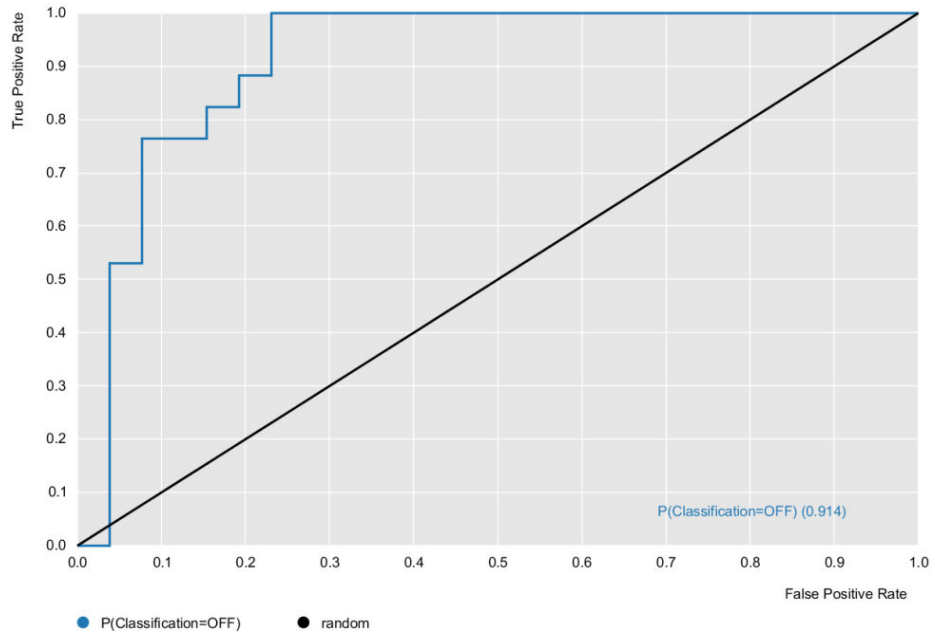


Figure 73 : Courbe ROC qui montre la performance du modèle à classifier correctement l'arrêt de l'eau chaude sanitaire pour les jours du week-end

Impact du planning de production

Nous créons un modèle représentant les jours de la semaine et un autre modèle caractérisant le samedi et dimanche. Nous remarquons une dégradation dans nos résultats alors que nous aurions pu croire à une amélioration puisque un planning de production pour la pompe à chaleur est différent entre les jours de la semaine et le week-end (Tableau 33).

Cela s'explique par le peu de données utilisées pour ce test. En séparant nos données d'entraînement, les différents algorithmes n'ont pas assez de données pour correctement modéliser le besoin en eau chaude sanitaire.

Algorithme	Modèles	MAE	MSE	RMSE	R ²
GBT	Global	0.043	0.008	0.092	0.338
	Semaine	0.062	0.022	0.149	0.666
	Week-end	0.07	0.023	0.15	0.697

Tableau 33 : Résultats statistiques pour la prédiction du besoin en eau chaude sanitaire entre les modèles jours de la semaine et week-end

Impact du nombre de variables

Nous faisons également évoluer notre nombre de variables pour y évaluer l'impact sur nos résultats et nous remarquons une dégradation significative de l'erreur à partir de 10 variables (Tableau 34). Pour cela, nous utilisons les données statistiques sur les arbres créés dans le Random Forest.

Nous vous donnons les 10 variables les plus utilisées pour la construction des arbres de décision pour la prédiction du besoin en chauffage pour les jours de la semaine :

- ⇒ Consommation électrique totale t (-1)
- ⇒ Maximum puissance t (-1)
- ⇒ Maximum Variations Puissance électrique totale t (-1) Phase 1
- ⇒ Moyenne de la puissance t (-1)
- ⇒ Moyenne Variations Puissance électrique totale t (-1) Phase 1
- ⇒ Déviation standard de la médiane de la Puissance électrique totale t (-24)
- ⇒ Médiane Puissance électrique totale t (-1)
- ⇒ Minimum Puissance électrique totale t (-1)
- ⇒ Besoin en eau chaude sanitaire t (-1)
- ⇒ Somme Puissance électrique totale t (-1)

Nombre de variables	MAE	MSE	RMSE	MSD	R ²
2400	0.038	0.012	0.109	0.006	0.744
1135	0.043	0.008	0.092	0.013	0.819
800	0.044	0.009	0.095	0.012	0.807
400	0.046	0.009	0.097	0.018	0.8
200	0.047	0.01	0.101	0.012	0.782
100	0.051	0.011	0.105	0.109	0.763
50	0.046	0.09	0.097	0.018	0.8
10	0.045	0.014	0.117	0.001	0.706
5	0.053	0.016	0.125	0.01	0.666

Tableau 34 : Résultats statistiques en sortie du GBT en fonction du nombre de variables pour la prédiction du besoin en eau chaude sanitaire pour le secteur résidentiel, modèle global

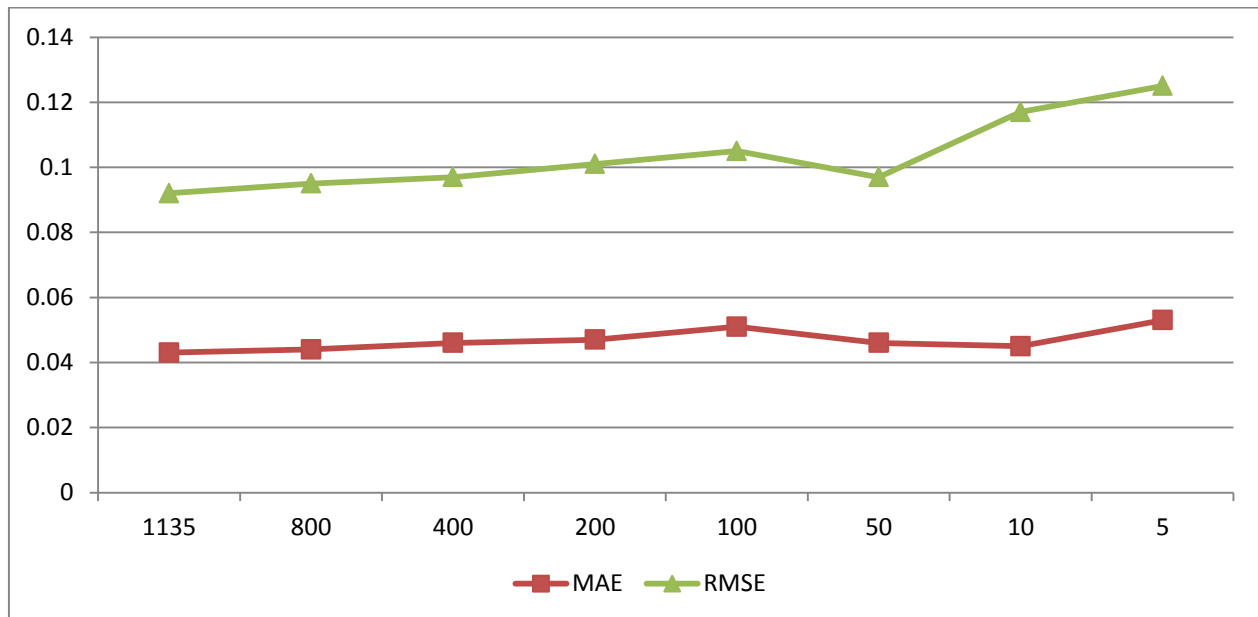


Figure 74 : Evolution de la précision de la prédiction en fonction du nombre de variables d'entrée pour la prédiction du besoin en eau chaude sanitaire

Impact de la collecte de données à la minute ou à l'heure

Nous testons ici la différence dans nos résultats entre les données à la seconde agrégées à la minute et une donnée par heure pour aller prédire l'heure suivante. Contrairement au test précédent où nous avons accès aux statistiques sur la puissance et sur les variations, nous utilisons dans ce test seulement la donnée à l'heure pour prédire la consommation électrique totale l'heure suivante.

Nous remarquons que l'erreur y est trois fois plus importante entre des données collectées à l'heure ou à la minute. Dans certains cas, le modèle à l'heure dévie totalement et les connaissances relatives aux variations (Tableau 35).

Algorithme	Modèles	Collecte	MAE	MSE	RMSE	R ²
GBT	Global	1 min	0.038 +/- 0.01	0.012	0.109 +/- 0.01	0.744
		1h	0.099 +/- 0.01	0.019	0.139 +/- 0.01	0.588

Tableau 35 : Résultats numériques par modèle mathématique pour une prédiction à t + 1 heure pour le modèle global, celui représentent les jours du lundi au vendredi et le week-end

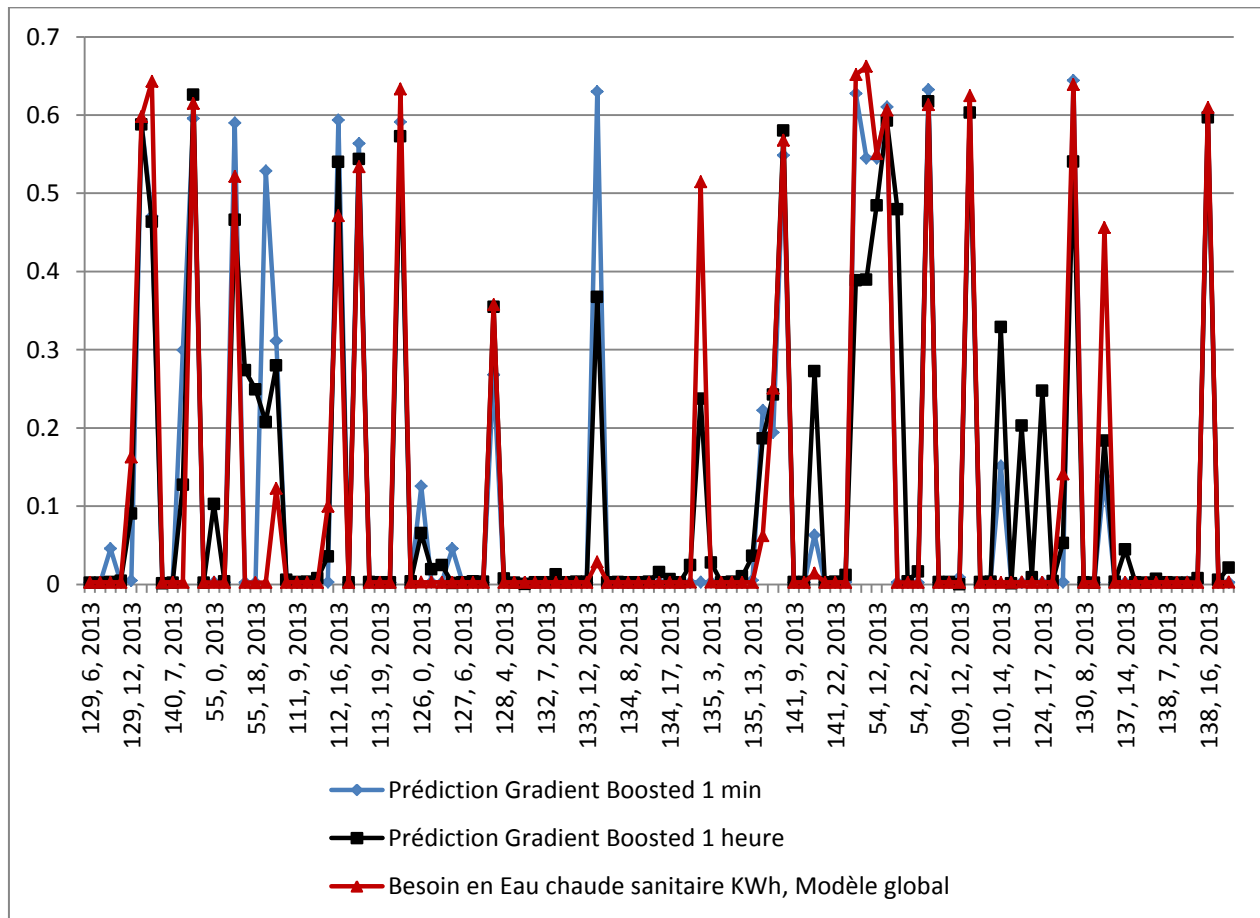


Figure 75 : Prédiction du besoin en eau chaude sanitaire en fonction d'une collecte des données à la minute ou à l'heure

5.3 Prédire le besoin en énergie dans le secteur tertiaire

5.3.1 Présentation du cas d'étude

Le deuxième cas d'étude se nomme le Technopôle et est situé dans la ville de Sierre. Il regroupe plusieurs entreprises aux secteurs d'activité variés (Tableau 36). Le bâtiment date des années 1980 aux années 2000 pour les dernières installations. Le bâtiment possède une centrale solaire photovoltaïque (PV) de 203.5 kWc et une batterie sur la base de modules Li-ion Titanate Leclanché SRB1021 [LEC-2015]. Les différentes applications réseau peuvent être réalisées et combinées avec une électronique de puissance pour contrôler les flux d'énergie. La configuration est une armoire double de 25 kWh composée de 14 modules de 2 kWh, d'un contrôleur, d'un système de disjonction et de l'équipement de refroidissement. Le chauffage et la climatisation sont régulés par des panneaux radiants et n'entrent pas dans la consommation électrique totale du site (Figure 76). Cette batterie est aujourd'hui raccordée au restaurant d'entreprise pour limiter ses pics de consommation en fonction de la production solaire et du pic détecté.

L'énergie liée au chauffage est fournie par une chaudière à gaz. Celle-ci est en cours de remplacement par une pompe à chaleur. L'énergie liée à la climatisation est fournie par pompage de l'eau du Rhône situé à côté du bâtiment.

