

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.

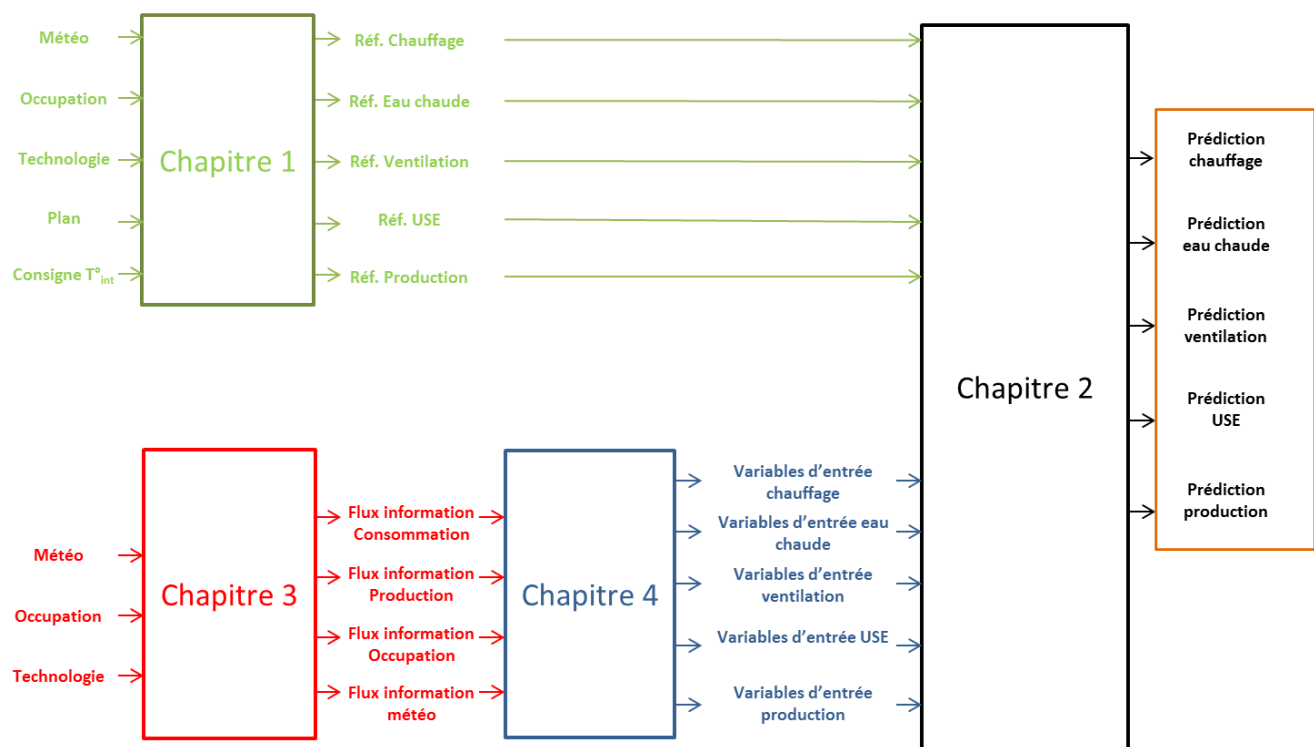


Figure 62 : Aide pour la lecture du document associée aux différents chapitres et au processus générale d'aide à la décision

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.