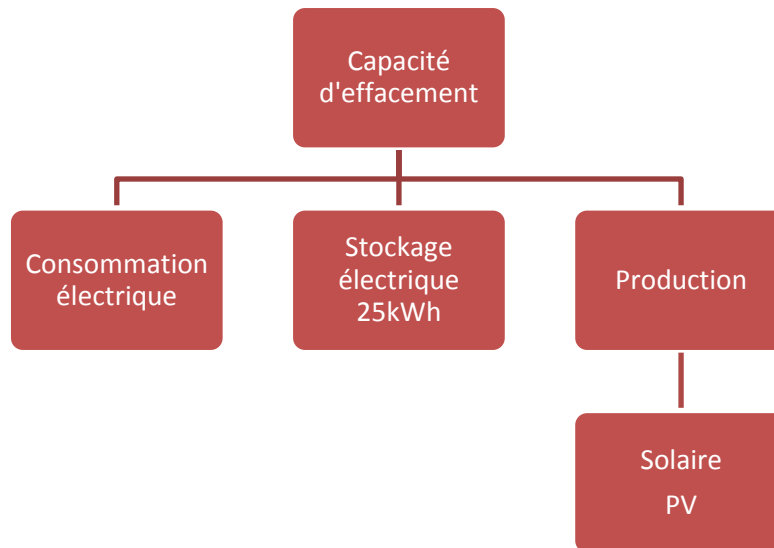


Figure 76 : Caractérisation de la capacité d'effacement qui nécessite la prédiction de la consommation électrique pour chaque entreprise connectée au système d'information et de la production solaire

Noms entreprises	Secteurs d'activité	
	Code NOGA	Libellés
TI Informatique	582900	Edition de logiciels
Consultec	692000	Activités comptables, fiduciaires
Schoechli Impression	181100	Imprimerie et services annexes
Eticolle	181204	Imprimerie (labeur)
Canal9	591100	Production de films cinématographiques, de vidéo et de programmes de télévision
	602000	Programmation de télévision et de télédiffusion
Netplus	612000	Télécommunication sans fil
HES-SO	854201	Haute école universitaire
TI Informatique	631100	Traitement de données, hébergement et activités connexes
Boomerang	619000	Activités de télécommunication
SecureIT	620900	Activités informatiques (Sécurité)
Propriétaire bâtiment : TechnoArk	631100	Traitement de données, hébergement et activités connexes
TelecomWatcher	702200	Conseil pour les affaires et autres conseils de gestion
Icare	620900	Activités informatiques (R&D)
Mikado	562900	Services de restauration (entreprises)

Tableau 36 : Entreprises connectées au système d'information avec un relevé de la consommation à la seconde

Nous présentons nos résultats pour deux entreprises : le Mikado et la HES-SO. Le Mikado est le restaurant du Technopôle de Sierre. Il est fréquenté par le personnel du restaurant (propriétaires, cuisiniers, serveurs, autres membres du personnel de cuisine) et par les clients qui travaillent dans le Technopôle ainsi que des personnes extérieures au Technopôle.

Le restaurant est ouvert pendant les jours de semaine du lundi au vendredi. Les heures d'ouverture sont de 7h à 17h et le restaurant est ouvert seulement le midi et les jours de la semaine.



Figure 77: Plan d'architecte du rez-de-chaussée du TechnoPôle à gauche, avec en orange la surface dédiée au restaurant d'entreprise

Le restaurant d'entreprise a une faible bande de base de 3kW, qui correspond à la consommation des groupes de froid. Cette entreprise ne consomme pas beaucoup d'électricité en soi, mais appelle beaucoup de puissance à certaines heures de la journée qui correspondent aux heures de préparation et de service des repas. Il s'agit de la consommation des fours, des plaques électriques et des micro-ondes.

Cela signifie que pour le dimensionnement d'un microgrid, la consommation va difficilement pouvoir être lissée, car il n'est pas envisageable pour un restaurant de couper les appareils cités précédemment aux heures de pointe. En revanche, la prédiction est très utile car elle peut permettre d'anticiper cette consommation et d'utiliser l'énergie électrique produite en toiture ou l'énergie stockée en batterie de capacité 25 kWh sur le site au moment du pic afin de limiter la pointe de consommation sans perturber l'activité de l'entreprise.

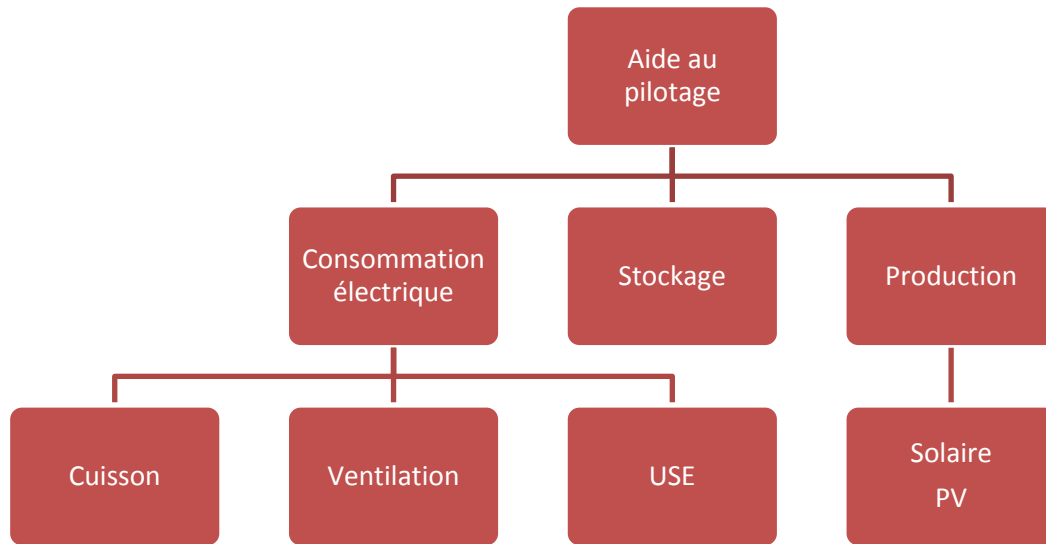


Figure 78 : Caractérisation de la capacité d'effacement qui nécessitent la prédiction du chauffage, de l'eau chaude sanitaire, de a consommation électrique globale et de la production solaire

Le deuxième cas d'étude est l'institut entrepreneariat et management. Pour ces bureaux, la consommation électrique globale comprend la consommation en luminaires et en usages spécifiques électriques utiles (Figure 79). Les heures d'ouverture sont de 7h à 19h. Nous avons :

$$E_{USE} t(i) = \sum_{A=1}^{A=N} (E_{Luminaires} + E_{Usage N}) t(i)$$

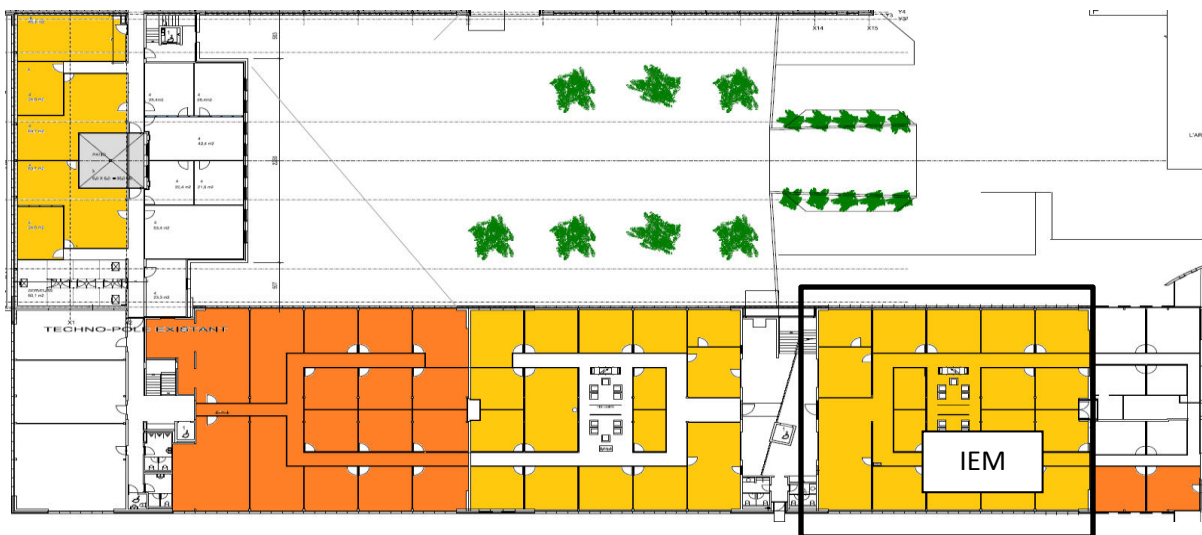


Figure 79 . Représentation des sous-systèmes d'étude dans le secteur d'activité bureaux

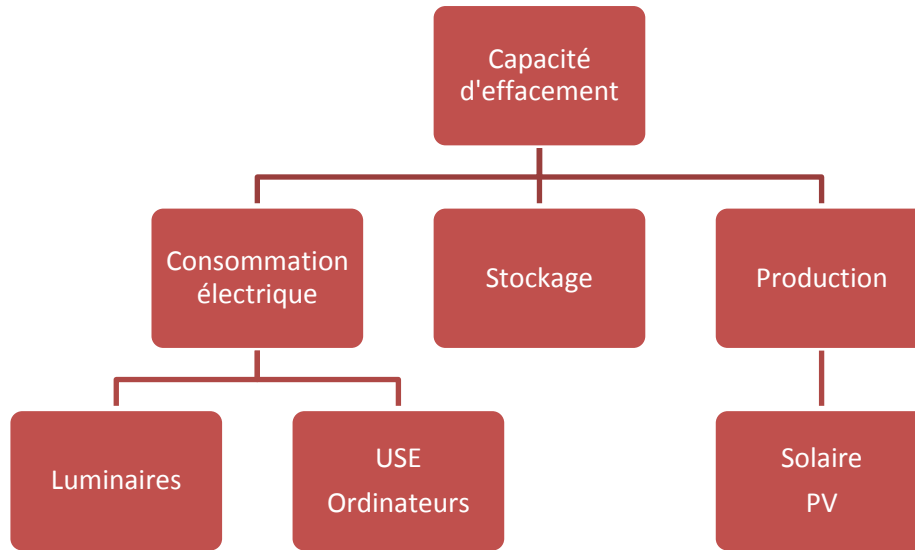


Figure 80 : Caractérisation de la capacité d'effacement qui nécessite la prédiction du chauffage, de l'eau chaude sanitaire, de la consommation électrique globale et de la production solaire

5.3.2 Prédire la consommation électrique globale

5.3.2.1 Secteur restauration

Données d'entrée

Pour le restaurant, la consommation électrique globale comprend la consommation en ventilation et en luminaires et en usages spécifiques électriques utiles. Nous avons :

$$E_{USE} t(i) = \sum_{A=1}^{A=N} (E_{Ventilation} + E_{Luminaires} + E_{Usage N}) t(i)$$

Notre vecteur d'entrée qui caractérise la consommation électrique globale regroupe les statistiques liées à la puissance et aux différentes variations au niveau du compteur global d'électricité. Les données météorologiques sont données par les valeurs réelles et prédites horaires des stations entourant notre cas d'étude. La temporalité de notre signal est gardée et délimitée par les variables m , j , js et l'heure. Pour les résultats de la classification, chaque heure est labélisée par une classe c qu'il s'agira de déterminer à partir de la modélisation :

$$E_{Consommation_électrique_totale} t(i) = [Date ;$$

$$\text{Statistiques Puissance Phase 1 } t(i), \dots, \text{Statistiques Puissance Phase 1 } t(i - 48),$$

$$\text{Statistiques Variations Phase 1 } t(i), \dots, \text{Statistiques Variations Phase 1 } t(i - 48)$$

$$\text{Données Réelles_MétéoStation } t(i), \dots, \text{Données Réelles_MétéoStation } t(i - 48),$$

$$\text{Données Prédites_MétéoStation } t(i + 1), \dots, \text{Données Prédites_MétéoStation } t(i - 48)$$

$$\text{Année}_j, \text{Mois}_m, \text{Jour}_j, \text{Jour_Semaine}_{js}, \text{Heure}_h, \text{Jour fériés}_{0,1}, \text{Vacances scolaires}_{0,1},$$

$$\text{Classes}_c]$$

Dans le secteur tertiaire et industriel, les heures d'arrivée et de départ du personnel et les jours travaillés comme informations d'entrée permettront d'affiner la qualité du set d'entraînement en enlevant des données que l'on peut considérer comme du bruit. Dans le secteur tertiaire, **les jours et les heures en dehors du planning de production sont enlevés** de notre set de données d'entraînement et de test. Pour le restaurant, le week-end est enlevé et seules les heures entre 6h et 17h sont gardées (Figure 74).

Pour notre étude, l'ensemble du set de données rassemble 4024 heures du 1^{er} septembre 2013 au 31 Octobre 2014 en enlevant les données. 3772 heures sont utilisées pour entraîner le modèle et 252 heures représentent notre set de test.

Résultats de la prédiction

Comme pour le secteur résidentiel, les ensembles d'arbres de décision et en particulier le Gradient Boosted Tree est l'algorithme qui nous fournit les meilleurs résultats. La moyenne des écarts en valeur absolue y est de 1.162 kWh pour le GBT (Tableau 37). La précision des modèles s'expliquent par la

répétition à des heures précises d'étapes pour préparer les repas. Des données a priori aujourd'hui collectées par le restaurant comme le nombre de clients prévus ou à des commandes spéciales (fêtes d'entreprise par exemple) connues en amont par le restaurant pourraient améliorer encore la précision de nos modèles mathématiques.

Modèles	MAE	MSE	RMSE	MSD	R ²
MLP	1.867	5.482	2.14	0.225	0.643
PNN	1.867	5.482	2.14	0.225	0.643
Régression linéaire	2.247	5.482	2.14	0.225	0.643
ARIMA	2.247	5.482	2.14	0.225	0.643
RF	1.488	4.494	2.12	0.124	0.946
GBT	1.162	4.282	1.92	0.112	0.978

Tableau 37 : Résultats statistiques en sortie des différents algorithmes pour la prédiction de la consommation électrique globale pour le secteur tertiaire, Restauration

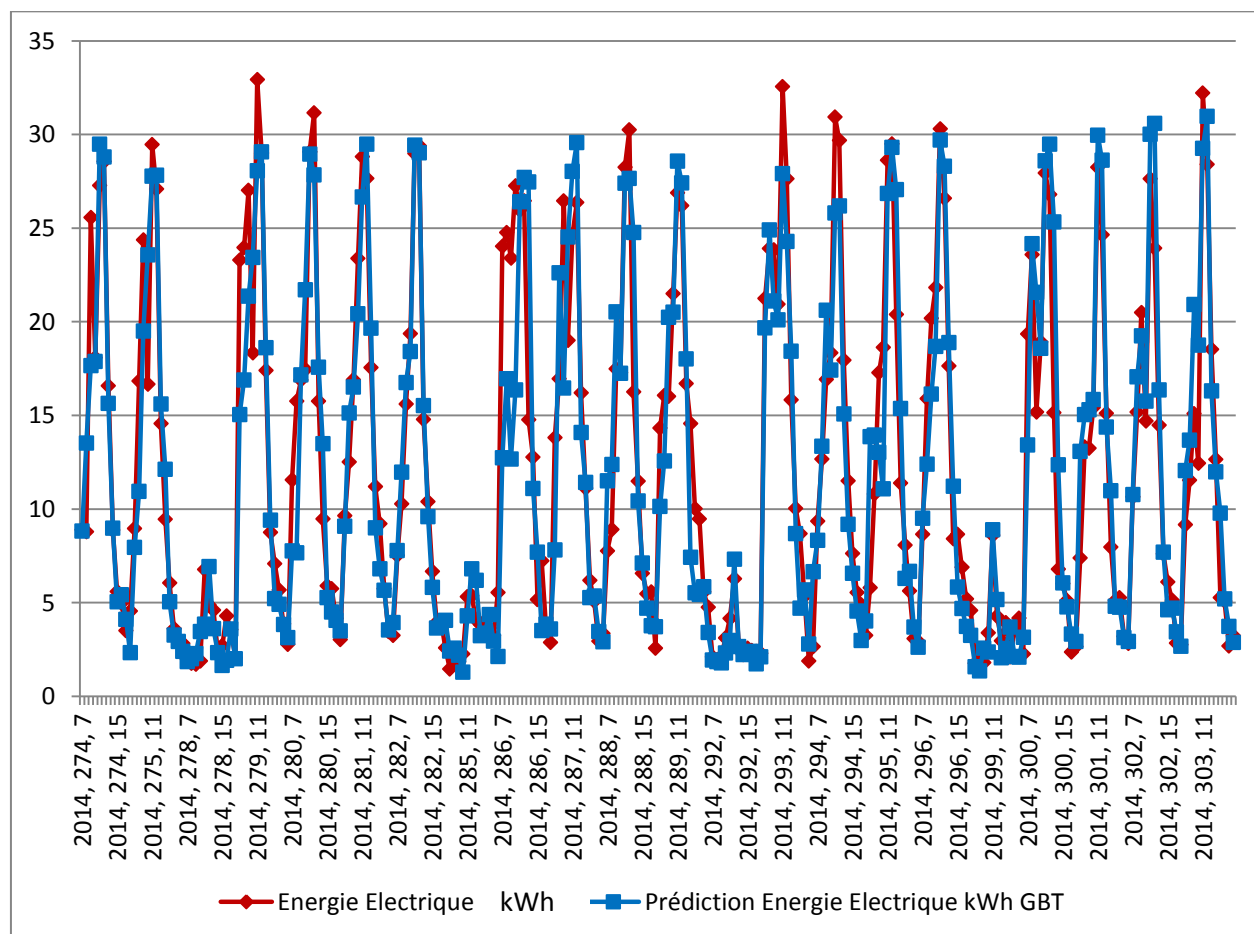


Figure 81 : Consommation électrique globale et la prédiction en sortie du gradient boosted tree en Kwh dans le secteur tertiaire, restauration (Année, jour de l'année, heure)

Résultats de la classification

L'objectif est de correctement assigner la bonne classe aux 1160 heures. L'ensemble du set de données représentent 2013 heures. 1989 heures sont utilisées pour entraîner le modèle par un tirage aléatoire. 1160 représentent notre set de test que nous devons prédire le plus précisément possible. En sortie de l'apprentissage non supervisé, la consommation électrique globale dans le secteur résidentiel est représentée par six classes (Tableau 38). La classe 6 représente une gaussienne qui représente le pic de puissance du restaurant pendant la période considérée.

Energie à prédire	Sortie apprentissage	Moyenne	Déviat ion standard	Répartition %
Classification de l'énergie électrique, Secteur résidentiel	0	2.304	0.535	10
	1	4.448	1.29	24
	2	8.961	2.392	18
	3	16.622	2.539	18
	4	22.982	2.449	9
	5	28.246	1.927	35
	6 : Pic de puissance	31.866	4.367	2

Tableau 38 : Classes représentant la consommation électrique globale dans le secteur tertiaire, secteur d'activité restauration en kWh

Le Gradient Boosted Tree réussit le mieux à classifier les six classes représentant la consommation électrique (Tableau 39). Sur les 1160 heures représentant notre set de données, 1124 heures ont été correctement classifiés. Si nous prenons plus spécifiquement la classe 6 représentant le pic de puissance, 10 heures étaient à classifier correctement, 8 d'entre elles l'ont été.

Modèles	Taux de classification	Déviat ion standard
PNN	84.617 +/- 1.2	1.6
MLP	88.847 +/- 1.3	1.5
SVM	91.837 +/- 1.1	1.6
Random Forest	96.124 +/- 0.7	1.5
Gradient Boosted Tree	98.074 +/- 0.6	0.8

Tableau 39 : Résultats de la classification par modèle mathématique

Impact de la collecte de données à la minute ou à l'heure

Nous testons également la différence dans nos résultats entre les données à la seconde agrégée à la minute et une donnée par heure pour aller prédire l'heure suivante. L'erreur moyenne des écarts augmentent de 0.5 kWh (Tableau 40). Contrairement au secteur résidentiel, la courbe de consommation électrique journalière moyennée par mois suit un même profil. L'impact de la fréquence de collecte des données y est beaucoup moins important que dans le secteur résidentiel. Dans le secteur tertiaire, les entreprises suivent un planning de production précis dans la journée contrairement aux habitudes de consommation humaines changeantes dans le secteur résidentiel.

Modèles	Fréquence	MAE	MSE	RMSE	MSD	R ²
Random Forest	1 min	1.488	4.494	2.12	0.124	0.946
Random Forest	1 heure	2.133	11.36	3.371	-0.273	0.863

Tableau 40 : Résultats statistiques en sortie du Gradient Boosted Tree en fonction de deux fréquences de mesures pour la prédiction de la consommation électrique globale pour le secteur tertiaire, Restauration

Impact du nombre de variables

Dans ce test, notre objectif est de déterminer l'impact de la quantité de nos données d'entrée dans la variation du résultat (Tableau 41).

Nombre de variables	MAE	MSE	RMSE	MSD	R ²
4160	1.488	4.494	2.12	0.124	0.946
300	2.034	8.668	2.944	0.101	0.895
150	1.882	7.583	2.754	-0.053	0.908
50	2.002	8.283	2.878	-0.343	0.899
15	2.22	10.767	3.281	-0.109	0.869

Tableau 41 : Résultats statistiques en sortie du Gradient Boosted Tree en fonction du nombre de variables pour la prédiction de la consommation électrique globale pour le secteur tertiaire, Restauration

Les 15 variables les plus déterminantes pour la construction des arbres de décision pour la prédiction de la consommation électrique globale pour ce système étudié sont :

- ⇒ Heure
- ⇒ Consommation électrique totale t (-1)
- ⇒ Consommation électrique totale t (-2)
- ⇒ Maximum Puissance électrique totale t (-1)
- ⇒ Maximum Puissance électrique totale t (-2)
- ⇒ Déviation standard t (-1)
- ⇒ Déviation standard t (-2)
- ⇒ Minimum Puissance électrique totale t (-1)
- ⇒ Médiane Puissance active t (-1)
- ⇒ Déviation standard Médiane Puissance active t (-1)
- ⇒ Maximum Variations puissance active t (-1)
- ⇒ Maximum Variations puissance active t (-2)
- ⇒ Déviation standard Variations puissance active t (-1)
- ⇒ Minimum Variations puissance active t (-1)
- ⇒ Déviation standard Médiane Variations puissance active t (-1)

5.3.2.2 Secteur bureaux

Données d'entrée

Notre vecteur d'entrée qui caractérise la consommation électrique globale regroupe les statistiques liées à la puissance et aux différentes variations au niveau du compteur global d'électricité. Les données météorologiques sont données par les valeurs réelles et prédites horaires des stations entourant notre cas d'étude. La temporalité de notre signal est gardée et délimitée par les variables m, j, js et l'heure. Pour les résultats de la classification, chaque heure est labélisée par une classe c où il s'agira à notre outil de définir l'heure suivante dans quelle classe se situe la consommation électrique :

$$E_{\text{Consommation_électrique_totale}} t(i) = [\text{Date ;} \\ \text{Statistiques Puissance phase 1 } t(i), \dots, \text{Statistiques Puissance phase 1 } t(i - 48), \\ \text{Statistiques Variations phase 1 } t(i), \dots, \text{Statistiques Variations phase 1 } t(i - 48) \\ \text{Données Réelles_MétéoStation } t(i), \dots, \text{Données Réelles_MétéoStation } t(i - 48), \\ \text{Données Prédites_MétéoStation } t(i + 1), \dots, \text{Données Prédites_MétéoStation } t(i - 48) \\ \text{Année}_j, \text{Mois}_m, \text{Jour}_j, \text{Jour_Semaine}_js, \text{Heure}_h, \text{Jour fériés}_{0,1}, \text{Vacances scolaires}_{0,1}, \\ \text{Classes}_c]$$

Concernant le secteur d'activité bureaux, le week-end est enlevé et seules les heures entre 7h et 19h sont gardées. L'ensemble du set de données représente 4024 heures : 3772 heures sont utilisées pour entraîner le modèle par un tirage aléatoire et 252 heures représentent notre set de test que nous devons prédire le plus précisément possible.

Résultats de la prédiction

Comme pour le secteur résidentiel, les ensembles d'arbres de décision et en particulier le Gradient Boosted Tree est l'algorithme qui nous fournit les meilleurs résultats. La moyenne des écarts en valeur absolue y est de 0.586 kWh (Tableau 42).

Modèles	MAE	MSE	RMSE	MSD	R ²
MLP	0.864	0.804	0.975	0.242	0.702
PNN	0.857	0.784	0.964	0.231	0.689
Régression linéaire	1.125	0.908	1.253	0.374	0.502
Random Forest regression	0.586	0.601	0.775	0.168	0.807
Gradient Boosted Tree	0.611	0.87	0.933	0.193	0.721

Tableau 42 : Résultats statistiques en sortie des différents algorithmes pour la prédiction de la consommation électrique globale pour le secteur tertiaire, secteur d'activité bureaux

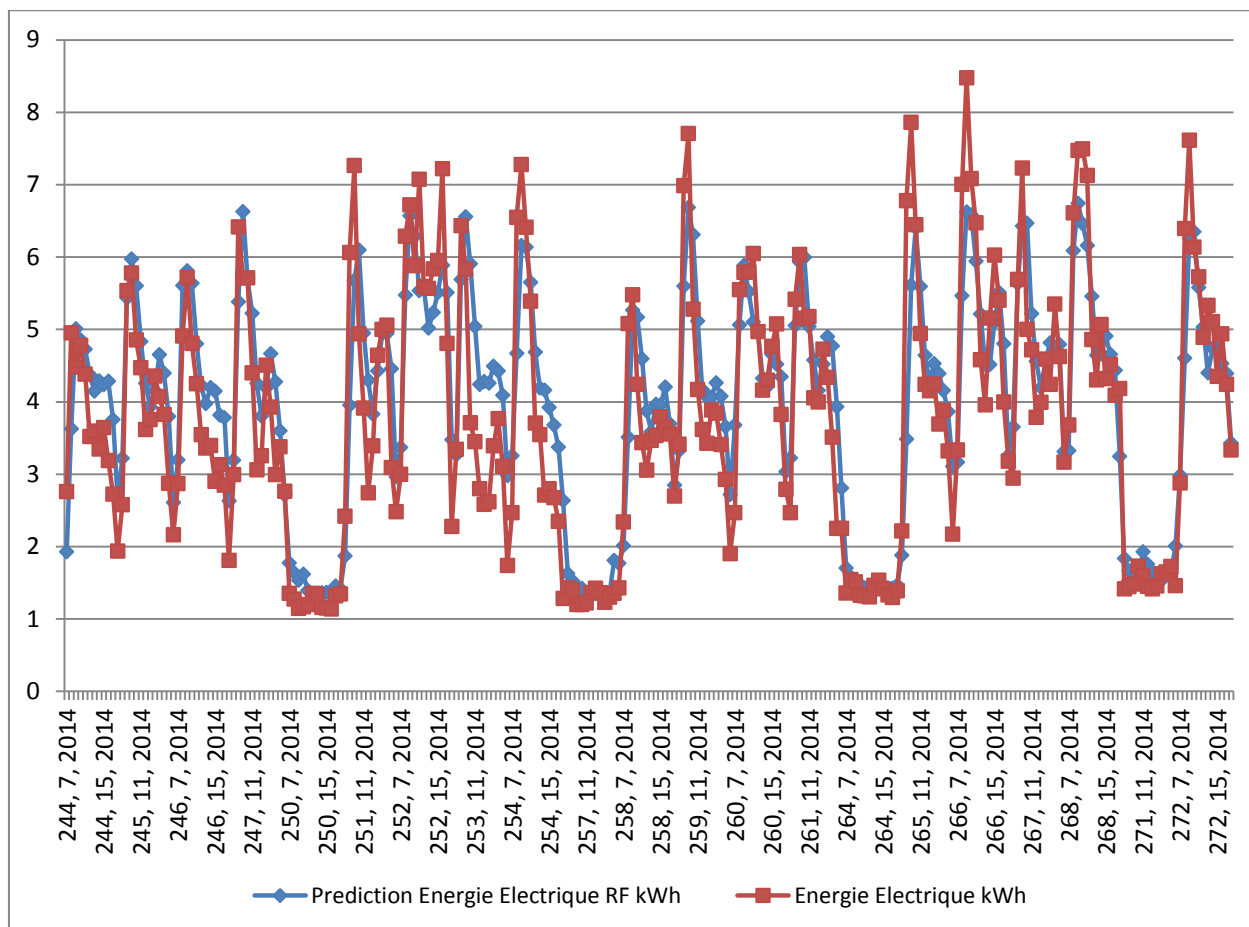


Figure 82 : Consommation électrique globale et la prédiction en sortie du Random Forest Tree en kWh dans le secteur tertiaire, secteur d'activité bureaux

Résultats de la classification

Pour la classification, l'objectif est de correctement assigner la bonne classe aux 1160 heures. L'ensemble du set de données représente 2013 heures. 1989 heures sont utilisées pour entraîner le modèle par un tirage aléatoire. 1160 représentent notre set de test que nous devons prédire le plus précisément possible. Nous obtenons 97% de classification réussie pour le secteur d'activité bureaux.

Modèles	Taux de classification	Déviat ion standard
PNN	84.617 +/- 1.3	2.5
MLP	88.847 +/- 1.3	2.4
SVM	89.937 +/- 1.2	1.6
Random Forest	94.048 +/- 0.8	1.8
Gradient Boosted Tree	97.056 +/- 0.6	1.4

Table 43 : Résultats de la classification par modèle mathématique pour une prédiction à t + 1 heure, secteur d'activité bureaux

Impact de la collecte de données à la minute ou à l'heure

Nous testons également la différence dans nos résultats entre les données à la seconde agrégée à la minute et une donnée par heure pour aller prédire l'heure suivante. Cependant, l'impact de la fréquence de collecte des données y est beaucoup moins important (Tableau 44). Contrairement au secteur résidentiel, la courbe de consommation électrique journalière moyennée par mois a le même profil.

Modèles	Fréquence	MAE	MSE	RMSE	MSD	R ²
Random Forest	1 min	0.586	0.601	0.775	0.168	0.807
Gradient boosted tree	1 heure	0.56	0.687	0.829	0.05	0.779

Tableau 44 : Résultats statistiques en sortie du GBT en fonction d'une collecte à la minute ou à l'heure pour la prédiction de la consommation électrique globale pour le secteur tertiaire, secteur d'activité bureaux

Impact du nombre de variables

Notre objectif est ainsi de déterminer le nombre minimum de variables qui nous garantissent toujours un même niveau de prédiction détaillé dans le tableau 45.