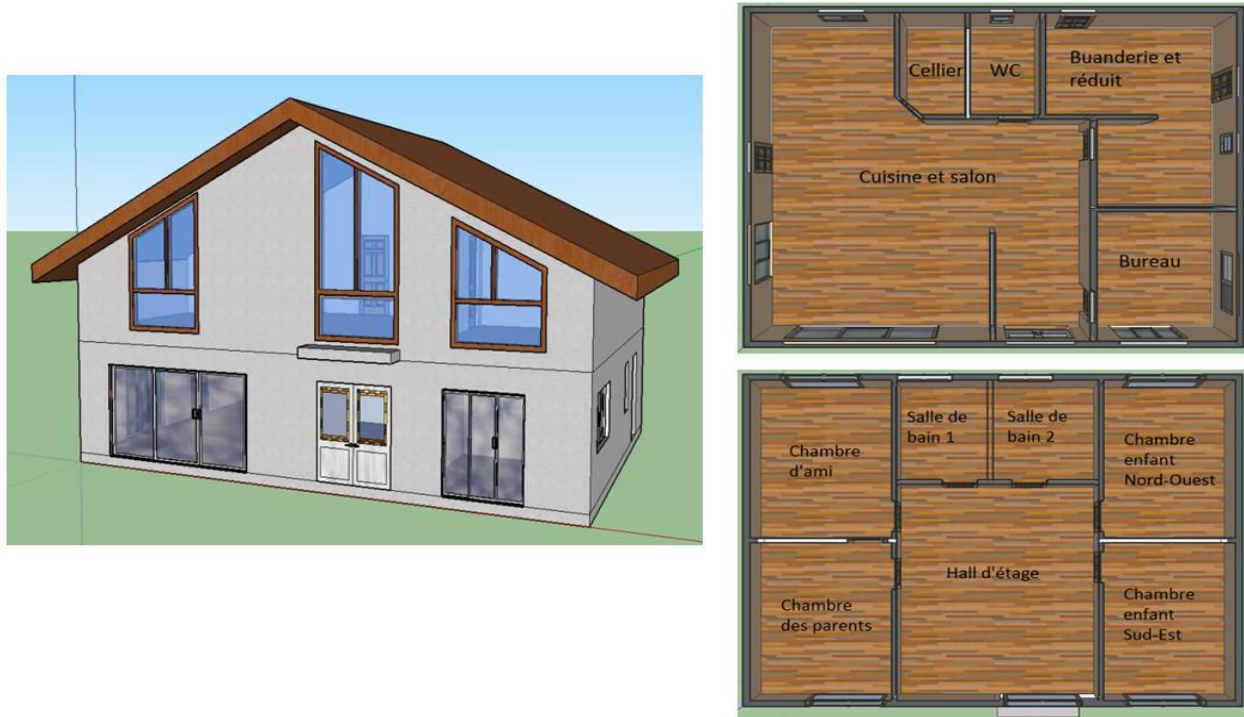


Figure 63 : Schématisation de la maison test sous Pleiades COMFIE

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).

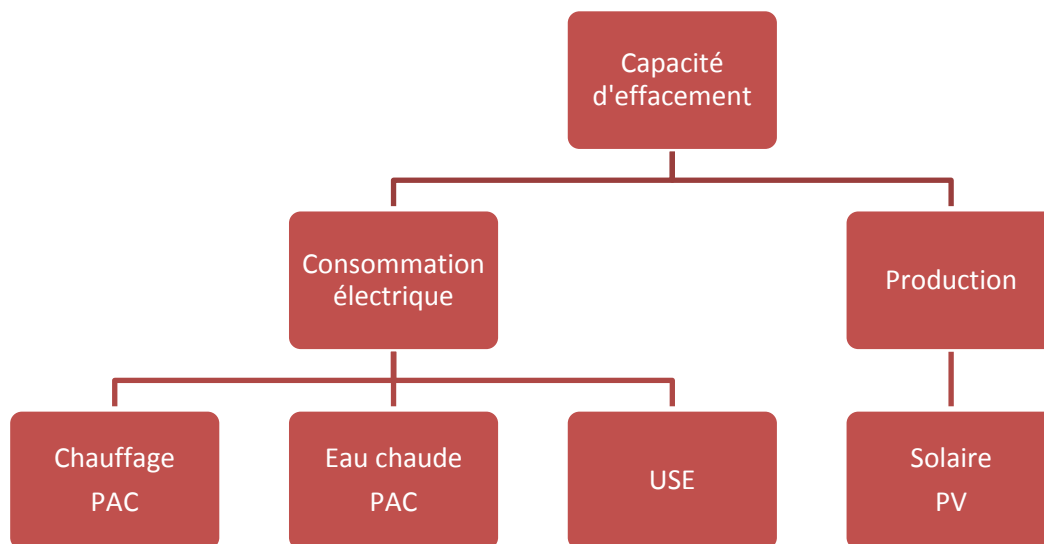


Figure 64 : Caractérisation de la capacité d'effacement qui nécessite la prédiction du chauffage, de l'eau chaude sanitaire, de la consommation électrique globale et de la production solaire

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 désagréations 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 $t(i)$:

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

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 :

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

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.

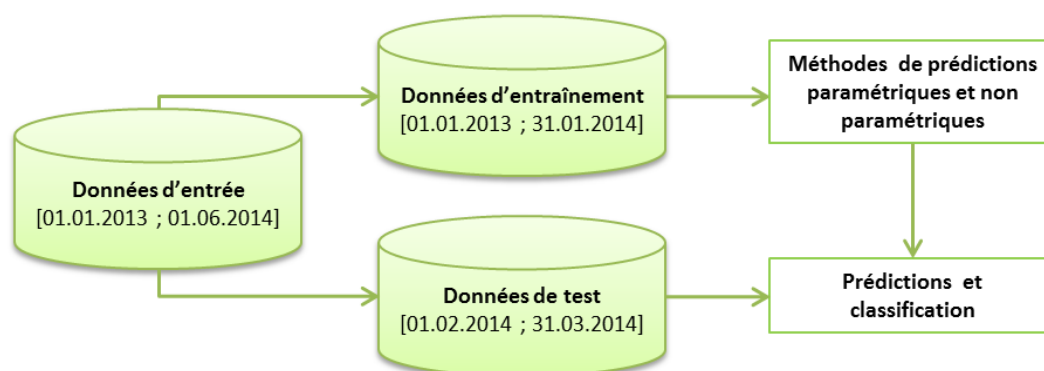


Figure 65 : Répartition des données pour la prédiction de la consommation électrique globale, secteur résidentiel

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	Déviati on 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	Déviati on 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_{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 classer 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.