Nombre de variables	MAE	MSE	RMSE	MSD	R ²
300	0.541	0.717	0.847	0.235	0.77
150	0.574	0.797	0.893	0.181	0.744
50	0.573	0.629	0.793	0.091	0.798
15	0.649	0.873	0.934	0.117	0.72

Tableau 45 : Résultats statistiques en sortie du GBT en fonction du nombre de variables pour la prédiction de la consommation électrique globale pour le secteur tertiaire, secteur d'activité bureaux

Les 15 variables les plus déterminantes pour la prédiction de la consommation électrique globale dans le secteur résidentiel pour les bureaux dans notre test sont :

- ⇒ Consommation électrique totale t (-1)
- ⇒ Consommation électrique totale t (-2)
- ⇒ Consommation électrique totale t (-24)
- ⇒ Maximum Puissance électrique totale t (-1)
- ⇒ Maximum Puissance électrique totale t (-2)
- ⇒ Maximum Puissance électrique totale t (-24)
- ⇒ Déviation standard t (-1)
- ⇒ Déviation standard t (-2)
- ⇒ Minimum Puissance électrique totale t (-1)
- ⇒ Minimum Puissance électrique totale t (-24)
- ⇒ Médiane Puissance active t (-1)

- ⇒ Médiane Puissance active t (-2)
- ⇒ Médiane Puissance active t (-24)

5.3.3 Prédire la production solaire

Au niveau de la prédiction de production solaire décentralisée, les différentes études concernent un large choix d'installations photovoltaïques et de type de données, (étude sur 12 cellules PV en laboratoire ou de 14 000m² de panneaux) [ZAM-2014] [GRA-2016] [GUL-2016]. D'après la bibliographie, les paramètres pouvant influencer la prédiction sont tellement nombreux, qu'il est difficile d'utiliser les résultats issus d'autres études. Ainsi, il peut être intéressant d'essayer des méthodes ou des paramètres même s'ils n'ont pas forcément fait leurs preuves dans d'autres études.

Il faut également noter qu'il existe de nombreux algorithmes d'apprentissage décrits dans le chapitre 2 avec la prédominance des réseaux de neurones. Les résultats obtenus sont de l'ordre de 12-15% (RMSE) pour une prévision à l'heure, et les moments les plus difficiles à prévoir concernent les moments consécutifs au passage d'épisodes météorologiques locaux (nuages, averses, orages).

Dans tous les cas étudiés, une seule station météo est considérée, par la suite j'ai donc choisi d'étudier l'impact de l'ajout de stations autour de mon centre de production de Sierre.

Données d'entrée

Nos résultats seront présentés sur la station photovoltaïque installée sur le toit du bâtiment tertiaire. Le bâtiment possède une centrale solaire photovoltaïque (PV) de 203.5kWc.

En sortie de la classification, chaque heure de notre vecteur d'entrée est caractérisé par une classe et un historique des données de productions et météorologiques sont également créés :

$$E_{ENRE} t(i + 1) = [Date ;$$

$$\begin{aligned} & \text{Statistiques Puissance}_{\text{Phase 1}} t(i), \dots, \text{Statistiques Puissance}_{\text{Phase 1}} t(i - 48), \\ & \text{Statistiques Variations}_{\text{Phase 1}} t(i), \dots, \text{Statistiques Variations}_{\text{Phase 1}} t(i - 48), \\ & \text{Données}_{\text{Réelles_MétéoStation-Local}} t(i), \dots, \text{Données}_{\text{Réelles_MétéoStation-Local}} t(i - 48), \\ & \text{Données}_{\text{Réelles_MétéoStation}} t(i), \dots, \text{Données}_{\text{Réelles_MétéoStation}} t(i - 48), \\ & \text{Données}_{\text{Prédites_MétéoStation}} t(i + 1), \dots, \text{Données}_{\text{Prédites_MétéoStation}} t(i - 48), \\ & \text{Année}_j, \text{Mois}_m, \text{Jour}_j, \text{Semaine}_{0,1}, \text{Heure}_h, \\ & \text{Classes}_c] \end{aligned}$$

Les données sont séparées en un set d'entraînement et un set de test. Nous utilisons les données depuis le mois de septembre 2013 au mois juillet 2014 pour entraîner nos modèles (Figure 83). Les mois d'août et de septembre 2014 représentent nos mois de test.

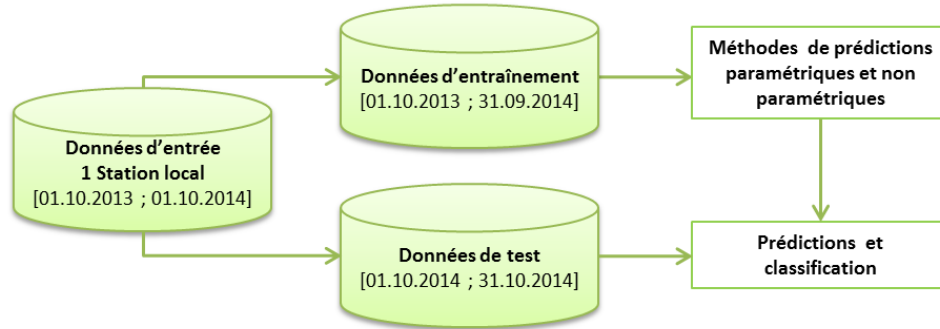


Figure 83 : Répartition des données pour la prédiction de la production décentralisée solaire

La présence d'une station météorologique reliée à une station photovoltaïque dépend de la puissance raccordée. Pour les petites stations inférieures à 20 kW, aucune station n'est présente. C'est pourquoi nous présentons différents tests.

Résultats de la prédiction

Pour ce premier test, nous utilisons seulement une station locale avec comme données d'entrée la température, la radiation lumineuse et la température des cellules par heure. Nous avons une MAE de 12.23 kWh basée pour une prédiction pour l'heure suivante (Tableau 46). Cependant, nous remarquons que les hautes puissances sont rarement prédites correctement. Ceci est la conséquence de la sous-évaluation de la radiation lumineuse en particulier lors des journées à fort ensoleillement du modèle de prédiction fournis par MétéoSuisse (Figure 84).

Modèles	MAE	RMSE	R ²
MLP	14.849	19.849	0.733
PNN	14.731	19.731	0.752
Régression linéaire	17.457	22.457	0.416
ARIMA	16.421	21.421	0.452
Random Forest regression	13.262	18.262	0.836
Gradient Boosted Tree	12.228	17.46	0.884

Tableau 46 : Résultats numériques par modèle mathématique pour une prédiction de la production solaire à t + 1heure, 1 station locale

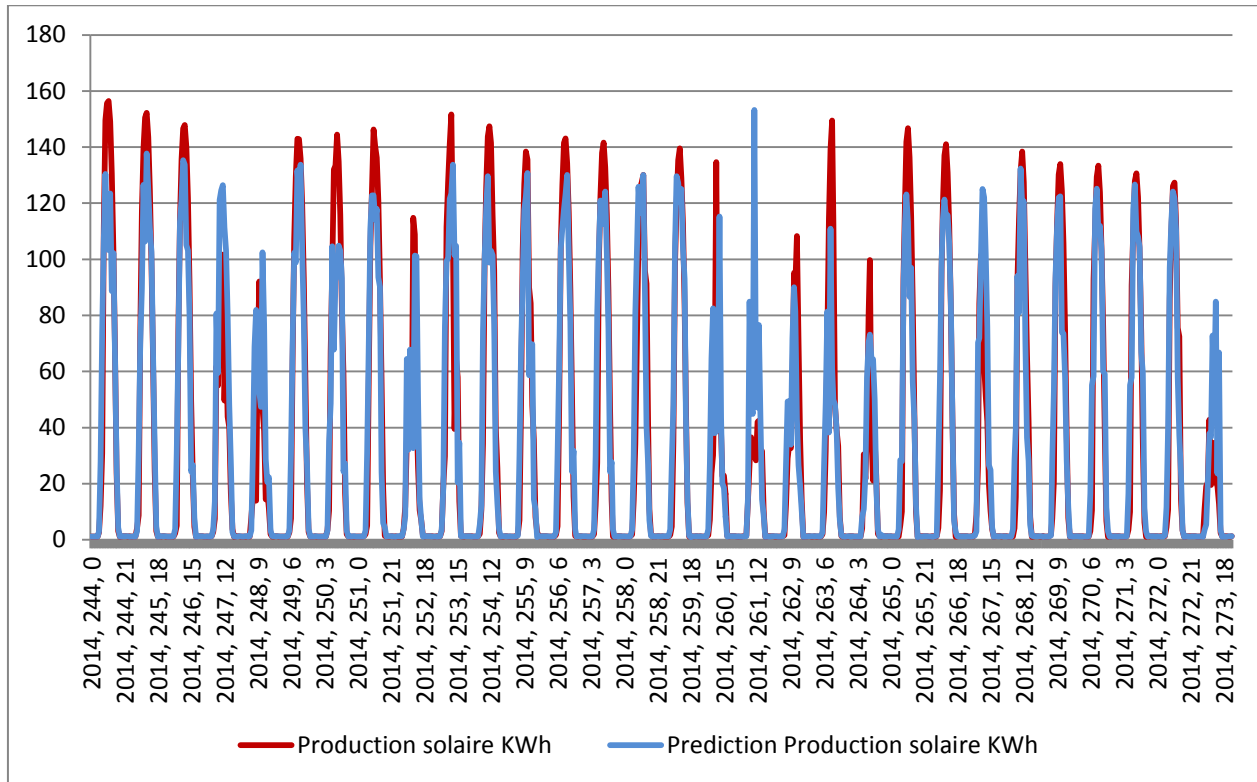


Figure 84 : Production solaire et la prédiction en sortie du GBT pour le mois de septembre avec la station de Sion distante de 15 km du site de production

Le second test regroupe les données de la station régionale à Sion distante de 15 km du site étudié. Sur cette station, nous avons les données réelles et prédites liées à la température et à la radiation lumineuse chaque heure avec notre propre modèle de prédiction météorologique décrit dans le chapitre 4.3. Nous remarquons que les erreurs relatives aux hautes puissances sont corrigées puisque notre modèle prédictif de radiation lumineuse corrige l'erreur fournie par MétéoSuisse en particulier sur les périodes à fort ensoleillement (Figure 85).

Modèle	Stations utilisées	MAE	RMSE	R ²
Gradient Boosted Tree	2 stations, données non corrigées	11.2	16.4	0.896
	2 stations, données corrigées	9.128	14.46	0.902
	6 stations, données corrigées	8.428	14.1	0.912

Tableau 47 : Résultats numériques par modèle mathématique pour une prédiction du besoin en chauffage à t + 1 heure pour les jours de la semaine

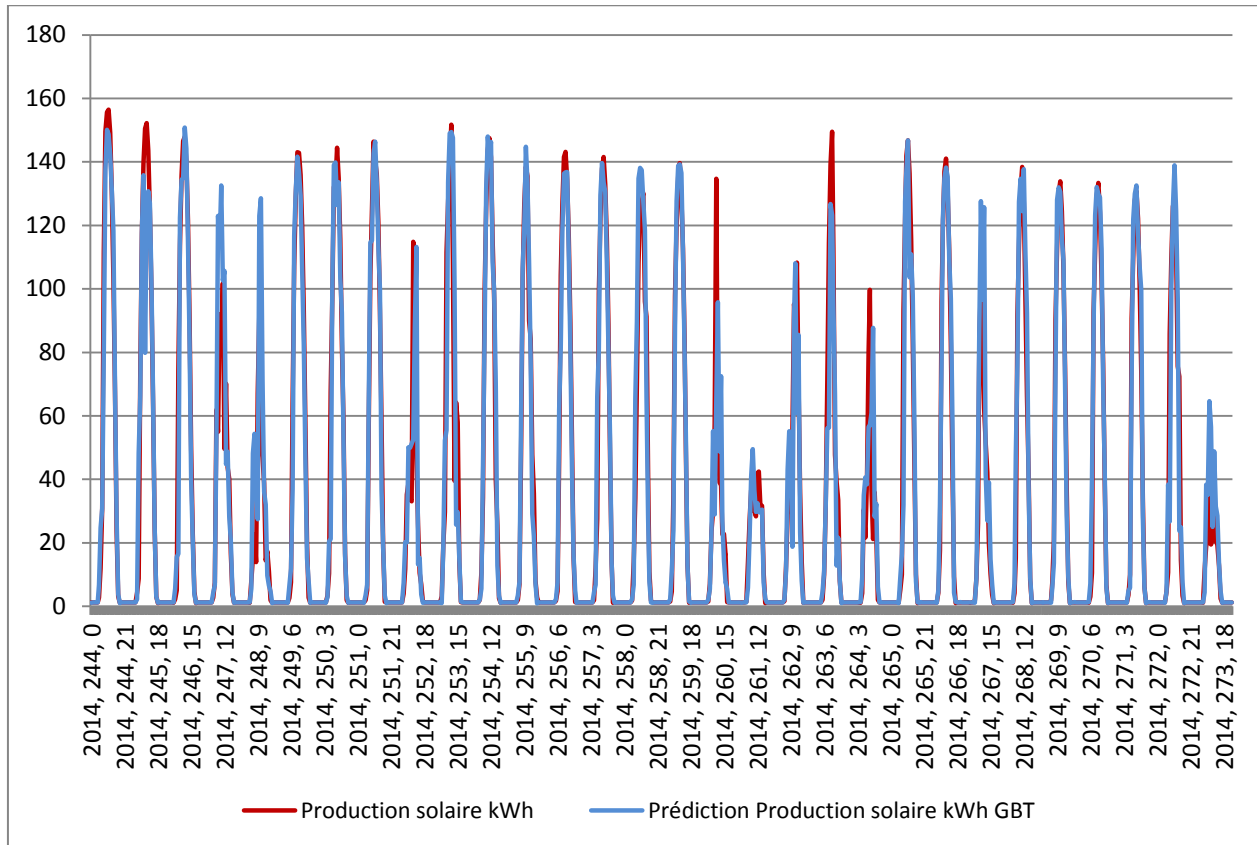


Figure 85 : Production solaire et la prédiction en sortie du gradient boosted tree pour le mois de septembre avec la station de Sion distante de 15 km du site de production

Résultats de la classification

A partir de l'énergie produite, de la luminosité locale à Sierre et de la luminosité régionale à Sion distante de 15 km du site de test, nous réalisons un apprentissage non supervisé (Figure 86). L'algorithme converge vers 11 classes regroupées en 5 classes (Tableau 46). Cet apprentissage nous donne des familles représentées par une gaussienne avec une moyenne et une déviation standard.

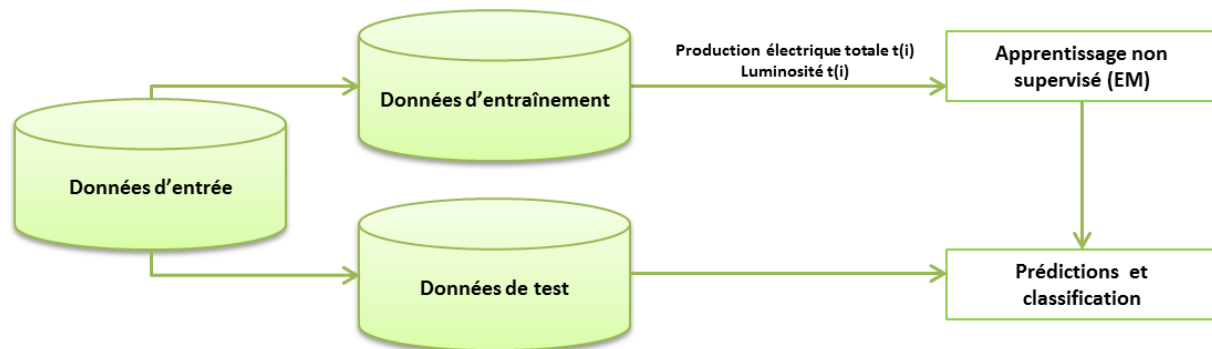


Figure 86 : Méthodes de classification des données pour la production décentralisée solaire

La répartition nous permet de définir si certaines familles ont besoin d'augmenter leur part dans leur représentation dans le set d'entraînement. L'objectif est de fournir un outil d'aide à la décision qui fournira à la fois des données numériques et par ses classification des tendances et des marges minimales et maximales calculées en fonction de la déviation standard et de la moyenne. Nous remarquons que le boosting par l'intermédiaire du GBT est adapté à la prédiction de production solaire avec plus de 92% de classification réussie (Tableau 48).

Energie à prédire	Sortie apprentissage	Moyenne	Déviati on standard	Répartiti on %
Production décentralisée solaire	0	1.0798	0.0794	18
		2.9921	1.3607	
		7.8636	3.8702	
	1	15.5819	6.9769	10
		21.4828	9.7723	
	2	46.0692	14.7446	24
		46.7811	15.2881	
	3	84.1565	12.5396	24
		101.8406	24.4283	
	4	140.2383	13.679	13
151.414		9.0872		

Tableau 48 : Nouvelles classes décrites par une moyenne, une déviation standard et leurs répartitions dans le set d'entraînement

Modèles	Taux de classification	Déviati on standard
SVM	90.28 +/- 1.1	2.3
Random Forest	88.75 +/- 1.3	1.4
Gradient Boosted Tree	92.09 +/- 0.9	1.5

Tableau 49 : Résultats de la classification ON-OFF par modèle mathématique pour une prédiction à t + 1 heure pour les jours de la semaine

5.4 Prédire les besoins en énergie dans le secteur industriel

5.4.1 Présentation du cas d'étude

Nous nous sommes basés sur le projet Adaptive IES [ADA-2014] qui nous a permis d'analyser les flux énergétiques au sein d'un site industriel pharmaceutiques et de dessiner des potentiels d'optimisation énergétiques basés sur les prédictions.

La modélisation de ces flux nous a permis de tirer des enseignements : une partie de l'énergie produite lors de la production est stockée sous forme d'eau chaude. Mais cette énergie n'est pas

valorisée. De plus, les températures, pression et humidité peuvent être optimisées par bâtiment et zone d'activité. Ces paramètres pourront ainsi évoluer en fonction de la météo extérieure, des plannings de production, des limites de confort et de sécurité en fonction des différentes zones dans les bâtiments. Nous pourrions ainsi comparer le potentiel de gains d'énergie en fonction de différents secteurs d'activité.

Dans cette première étude, nous nous sommes attachés à sélectionner le système le plus énergivore. Selon les coûts d'exploitation, le système HVAC est le système le plus énergivore, pour 30 à 40% de la consommation d'énergie totale du site. Cela représente entre 60000 et 80000 Franc de coût pour une seule unité, celle qui contrôle le bâtiment B1.

L'objectif des systèmes HVAC est d'assurer la gestion des boucles de régulation en pression, en température et en humidité dans les salles du bâtiment B2 DEBIOPHARM RESEARCH & MANUFACTURING à Martigny. Ces salles de production sont alimentées en air depuis quatre CTA. Une de prétraitement CTA00 et trois autres pour la gestion des différentes zones. Les CTA01 et CTA02 sont équipées d'extracteur d'air tandis que la CTA03 fonctionne en recyclage.

La gestion des boucles de régulation est assurée par un automate programmable de type Siemens S7-300 placé dans l'armoire électrique principale. Cet automate permet de contrôler en temps réel le fonctionnement de l'installation (états et mesures), de gérer les alarmes et les défauts, de renseigner les consignes de fonctionnement et de consulter les archives.

Le système HVAC comprend plusieurs étapes qui dépendent des contraintes mises sur les trois paramètres et de l'air extérieur qui entre dans le système (Figure 81). Ainsi l'air extérieur représente notre variable de transition : c'est en fonction des données météorologiques que certaines étapes du processus vont devoir s'enclencher. La pression, le gradient de pression entre différentes zones, le taux de renouvellement, la température sont des variables d'étapes. D'ailleurs, ces paramètres sont le plus souvent fixes dans le temps.

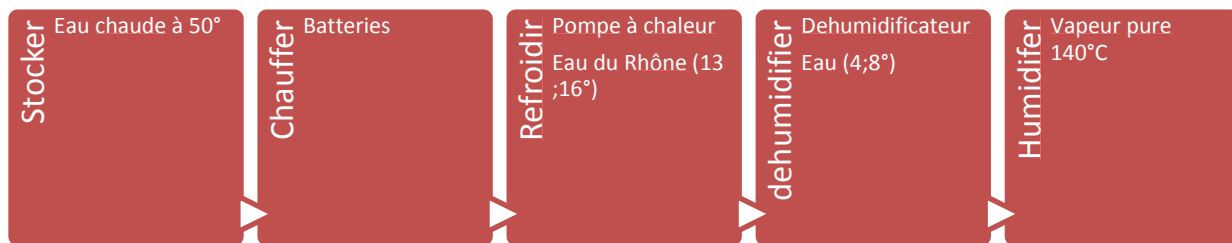


Figure 87 : Processus de régulation d'air sur le site industriel

Nous avons défini les spécifications sur la température, l'humidité et sur la pression selon des critères définis avec l'entreprise DebioPharm :

- ⇒ Confort du personnel
- ⇒ Qualité des produits

- ⇒ Qualité de la zone
- ⇒ Equipements
- ⇒ Processus de fabrication
- ⇒ Stockage de produits inflammables ou dangereux

Nous avons défini les contraintes sur les critères énoncés en lien avec nos paramètres de contrôle que sont la température, l'humidité et la pression :

- ⇒ La température et l'humidité relative sont définies en fonction du produit et des activités mais ils ne doivent pas interférer avec les standards de propretés requis
- ⇒ Une pression positive devrait être maintenue. Entre des locaux de classes différentes, une surpression comprise entre 10-15 Pa est recommandée.
- ⇒ La température et l'humidité ambiante ne doivent pas être trop élevées pour le confort du personnel.
- ⇒ Il convient que la surpression différentielle entre des salles ou zones propres de niveau de propreté différent se situe normalement entre 5 et 20 Pa, afin de faciliter l'ouverture des portes et d'éviter des flux de transfert inopinés à cause des turbulences.

Il est recommandé d'appliquer en régime réduit les spécifications suivantes :

- ⇒ Température : $\geq 5^{\circ}\text{C}$ et $\leq 25^{\circ}\text{C}$
- ⇒ Humidité relative : $\leq 75\%$
- ⇒ Surpressions : une surpression positive est requise. Les valeurs à appliquer doivent être évaluées selon une analyse de risque.

5.4.2 Prédire le besoin en chauffage, ventilation, humidité et pression

Notre vecteur d'entrée qui caractérise la consommation électrique globale regroupe les statistiques liées à la puissance active et réactive collectées au niveau du HVAC.

Pour définir la présence humaine dans le bâtiment, nous nous sommes basés sur le planning de production qui correspond au planning d'occupation du personnel dans l'entreprise. Nous avons retenu trois périodes :

- ⇒ [00 :00, 07 :00] et [17 :00, 23 :59] qui correspond à la période où le personnel n'est pas présent dans l'entreprise
- ⇒ [09 :00, 11 :00] et [14 :00, 15 :30] qui correspond à la période où l'ensemble du personnel est obligatoirement au sein de l'entreprise (nombre constant de personnes)
- ⇒ [07 :00, 09 :00] et [11 :00, 14 :00] [15 :30, 17 :00] qui correspond à la période où le nombre de personnes est variable

490 heures sont utilisées pour entraîner les modèles concernant la période où le nombre de personnes est constant. Pour cette période, la consommation électrique est représentée par 8 classes.

735 heures sont utilisées pour entraîner les modèles concernant la période où le nombre de personnes est variable et la consommation électrique y est représentée par 5 classes. L'ensemble des résultats sont décrits dans l'annexe 2 et sont résumés dans les tableaux 50 et 51.

Modèles	Taux de classification	Déviati on standard
Decorate	94.7 +/- 0.4	1.3
Logitboost	95.7 +/- 0.6	1.4
Bagging	92.8 +/- 0.9	1.5
J48	94.7 +/- 0.8	1.6
Random forest	94.5 +/- 0.7	1.3
Random tree	94.9 +/- 0.4	1.2

Tableau 50 : Résultats de la classification de la consommation électrique quand le nombre de personnes est constant

Modèles	Taux de classification	Déviati on standard
Decorate	96.9 +/- 0.9	2.3
Logitboost	97.1 +/- 0.4	1.4
Bagging	96.7 +/- 0.9	1.5
J48	97.5 +/- 0.7	1.6
Random forest	97.9 +/- 0.4	1.4
Random tree	98.5 +/- 0.3	1.1

Tableau 51 : Résultats de la classification de la consommation électrique quand le nombre de personnes est variable

Conclusion générale

Pour une participation active des charges, un pilotage plus précis, ponctuel et local sera mieux adapté permis par le développement des nouvelles technologies d'informations et de communications. La mise en place de ces plans de pilotage requiert certains points :

- ⇒ l'évolution des plans de délestages actuels vers des plans plus fins avec un traitement individuel des charges
- ⇒ la mise en place de suivi des usages énergétiques pouvant être pilotés
- ⇒ la possibilité d'implanter des algorithmes capables de prédire et de s'adapter aux fluctuations brutales locales de consommation et de production d'énergie.
- ⇒ l'acceptation de ces derniers points par les clients et leurs implications à l'intérieur du système.

Les études liées au pilotage de charges concernent essentiellement le chauffage ou plus récemment le refroidissement. Des algorithmes de pilotage ont pu être testés sur des données issues de logiciel puissants nommés dans le chapitre 1 comme Pleiades COMFIE ou bSol. Cette focalisation sur le chauffage et refroidissement a été liée, avant l'arrivée des microgrids, entre autre au manque de données réelles horaires sur le besoin en énergie d'un système. De plus, le pilotage du chauffage est lié au confort des personnes, notion difficile à généraliser pour tout type de personnes, dans des bâtiments énergétiquement très différents et dans des secteurs d'activité tout aussi différents.

La révolution attendue est celle des Smart Grids ou Réseaux Intelligents et plus récemment des microgrids. Elle répond à plusieurs challenges : faciliter l'intégration des énergies renouvelables, maîtriser la demande énergétique ou encore limiter le vieillissement du réseau. Un réseau intelligent est un réseau capable de communiquer et de prendre des décisions de façon autonome. Ceci est possible grâce aux **capteurs permettant de récolter les données** appropriées **en temps réel** à plusieurs niveaux du réseau. Au niveau du consommateur, la consommation d'électricité est historiquement enregistrée à l'aide de compteurs électromécaniques à simple ou double tarif. La lecture des compteurs s'effectue chez le consommateur, une ou deux fois par année. Aujourd'hui, les compteurs numériques sont capables de relever la consommation, la production ainsi que tout paramètre lié à l'état du réseau à chaque seconde près.

Nous avons présenté une méthode générale de prédiction locale des flux énergétiques d'un système tel qu'une maison, un restaurant ou une industrie. C'est l'un des objectifs de la collecte massive des données qui se situera au niveau des compteurs globaux d'électricité. Nous voulons élargir cette approche appliquée au chauffage/refroidissement, pour **se situer au niveau du compteur global d'électricité**, basé sur des données réelles, dans des secteurs différents (résidentiel, tertiaire, industriel) avec des activités différentes (restauration, bureaux, pharmaceutique). A partir de ce point de départ, plusieurs flux énergétiques sont représentés qui dépendent du système étudié.

Pour les systèmes avec une production décentralisée, une prédiction liée à ces flux permettra de créer des **scénarios de pilotage** liés à **l'eau chaude et au chauffage**. Une prédiction précise de ces charges devient ainsi nécessaire pour valider les scénarios liés au pilotage de l'eau chaude sanitaire ou du chauffage/refroidissement.

Nous avons voulu notre **approche la moins intrusive possible** avec comme information a priori le planning de production dans le secteur tertiaire et industriel qui nous permet d'affiner la qualité du set d'entraînement des modèles mathématiques et de définir des périodes de pilotage. Notre **méthodologie** se veut **générale** dans le cas de la prédiction d'énergie et a été testée sur un cas résidentiel, tertiaire et industriel.

Au vue de nos expériences pour la construction du système d'information et la difficulté de le maintenir tout au long de la durée de vie d'un bâtiment, de l'anxiété générée par la collecte du plus grand nombre de données sur toutes les personnes présentes sur la planète par des entreprises comme Google et d'un point de vue économique pour limiter le nombre de capteurs à installer, nous avons voulu créer un outil totalement non intrusif. **Seules les données au niveau du compteur global de consommation et de production sont collectées**. Aucun capteur de présence, de température intérieure ou extérieure local n'est utilisé pour nos prédictions. Seules les données de prédiction peuvent être envoyées à un niveau supérieur si besoin.

Nous avons commencé par créer notre système d'information qui permet la collecte et le stockage des données de consommation, de production et météorologiques. Nous avons testé plusieurs protocoles de communication sans fil ou câblés avec des collectes de données de la seconde à l'heure. Nous avons utilisé un logiciel d'analyse KNIME qui nous a permis de faire la connexion entre nos différentes bases de données et les applications qui découlent des différentes prédictions calculées et de traiter des masses de données importantes structurées et non structurées « Big Data ».

Une méthode non intrusive d'identification des systèmes de chauffage et d'eau chaude à partir de la puissance active et réactive a été créée qui permet au système à la fois de limiter le stockage des données de consommation et de fournir des informations complémentaires sur le suivi des consommations aux acteurs secondaires de notre outil, les consommateurs finaux.

Nous fournissons à la fois les résultats de nos prédictions sous forme statistiques et les résultats de la classification. Selon les acteurs, les besoins ne sont pas les mêmes : une classification est une information nécessaire et suffisante pour enclencher des plans de délestage plus massif de la part du fournisseur d'électricité en s'assurant du nombre de délestages possibles liés à l'eau chaude et au chauffage. Pour un pilotage en mode autoconsommation, une mesure précise de l'erreur est nécessaire.

Plusieurs méthodes et algorithmes ont été testés et nous pouvons conclure que les méthodes de sur ou sous-dimensionnement du set d'entraînement avec des tirages aléatoires semblent particulièrement adaptées à notre problématique en particulier le **Boosting** utilisé dans notre outil par l'algorithme Gradient Boosting Tree. Ces techniques sont très utiles dans le cadre de la prédiction énergétique puisque nous sommes souvent confrontés à sous-échantillonnage d'une partie des classes à **prédire** comme **les pics de consommations ou de productions** qui ne représentent que quelques % des heures dans la part totale du set d'entraînement ou de test. Le fait de tirer aléatoirement et de dupliquer

certaines données du set d'entraînement comble une partie de notre problème. Il permet ainsi dans le cas de la prédiction énergétique et en particulier des données qui sont sous-représentés dans le set d'entraînement d'avoir une plus grande visibilité pour nos algorithmes itératifs.

D'un point de vue système d'information, **l'analyse combinatoire Random Forest et Gradient Boosting Tree** nous permet de limiter le nombre de données en transit puisque des dizaines de variables suffisent dans certains cas pour garantir un même niveau de précision contrairement à un état initial avec plus de 4000 variables. Ceci est possible par l'interprétation des **ensembles d'arbres** par leurs **résultats statistiques** sur la création des arbres.

Les résultats que ce soit en classification ou statistiques, montrent que la collecte des données sur plusieurs mois à la minute est suffisante pour atteindre des niveaux de classification au-dessus des 95 % peu importe le secteur d'activité et le flux énergétique étudié et des MAE faibles en particulier pour l'eau chaude sanitaire et la consommation électrique globale. Dans le cadre du chauffage, il est difficile d'interpréter des interventions humaines telles que la mise en route d'un poêle à bois. Il nous faudrait compléter notre système par la mesure locale des données météorologiques en particulier si les apports solaires représentent une grande part dans l'apport des besoins en chauffage.

De plus l'identification de la présence humaine devient un paramètre essentiel tant pour le pilotage que pour la prédiction de la consommation énergétique. Pour créer un système non intrusif, il est plus facile d'appréhender cette problématique dans le secteur résidentiel et tertiaire en intégrant les plannings de production. Cela permet de cibler l'entraînement des modèles sur ce qui nous intéresse ici, prédire la consommation électrique globale pendant les heures d'ouverture. Nous pensons qu'il faudrait dans un premier temps **se focaliser sur les microgrids dans le secteur tertiaire et industriel où la variable humaine est plus prévisible par la définition des plannings d'occupation des entreprises.**

Pour la production décentralisée, il est encore plus utile de corriger les erreurs météorologiques liées à la prédiction de température et de luminosité. De plus, les modèles utilisent toujours les données prédites avant les données réelles historiques pour fournir une prédiction : l'apport des prédictions corrigées de stations aux alentours augmentent la prédiction de la production de 1.8 kWh en MAE pour avoir une erreur inférieure à 5%.

Nous avons également testé la précision de nos résultats en fonction d'une collecte des données à la minute ou à l'heure pour une prédiction à l'heure suivante pour chacun de nos flux énergétiques. Nous pouvons conclure qu'une **étude à la minute est nécessaire dans le cas du secteur résidentiel.**

Pour le chauffage, l'erreur est multipliée par deux entre des données collectées à la minute ou à l'heure. Pour la consommation électrique totale, l'erreur est 5 fois plus grande. Les habitudes de consommation ne sont pas aussi lissées que dans une entreprise qui suit un planning de production et d'occupation plus précis. Dans le cas du secteur tertiaire, les facteurs d'erreur sont de 0.25 ce qui est faible comparé au secteur résidentiel. Tout dépend de l'objectif de la prédiction : l'autoconsommation locale ou l'envoi de ces prédictions à un niveau plus élevé de pilotage pour agrégation des données.

Nous présentons nos résultats sur un horizon à l'heure qui est un pas de temps de référence énergétique en particulier pour le chauffage et l'eau chaude sanitaire. Notre outil peut ainsi proposer des scénarios de pilotage local basé sur les prédictions de consommations totales et par charges (chauffage et eau chaude sanitaire) et de productions ou envoyer ces informations de prédictions à un centre de pilotage. Un suivi des contraintes techniques à la minute et une prédiction au quart d'heure pour atteindre les marchés boursiers sont à intégrer pour fournir un outil complet.

Bibliographie

[ABB-2013] China State Grid and BYD launch world's largest battery energy storage station. Wire, Business. 2011. Fourniture d'un système de stockage d'énergie par batterie à l'un des plus grands distributeurs d'énergie suisses. ABB, 2013

[ADE-2004] Bâtiments de logements HQE, ADEME, Agence de l'environnement et de la maîtrise de l'énergie, 2004

[AES 2013] Différentes approches pour le monitoring du réseau de distribution. Bulletin 5/2016, 2016.

[ANG-2004] Angioletti R. et Desprez H., « Maîtrise de l'énergie dans les bâtiments-Définitions, usages, consommations », Techniques de l'Ingénieur, BE 9 020, 2004

[ANT-2013] Antonio J Jara, Luc Dufour, Gianluca Rizzo, Dominique Genoud, Alexandre Cotting, Yann Bocchi and François Chabbey, I-BAT : A Data-intensive Solution based on the Internet of Things to Predict Energy Behaviours in Microgrids , International Journal of Data Warehousing and Mining (IJDWM), 2014

[ARB-2013] Thomas Berthou, Développement de modèles de bâtiment pour la prévision de charge de climatisation et l'élaboration de stratégies d'optimisation énergétique et d'effacement, Ecole Nationale Supérieure des Mines de Paris, 2013

[BAK-2012] Bakshi, Kapil. "Considerations for big data: Architecture and approach." Aerospace Conference, 2012 IEEE. IEEE, 2012

[BEL-2010] Modélisation de la détection de présence humaine, Antoine BELCONDE, Université d'Orléans, 2010

[BEN-2015] Frederick Benaben. Conception de Système d'Information de Médiation pour la prise en charge de l'Interopérabilité dans les Collaborations d'Organisations. Artificial Intelligence [cs.AI]. Institut National Polytechnique de Toulouse, 2012

[BER-2010] Guide to Intelligent Data Analysis, How to Intelligently Make Sense of Real Data, Berthold, M.R., Borgelt, C., Höppner, F., Klawonn, F., 2010

[BLU-2014] Le smart metering pour maîtriser sa consommation énergétique, site internet : <https://www.bluewin.ch/fr/infos/suisse/2014/8/21/le-smart-metering-pour-maitriser-sa-consommation-d-energie.html> , 2014

[BRE-2001] Leo Breiman, Random forests. Machine Learning, 45(1):5–32., 2001

[BSO-2015] Logiciel de simulation énergétique, <http://www.physeos.ch>

[CAR-2010] Ensemble d'arbres de décision, <https://scaron.info/doc/intro-arbres-decision/intro.pdf>, Stéphane Caron, 2011

[CHE-2016] "Extended modeling procedure based on the projected sample for forecasting short-term electricity consumption." Che-Jung Chang, Jan-Yan Lin, Meng-Jen Chang, 2016

[CRE-2015] Communauté de régulation de l'énergie, <http://www.smartgrids-cre.fr> 2015

[DAT-2010] Stéphane Tufféry, Data Mining et statistique décisionnelle, Paris, éditions Technip, 3^e éd., 705 p, 2010

[DAV-2010] THE SMART GRID IN 2010: MARKET SEGMENTS, APPLICATIONS AND INDUSTRY PLAYERS, DAVID J. LEEDS |GTM RESEARCH, Smart Grid ,White Paper ,Deploying a smarter grid through cable solutions and services, NEXANS.2010

[DUF-2015] L.Dufour, D.Genoud, A.Jara, J.Treboux, B.Ladevie, JJ Bezian "A non-intrusive model to predict the flexible energy in a residential building", 2015 IEEE Wireless Communications and Networking Conference (WCNC) - Workshop - Energy Efficiency in the Internet of Things, and Internet of Things for Energy Efficiency, New orleans USA, 2015

[EDF-2013] La prévision de production éolienne et photovoltaïque à EDF Julien Najac, EDF R&D Séminaire In'Tech, INRIA, Grenoble, 2012

[ELF-2015] Projet ELF4U, - ELECTRICITY LOAD FORECASTING, <https://www.hevs.ch/fr/rad-instituts/institut-informatique-de-gestion/projets/elf4u-electricity-load-forecasting-10607>

[ENE-2004] Enertech, Bâtiments de logements HQE économes en énergie et en eau chaude, ADEME, Agence de l'environnement et de la maîtrise de l'énergie, 2004

[ENT-2015] Projet Européen ENTROPY, HES-SO Valais Wallis. Projet Européen dans le domaine de l'eenergy, <http://www.hevs.ch/fr/rad-instituts/institut-informatique-de-gestion/actualites/projet-europeen-dans-le-domaine-de-l-eenergy-10848>, 2016

[EU-2004] Eudeep project, Keynote paper, WP3- S5, « Methodologies to build market configurations suited to DER deployment – Trading of DER in competitive market », 2004

[EWE-2015] Association des entreprises électriques suisses, bulletin sur l'énergie éolienne, https://www.strom.ch/fileadmin/user_upload/Dokumente_Bilder_neu/010_Downloads/Basiswissen-Dokumente/15_Energie_%C3%A9olienne_fr.pdf

[FRE-1999] "Greedy Function Approximation: A Gradient Boosting Machine" by Jerome H. Friedman, 1999

[FRE-2001] RANDOM FORESTS Leo Breiman Statistics Department University of California Berkeley, CA 94720 January 2001

[GRA-2016] "Comparison of Photovoltaic plant power production prediction methods using a large measured dataset." G. Graditi, S. Ferlito, G. Adinolfi, 2016

[GRE-2014] Green value projet, 2014.

[GUE-2009] M. El Guedri, J. Bect, C. Lajaunie, G. Fleury, R. Seraoui, T. Garcia et A. Girard. RJMCMC point process sampler for single sensor source separation : an application to electric load monitoring. In Proceedings of the 17th European Signal Processing Conference, Glasgow, Royaume-Uni, 2009.

[GUE-2013] G. Guerassimoff, N. Maïzi. Smart Grids, Au-delà du concept, comment rendre les réseaux plus intelligents. s.l. : Mines ParisTech, 2013.

[GUL-2016] "Photovoltaic panel and array static models for power production prediction: Integration of manufacturers' and on-line data." Marko Gulin, Tomislav Pavlovic, Mario Vacsak, 2016

[GWE- 2016,] Institut Fraunhofer ISE , les statistiques de l'éolien en suisse et dans le monde

[HAR-1992] G.W. Hart. Nonintrusive appliance load monitoring. Proceedings of the IEEE, vol. 80,no. 12, pages 1870 –1891, 1992

[HEN-2007] Un système avancé de gestion d'énergie dans le bâtiment pour coordonner production et consommation, Duy-Long HA, Institut Polytechnique de Grenoble, 2007

[IBA-2013] IBATS, information system for collecting and processing data to control the energy consumption of buildings, http://www.hevs.ch/media/hes_so/document/0/2013.04.30_ibats_en.pdf

[IZA 2006] Plan d'expérience pour les différentes simulations énergétiques, 2006

[JAZ-2002] S. JAZOULI et S. BARROIS : Calage d'un modèle thermique de bâtiment "r3c2" à partir de la courbe de charge et de la météo. Rapport technique, EDF, 2002

[KEE-2010] SVM : Fast Training of Support Vector Machines using Sequential Minimal Optimization, by John C. Platt and Improvements to Platt's SMO Algorithm for SVM Classifier Design, by S. S. Keerthi et. al.

[LAD-2015] Cours de Bruno Ladevie, Bâtiment à énergie positive, Ecole des Mines Telecom Albi-Carmaux

[LEC-2015] Entreprise Leclanché en suisse, site internet : <http://www.leclanche.com/markets-solutions/microgrid-solutions-utilities-grid/microgrids/>

[LID-2010] Microgrids research: A review of experimental microgrids and test systems, N.W.A. Lidula *, A.D. Rajapakse 1.2010

[MIN -2015] Ministère de la transition écologique et solidaire, chiffre clés de l'énergie, <http://www.statistiques.developpement-durable.gouv.fr>, 2015

[MOU-2016] "Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review." Moulay Larbi Chalal , Medjdoub Benachir, Michael White, Raid Shrahily, 2016

[NIC-2014] Nice Grid le stockage d'énergie, site internet : 2014. <http://www.nicegrid.fr/nice-gridlestockagedenergie10.htm>. 2014

[NIK-2012] Virtual Power Plants general review: structure, application and optimization. L. Nikonowicz, J. Milewski. 2012

[NOS-2015] Cattell, Rick. "Scalable SQL and NoSQL data stores." ACM SIGMOD Record 39.4 (2011): 12-27.

[OBJ-2015] Logiciel de simulation énergétique, site internet : <http://www.objective-a.com/>

[OBS-2013] La production d'électricité d'origine renouvelable dans le monde, <http://www.energies-renouvelables.org/>

[OFF-2015] Consommation finale d'énergie par agent énergétique, OFFEN, office nationale de la statistique, 2015

[PAR-2004] Le perceptron multicouche et son algorithme de rétropropagation des erreurs Marc Parizeau Département de génie électrique et de génie informatique Université Laval, 10 septembre 2004

[PLE-2015] Logiciel de simulation énergétique, site internet : <http://www.izuba.fr/logiciel/pleiadescomfie>

[REB-2015] Modeling and forecasting energy consumption for residential buildings in Algeria using bottom-up approach." Rébha Ghedamsi, Nouredine Settou, Abderrahmane Gouareh, Adem Khamouli, Nadia Saifi, Bakhta Recioui, Boubekker Dokkar, 2015

[RID-2015] Ridi, A., Gisler, C. & Hennebert, J (2014). ACS-F2 - A new database of appliance consumption signatures. In Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of, pages 145-150. IEEE

[RIV-2013] D. Rivola, A. Giusti, M. Salani, A. E. Rizzoli, R. Rudel, L. M. Gambardella. A decentralized approach to demand side load management: the Swiss2Grif project. 2013

[ROD-2014] Non-intrusive appliance load curve disaggregation for service development, P Ferrez, P Roduit, Energy Conference (ENERGYCON), 2014 IEEE International, 813-820

[RTE-2004] Document RTE, « Memento de la sûreté du système électrique », 2004

[SHA-2010] "C4.5 Programs for machine learning", by J.R. Quinlan and in "SPRINT: A Scalable Parallel Classifier for Data Mining", by J. Shafer, R. Agrawal, M. Mehta, 2010

[SIA-380] Normes SIA, site internet : <http://www.sia.ch/fr/services/sia-norm/>

[SUR-2004] Surdu C. & Enacheanu O, Valorisation desservices systèmes” Rapport IDEA, 2004

[TRO-2015] Wikipédia, Site internet : https://fr.wikipedia.org/wiki/Mur_Trombe

[VIC-2013] Victoria Moreno, Luc Dufour, Antonio F Gómez Skarmeta, Antonio J Jara, Dominique Genoud, Bruno Ladevie and Jean-Jacques Beziau, Big data: the key to energy efficiency in smart buildings, Soft computing manuscript, 2015

[ZAM-2013] “A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production.” M. Zamo, O. Mestre, P. Arbogast, O. Pannekoucke, 2013

[ZHA-2016] Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique.” Fan Zhanga, Chirag Debb, Siew Eang Leeb, Junjing Yangb, Kwok Wei Shah, 2016

Table des illustrations

Figure 1: Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision	12
Figure 2: Emissions de CO2 issues de la combustion d'énergie (en millions de tonnes).....	16
Figure 3 : Evolution de la consommation des différentes sources d'énergie en France [MIN-2015].....	18
Figure 4 : Répartition de la consommation des différentes sources d'énergie en France, [MIN-2015].....	18
Figure 5 : Evolution de la consommation des différentes sources d'énergie en Suisse, 2014, [OFF-2014]	19
Figure 6 : Evolution de la consommation d'énergie finale par secteur en France de 1970 à 2014	20
Figure 7 : Evolution de la consommation d'électricité finale par secteur en France corrigée des variations climatiques	21
Figure 8 : Structure de la production d'électricité mondiale en 2013, [OBS-2013]	22
Figure 9 : Structure de la production électrique d'origine renouvelable mondiale en 2013, [OBS-2013] .	22
Figure 10 : Part des différentes sources de production dans le mix électrique Suisse en 2012.....	23
Figure 11 : Taux de croissance des différentes énergies entre 2013 et 2015, [MIN-2015]	24
Figure 12 : Projections de la part des différentes énergies renouvelables selon trois scénarios pour la Suisse [AES-2013]	24
Figure 13 : Réseau de production et d'acheminement d'électricité en Suisse, [AES-2013]	25
Figure 14 : Erreur de prédiction sur la production éolienne due à une erreur de prévisions météorologiques, [EDF-2013]	26
Figure 15 : Moyens de productions utilisés en fonction des ajustements possibles par rapport aux horizons de prédiction [EDF-2013].....	26
Figure 16 : Représentation du microgrid Nice Grid avec les différents niveaux de pilotage, [NIC-2014]...	28
Figure 17 : Domaines d'application de la thèse pour la mise en place d'un pilotage énergétique décentralisé	31
Figure 18 : Représentation d'un microgrid et de son insertion au smartgrid [IBA-2012].....	32
Figure 19 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision	34
Figure 20 : Représentation de l'enveloppe du bâtiment à prendre en considération pour un bilan énergétique	35
Figure 21 : Flux globaux d'énergie pour caractériser le besoin en chauffage [LAD-2013].....	37
Figure 22 : Types de transmissions possibles des flux énergétiques	38
Figure 23 : Représentation des évènements détectés dans le plan Puissance Active en abscisse, Puissance Réactive en ordonnée pour une maison à gauche, deux maisons à droite, Etats Unis, [HAR-1992].....	45
Figure 24 : Types d'installations différentes de récupération du rayonnement solaire pour produire de la chaleur [TRO-2015]	47
Figure 25 : Nouveau type de capteurs solaires qui intègre l'électricité et la chaleur [DUA-2014]	47
Figure 26 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision	50
Figure 27 : Représentation d'un réseau de neurones [BER-2010]	53

Figure 28: Paramètres de configuration de l'algorithme PNN qui correspondent à la délimitation de la zone de conflit [BER-2010]	54
Figure 29 : Représentation de la notion de marge γ dans le cas de séparation de deux classes [BER-2010]	55
Figure 30 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision	61
Figure 31 : Processus de mise en place du système d'information pour le pilotage énergétique pour nos systèmes d'étude.....	61
Figure 32 : Représentation des acteurs principaux dans notre outil	63
Figure 33 : Triangle des puissances indiquant la relation entre la puissance active, réactive et apparente	63
Figure 34 : Séparation de la facture d'électricité selon la fourniture et l'acheminement hors pic de puissance de l'électricité, la TVA, les différentes taxes et le pic de puissance qui comprend la part active et réactive pour le restaurant d'entreprise.....	64
Figure 35 : Quantités d'énergies électriques consommées moyennes sur l'année 2013 et leurs pics de puissance associés pour trois consommateurs d'un des sites d'étude, Canal 9 (Studio de télévision), Eticolle (Fabrication d'étiquettes) et le Mikado (Restaurant d'entreprise)	65
Figure 36 : Courbes de charges journalières moyennes en semaine et le weekend du restaurant d'entreprise moyennées sur le mois de Mars 2013	65
Figure 37 : Courbes de charges journalières moyennes en semaine et le weekend de l'entreprise de télévision Canal 9 moyennées sur le mois de Mars 2013...	
Figure 38 : Acteurs secondaires du système étudié qui est utilisé pour différencier les systèmes.....	66
Figure 39 : Caractérisation de la capacité d'effacement qui nécessite la prédiction du chauffage, de l'eau chaude sanitaire, de la consommation électrique globale et de la production solaire	67
Figure 40 : Représentation d'un beckhoff qui fait le lien entre les compteurs d'électricité et les bases de données.....	70
Figure 41 : Stations météorologiques connectées à notre système d'information	73
Figure 42 : Evolution des courbes de charges journalières moyennées sur le mois de Mars 2013 caractérisant les jours de la semaine, le samedi et le dimanche pour une industrie de fabrication d'étiquettes	75
Figure 43 : Echanges d'information mises en place dans le cadre du projet ELF4U, [ELF-2015]	76
Figure 44 : Représentation du traitement des données à l'intérieur du logiciel d'analyse KNIME [HES-2012].....	79
Figure 45 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision	81
Figure 46 : Puissance active et réactive à la seconde en sortie du compteur d'électricité dans le secteur résidentiel sur une phase	82
Figure 47 : Puissance active et réactive sans filtre au-dessus et après un filtre médian de taille 10	83
Figure 48 : Evènements détectés sur chacune des phases en sortie du compteur global entre le 1 Janvier et le 18 Juin 2014.....	86
Figure 49 : Identification des appareils électriques par usages à partir de la consommation électrique globale en sortie du compteur et d'une base de données d'apprentissage	88

Figure 50 : Donnée d'entrée et de sortie de l'algorithme de décomposition de la signature électrique d'une pompe à chaleur en 2 états.....	89
Figure 51 : Statistique et distribution de l'erreur de prédiction en luminosité et en température, station Aigle.....	90
Figure 52 : Statistique et distribution de l'erreur de prédiction en luminosité et en température, station Sion.....	90
Figure 53: Méthodologie de réduction de l'erreur prédite météorologique provenant de MétéoSuisse..	91
Figure 54 : Répartition des données d'entrée pour la correction des erreurs météorologiques	91
Figure 55: Statistiques et distribution de la nouvelle erreur de prédiction en luminosité et en température, station Aigle	92
Figure 56: Statistiques et distribution de la nouvelle erreur de prédiction en luminosité et en température, station Sion	92
Figure 57 : Méthodes de classification des données pour le besoin en eau chaude sanitaire et en chauffage	93
Figure 58 : Méthodes de classification des données pour la consommation électrique totale	95
Figure 59 : Méthodes de classification des données pour la production décentralisée solaire	96
Figure 60 : Représentation du regroupement des classes dans la classification de la production décentralisée solaire	97
Figure 61 : Méthode d'analyse de l'impact du nombre de variables d'entrée sur la précision de la prédiction	99
Figure 62 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision	101
Figure 63 : Schématisation de la maison test sous Pleiades COMFIE	102
Figure 64 : Caractérisation de la capacité d'effacement qui nécessite la prédiction du chauffage, de l'eau chaude sanitaire, de la consommation électrique globale et de la production solaire.....	103
Figure 65 : Répartition des données pour la prédiction de la consommation électrique globale, secteur résidentiel.....	105
Figure 66 : Répartition des données pour la prédiction du besoin en chauffage.....	109
Figure 67 : Besoin et prédiction en chauffage en sortie du GBT pour les jours de la semaine (Année, jour de l'année, heure)	111
Figure 68 : Besoin en chauffage en kWh et la prédiction en sortie du GBT sur nos données de test pour les jours du week-end	112
Figure 69 : Evolution de la précision de la prédiction en fonction du nombre de variables d'entrée pour la prédiction du besoin en chauffage.....	114
Figure 70 : Répartition des données pour la prédiction du besoin en eau chaude sanitaire	116
Figure 71 : Besoin en eau chaude sanitaire et la prédiction associée sur les heures représentant nos données de test.....	116
Figure 72 : Courbe ROC qui montre la performance du modèle à classier correctement l'allumage de l'eau chaude sanitaire pour les jours de la semaine	118
Figure 73 : Courbe ROC qui montre la performance du modèle à classier correctement l'arrêt de l'eau chaude sanitaire pour les jours du week-end	119

Figure 74 : Evolution de la précision de la prédiction en fonction du nombre de variables d'entrée pour la prédiction du besoin en eau chaude sanitaire	120
Figure 75 : Prédiction du besoin en eau chaude sanitaire en fonction d'une collecte des données à la minute ou à l'heure	121
Figure 76 : Caractérisation de la capacité d'effacement qui nécessite la prédiction de la consommation électrique pour chaque entreprise connectée au système d'information et de la production solaire....	122
Figure 77: Plan d'architecte du rez-de-chaussée du TechnoPôle à gauche, avec en orange la surface dédiée au restaurant d'entreprise	124
Figure 78 : Caractérisation de la capacité d'effacement qui nécessitent la prédiction du chauffage, de l'eau chaude sanitaire, de la consommation électrique globale et de la production solaire	125
Figure 79 . Représentation des sous-systèmes d'étude dans le secteur d'activité bureaux.....	125
Figure 80 : Caractérisation de la capacité d'effacement qui nécessite la prédiction du chauffage, de l'eau chaude sanitaire, de la consommation électrique globale et de la production solaire	126
Figure 81 : Consommation électrique globale et la prédiction en sortie du gradient boosted tree en Kwh dans le secteur tertiaire, restauration (Année, jour de l'année, heure)	128
Figure 82 : Consommation électrique globale et la prédiction en sortie du Random Forest Tree en kWh dans le secteur tertiaire, secteur d'activité bureaux	132
Figure 83 : Répartition des données pour la prédiction de la production décentralisée solaire.....	135
Figure 84 : Production solaire et la prédiction en sortie du GBT pour le mois de septembre avec la station de Sion distante de 15 km du site de production	136
Figure 85 : Production solaire et la prédiction en sortie du gradient boosted tree pour le mois de septembre avec la station de Sion distante de 15 km du site de production.....	137
Figure 86 : Méthodes de classification des données pour la production décentralisée solaire	137
Figure 87 : Processus de régulation d'air sur le site industriel.....	139

Table des tableaux

Tableau 1 : Puissance thermique dégagée par les habitants, norme SIA 380/1.....	40
Tableau 2 : Plan d'expérience réalisé sur les différentes simulations de chauffage [IZA-2006].....	43
Tableau 3 : Besoin en eau chaude sanitaire selon le nombre d'équipements associé à un nombre de personnes n [ANG-2004]	44
Tableau 4 : Tableau des fonctionnalités de l'outil.....	69
Tableau 5 : Données récoltées chaque seconde au niveau du transformateur basse tension pour mesurer la production photovoltaïque	70
Tableau 6 : Données récoltées chaque seconde au niveau de la transformatrice basse tension pour mesurer la production photovoltaïque	71
Tableau 7 : Données récoltés chaque seconde au niveau du compteur global dans le secteur résidentiel	71
Tableau 8 : Données météorologiques utilisées dans notre outil de prédiction	74
Tableau 9 : Techniques d'identification de la présence humaine	74
Tableau 10 : Données de présence utilisées dans notre outil de prédiction pour identifier les périodes de production	75
Tableau 11 : Accès aux données collectées par les capteurs dans le secteur tertiaire.....	77
Tableau 12 : Tableau récapitulatif des logiciels d'analyse de données [DAT-2015]	78
Tableau 13 : Base de données d'appareils électriques [RID-2015]	88
Tableau 14 : Base de données d'appareils électriques dans la maison test [DUF-2015].....	88
Tableau 15 : Répartition des différentes familles caractérisant la consommation électrique dans nos différents cas d'étude.....	96
Tableau 16 : Classes décrites par une moyenne, une déviation standard et leurs répartitions dans le set d'entraînement caractérisant la production solaire en kWh.....	97
Tableau 17 : Nouvelles classes décrites par une moyenne, une déviation standard et leurs répartitions dans le set d'entraînement	97
Tableau 18 : Résultats statistiques en sortie des différents algorithmes pour la prédiction de la consommation électrique globale pour le secteur résidentiel	105
Tableau 19 : Classes pour caractériser la consommation électrique globale pour les jours de la semaine pour le secteur résidentiel en kWh	106
Tableau 20 : Classes pour caractériser la consommation électrique globale dans le secteur résidentiel pour le week-end en kWh	106
Tableau 21 : Résultats de la classification de la consommation électrique globale dans le secteur résidentiel.....	106
Tableau 22 : Résultats statistiques en sortie du GBT en fonction de deux fréquences de mesures pour la prédiction de la consommation électrique globale pour le secteur résidentiel	107
Tableau 23 : Résultats statistiques en sortie du GBT en fonction du nombre de variables pour la prédiction de la consommation électrique globale pour le secteur résidentiel	108
Tableau 24 : Résultats numériques par modèle mathématique pour une prédiction du besoin en chauffage à t + 1 heure pour les jours de la semaine	110

Tableau 25 : Résultats de la classification représentant le chauffage pour les jours de la semaine	110
Tableau 26 : Résultats statistiques de la prédiction du besoin en chauffage pour le modèle semaine et week-end	111
Tableau 27 : Résultats de la classification représentant le chauffage pour les jours du week-end	113
Tableau 28 : Résultats numériques par modèle mathématique pour une prédiction du besoin en chauffage à t + 1 heure pour le week-end	113
Tableau 29 : Résultats statistiques en sortie du GBT en fonction du nombre de variables pour la prédiction du besoin en chauffage pour le secteur résidentiel	113
Tableau 30 : Résultats statistiques en sortie du GBT en fonction du nombre de variables pour la prédiction du besoin en chauffage pour le secteur résidentiel, modèle week-end	113
Tableau 31 : Résultats de la prédiction par modèle mathématique pour une prédiction à t + 1 heure pour les jours de la semaine	116
Tableau 32 : Résultats de la classification ON-OFF par modèle mathématique pour une prédiction à t + 1 heure pour les jours de la semaine	118
Tableau 33 : Résultats statistiques pour la prédiction du besoin en eau chaude sanitaire entre les modèles jours de la semaine et week-end.....	119
Tableau 34 : Résultats statistiques en sortie du GBT en fonction du nombre de variables pour la prédiction du besoin en eau chaude sanitaire pour le secteur résidentiel, modèle global.....	120
Tableau 35 : Résultats numériques par modèle mathématique pour une prédiction à t + 1 heure pour le modèle global, celui représentant les jours du lundi au vendredi et le week-end.....	121
Tableau 36 : Entreprises connectées au système d'information avec un relevé de la consommation à la seconde	123
Tableau 37 : Résultats statistiques en sortie des différents algorithmes pour la prédiction de la consommation électrique globale pour le secteur tertiaire, Restauration.....	128
Tableau 38 : Classes représentant la consommation électrique globale dans le secteur tertiaire, secteur d'activité restauration en kWh.....	129
Tableau 39 : Résultats de la classification par modèle mathématique.....	129
Tableau 40 : Résultats statistiques en sortie du Gradient Boosted Tree en fonction de deux fréquences de mesures pour la prédiction de la consommation électrique globale pour le secteur tertiaire, Restauration	130
Tableau 41 : Résultats statistiques en sortie du Gradient Boosted Tree en fonction du nombre de variables pour la prédiction de la consommation électrique globale pour le secteur tertiaire, Restauration	130
Tableau 42 : Résultats statistiques en sortie des différents algorithmes pour la prédiction de la consommation électrique globale pour le secteur tertiaire, secteur d'activité bureaux	131
Table 43 : Résultats de la classification par modèle mathématique pour une prédiction à t + 1 heure, secteur d'activité bureaux.....	133
Tableau 44 : Résultats statistiques en sortie du GBT en fonction d'une collecte à la minute ou à l'heure pour la prédiction de la consommation électrique globale pour le secteur tertiaire, secteur d'activité bureaux.....	133
Tableau 45 : Résultats statistiques en sortie du GBT en fonction du nombre de variables pour la prédiction de la consommation électrique globale pour le secteur tertiaire, secteur d'activité bureaux	133

Tableau 46 : Résultats numériques par modèle mathématique pour une prédiction de la production solaire à $t + 1$ heure, 1 station locale	135
Tableau 47 : Résultats numériques par modèle mathématique pour une prédiction du besoin en chauffage à $t + 1$ heure pour les jours de la semaine	136
Tableau 48 : Nouvelles classes décrites par une moyenne, une déviation standard et leurs répartitions dans le set d'entraînement	138
Tableau 49 : Résultats de la classification ON-OFF par modèle mathématique pour une prédiction à $t + 1$ heure pour les jours de la semaine	138
Tableau 50 : Résultats de la classification de la consommation électrique quand le nombre de personnes est constant.....	142
Tableau 51 : Résultats de la classification de la consommation électrique quand le nombre de personnes est variable	142

Title : Decentralized energy management by predictions

Keywords : Decentralized energy management; Energy prediction ; Information system ; Microgrid ; Decentralized energy prediction ; Machine learning ; Knime

Summary :

The world today is facing unprecedented environmental challenges and the global warming requires the states to react. The energy model which we inherit is today to be rethought. It is based on a variable power demand, constantly increasing and the supply of which depends on a point of production distant from the source of consumption. From an electrical point of view, the insertion of variable production from wind and photovoltaic increases pressure on the network.

This work presents a data-intensive solution to manage energy flux after a low transformer voltage named microgrid concept. A microgrid is an aggregation of building with a decentralized energy production and or not a storage system. These microgrids can be aggregate to create an intelligent virtual power plant.

However, many problems must be resolved to increase the part of these microgrid and the renewable resource in an energy mix. The physic model can not integrate and resolve in a short time the quickly variations. The intelligent district can be integrating a part of flexibility in their production with a storage system. This storage can be electrical with a battery or thermal with the heating and the hot water.

For a virtual power plant, the system can be autonomous when the price electricity prediction is low and increase the production provided on the market when the price electricity is high. For an energy supplier and with a decentralized production building distant of a low transformer voltage, regulations with a storage capacity enable a tension regulation. Finally, the auto-consumption becomes more and more interesting combined with a low electrical storage price and the result of the COP 21 in Paris engages the different country toward the energy transition. In these cases, a flexibility is crucial at the building level but this flexibility is possible if, and only if, the locally prediction are correct to manage the energy.

The main novelties of our approach is to provide an easy implemented and flexible solution to predict the consumption and the production at the building level based on the machine learning technique and tested on the real use cases in a residential and tertiary sector. A new evaluation of the consumption is realized: the point of view is energy and not only electrical. The energy consumption is decomposed between the heating consumption, the hot water consumption and the electrical devices consumption. A prediction every hour is provided for the heating and the hot water consumption to estimate the thermal storage capacity. A characterization of Electrical devices consumption is realized by a non-intrusive disaggregation from the global load curve. The heating and the hot water are identifying to provide a non-intrusive methodology of prediction. Every day, the heating, the hot water, the household appliances, the cooling and the standby are identified.

These predictions allow a shift of energy demand according the different predictions at $t + n$, n between 15 minutes and one day, and the system users (the final consumers and the aggregator). It is therefore to set up an information system non-intrusive, intelligent, robust and simple. To limit the intrusiveness of the systems, they are both adding intelligence as identify useful devices from a single measurement point but also collect and analyze information locally. The private data is not transmitted to the outside. Only the information's of energy predictions are sent at a higher level for the aggregation of the district data. The intelligence is also in learning method predictive: Support Vector Machine, Decision Tree Ensemble and Hidden Markov Model. Robustness is studied from a technology point of view (more communication protocol have been tested), technical (Several methods for collecting) and a data storage point of view (the limit frequency of collection). The ease of use results in a simple installation. Limit the frequency of collection and minimize the number of prediction model input data are essential points. One goal is to find generalization points to limit the number of sensors: identical load curve useful specific uses in a particular sector of activity, presence detection from the global burden of curve electricity. Evaluation of the results and relevance of the predictions made on the basis of reference models such as detailed thermal models for heating or SIA standards. These reference models need a priori information's about the different buildings: activity sector, construction year, surface, people number. The Information systems are implemented on real cases and are present at different scales: a house, a tertiary industrial district, an industrial site.

Titre : Contribution à la mise au point d'un pilotage énergétique décentralisé par prédiction

Mots clés : Pilotage énergétique décentralisé ; Prédiction énergétique ; Système d'information ; Production décentralisée ; Analyse intelligente des données ; Knime

Résumé :

Réchauffement climatique, accidents nucléaires, généralisation de l'électrification.... Le monde de l'énergie est face à des défis importants qui nous ramènent au fonctionnement global de notre société et au rapport des citoyens avec celle-ci : comment satisfaire les besoins en énergie d'une population de 9 milliards d'êtres humains en 2050, de façon économiquement viable et en minimisant l'impact sur l'environnement. En Europe, les États membres ont convenu d'un nouveau cadre pour l'énergie et le climat à l'horizon 2030, comprenant des objectifs stratégiques à l'échelle européenne pour la période 2020-2030. Ces objectifs visent à aider l'Union à mettre en place un système énergétique plus compétitif, plus sûr et plus durable et à réaliser son objectif à long terme de réduction des gaz à effet de serre fixé pour 2050. Les objectifs clés à l'horizon 2030 sont les suivants: réduire d'au moins 40 % les émissions de gaz à effet de serre (par rapport aux émissions de 1990), atteindre une part d'au moins 27 % d'énergies renouvelables et améliorer d'au moins 27 % l'efficacité énergétique.

Pour répondre à nos engagements, la transition énergétique passera par une maîtrise de l'insertion de production d'énergie d'origine éolienne et photovoltaïque. D'autant plus que la majorité des installations solaires installées en France et en Suisse sont de faibles puissances, raccordées au réseau basse tension, lieu où l'information manque. Les modèles prédictifs historiques centralisés et paramétriques ont du mal à appréhender les variations brutales de productions et de consommations et un maillage plus fin devient utile pour comprendre les scénarios d'occupation du bâtiment.

Notre objectif dans cette thèse est de **favoriser la maîtrise de la demande locale en énergie**. Cela passera par la mise en place d'un **bilan énergétique prédictif** pour l'heure suivante qui permettra d'avoir des possibilités de **scénarios de surproductions ou surconsommations** sur un quartier donné à **court terme au niveau de chaque flux énergétiques (t + 1 heure)**. Cela implique une prédiction fine des différents flux énergétiques d'un système en particulier le chauffage et l'eau chaude qui représentent le plus gros potentiel de flexibilité dans les bâtiments **et permettra ainsi de limiter le besoin en stockage électrique**. Le besoin en énergie d'un système comme une maison est ainsi décomposé en un besoin en **chauffage**, en un besoin en **eau chaude sanitaire**, en **luminaires**, en besoin de **ventilation** et en **usages spécifiques électriques** utiles. Le système peut posséder une **production décentralisée** et un système de stockage ce qui augmentera sa capacité d'effacement.

Pour réaliser un bilan, nous devons calculer les différents flux énergétiques à l'intérieur de notre système : les déperditions par l'enveloppe et la ventilation, les gains internes solaires, des personnes et des appareils, le stockage, la production d'eau chaude sanitaire (ECS) et les usages spécifiques électriques utiles (USE). Sur certains de ces points, nous pouvons évaluer assez précisément et en fonction du temps les quantités d'énergie échangées. Pour les autres (ECS, USE, gains internes, stockage), la bibliographie nous donne que des méthodes globales et indépendantes du temps. Il n'est donc pas possible d'envisager une méthode correspondant au pas de temps souhaité. Ceci impose la mise au point d'une méthode prédictive et apprenante dont nos modèles de simulation énergétique

seront le point de référence. Il n'en reste pas moins que ces modèles permettent la compréhension du comportement énergétique du système.

L'outil devrait être non intrusif, personnalisé, robuste et simple. Pour limiter le caractère intrusif de l'outil, il s'agit à la fois d'ajouter de l'intelligence, comme par exemple l'identification des appareils utiles à partir d'un seul point de mesure mais aussi la collection et l'analyse d'informations locales. Les données privées ne sont pas transmises vers l'extérieur. Seules les informations de prédictions énergétiques sont envoyées à un niveau supérieur pour agrégation des données des quartiers. Nous testerons également différentes méthodes de prédictions : les méthodes paramétriques comme les régressions linéaires ou les méthodes non paramétriques par apprentissage comme l'utilisation des réseaux de neurones ou des arbres de décision.

Notre outil a été testé sur des cas réels dans le secteur résidentiel, tertiaire et industriel. Sur le secteur résidentiel, l'ensemble des flux énergétiques (chauffage, eau chaude, électrique, production décentralisée renouvelable) ont été caractérisés et prédits. Le taux d'exactitude des précisions des classifications atteint 98% +/-1.2% dans le cadre de la classification.

Ensuite, sur le secteur tertiaire, une prédiction électrique est réalisée chaque quart d'heure par secteur d'activité (bureaux, restauration, studio de télévision). Les résultats de classification peuvent atteindre 98.4% +/-0.6% dans le secteur d'activité bureaux avec un an de données d'entraînement. Enfin, nos études ont porté sur le secteur industriel (pharmaceutique) sur la prédiction des HVAC (gestion de la température, humidité et pression) à partir de la température extérieure et du planning de production avec des résultats de 98.5% +/- 1.6%.

Titre : Contribution à la mise au point d'un pilotage énergétique décentralisé par prédiction

Résumé :

Comment satisfaire les besoins en énergie d'une population de 9 milliards d'êtres humains en 2050, de façon économiquement viable tout en minimisant l'impact sur l'environnement. Une des réponses est l'insertion de production d'énergie propre d'origine éolienne et photovoltaïque mais leurs totales dépendances aux variations climatiques accentuent une pression sur le réseau. Les modèles prédictifs historiques centralisés et paramétriques ont du mal à appréhender les variations brutales de productions et de consommations.

La révolution internet permet aujourd'hui une convergence entre le numérique et l'énergie. En Europe et depuis cinq ans, l'axe d'étude est celui de la maîtrise locale de l'électricité. Ainsi plusieurs quartiers intelligents ont été créés et les modèles utilisés de pilotage et de prédiction restent souvent la propriété des partenaires des projets. Dans cette thèse, il s'agit de réaliser un bilan énergétique chaque heure pour prédire l'ensemble des vecteurs énergétiques d'un système. Le besoin en énergie d'un système comme une maison est décomposée en un besoin en chauffage, en un besoin en eau chaude sanitaire, en un besoin en luminaires, en besoin de ventilation et en usages spécifiques électriques utiles. Le système peut posséder une production décentralisée et un système de stockage ce qui augmentera sa capacité d'effacement. Pour le centre de pilotage, l'objectif est d'avoir une possibilité de scénarios de surproductions ou surconsommations sur un quartier donnée à court terme. Nous considérerons dans cette thèse un horizon à l'heure pour notre bilan énergétique. Cela implique une prédiction fine des différents flux énergétiques d'un système en particulier le chauffage et l'eau chaude qui représente le plus gros potentiel de flexibilité dans les bâtiments.

Pour réaliser un bilan, nous devons calculer les différents flux énergétiques à l'intérieur de notre système : les déperditions par l'enveloppe et la ventilation, les gains internes solaires, des personnes et des appareils, le stockage, la production d'eau chaude sanitaire, les usages spécifiques électriques utiles. Sur certains de ces points, nous pouvons évaluer assez précisément et en fonction du temps les quantités d'énergie échangées. Pour les autres (ECS, USE, gains internes, stockage), la bibliographie nous donne que des méthodes globales et indépendantes du temps. Il n'est donc pas possible d'envisager une méthode correspondant au pas de temps souhaité. Ceci impose la mise au point d'une méthode prédictive et apprenante dont nos modèles de simulation énergétique seront le point de référence. Il n'en reste pas moins que ces modèles permettent la compréhension du comportement énergétique du système.

L'outil se devra non intrusif, personnalisé, robuste et simple. Pour limiter le caractère intrusif de l'outil, il s'agit à la fois d'ajouter de l'intelligence comme par exemple l'identification des appareils utiles à partir d'un seul point de mesure mais aussi la collection et l'analyse d'informations localement. Les données privées ne sont pas transmises vers l'extérieur. Seules les informations de prédictions énergétiques sont envoyées à un niveau supérieur pour agrégation des données des quartiers. L'intelligence est également au niveau des prédictions réalisées issues de méthodes d'apprentissage comme l'utilisation des réseaux de neurones ou des arbres de décision. La robustesse est étudiée d'un point de vue technologie (plusieurs protocoles de communication ont été testés), techniques (plusieurs méthodes de collecte) et d'un point de vue du stockage de données (limiter la fréquence de collecte). La simplicité d'usage engendre une simplicité d'installation minimiser le nombre de données d'entrée tout en gardant une précision souhaitable sera notre principal axe d'optimisation.

Mots-clés : Pilotage énergétique décentralisé ; Prédiction énergétique ; Système d'information ; Production décentralisée ; Analyse intelligente des données ; Knime