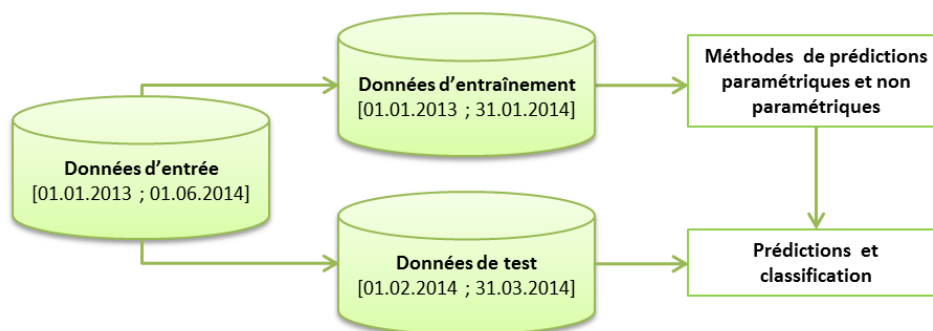


Figure 66 : Répartition des données pour la prédiction du besoin en chauffage

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

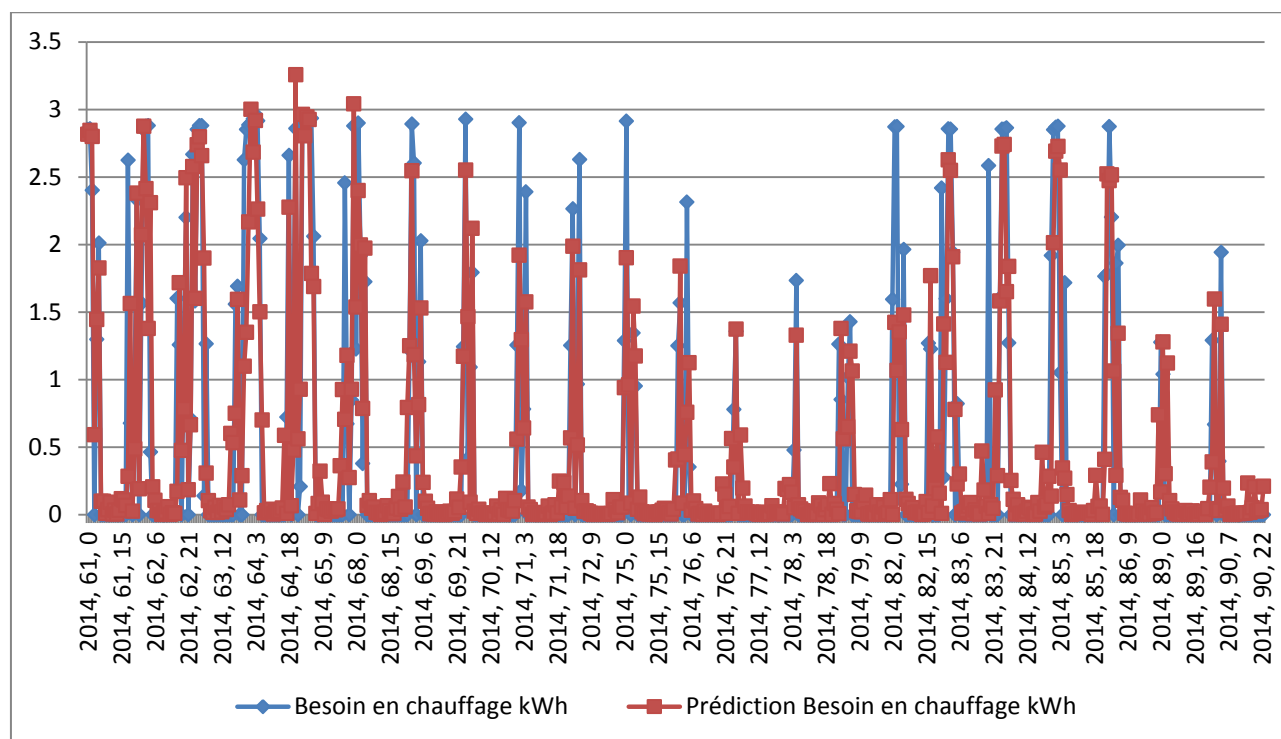


Figure 67 : Besoin et prédiction en chauffage en sortie du GBT pour les jours de la semaine (Année, jour de l'année, heure)

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

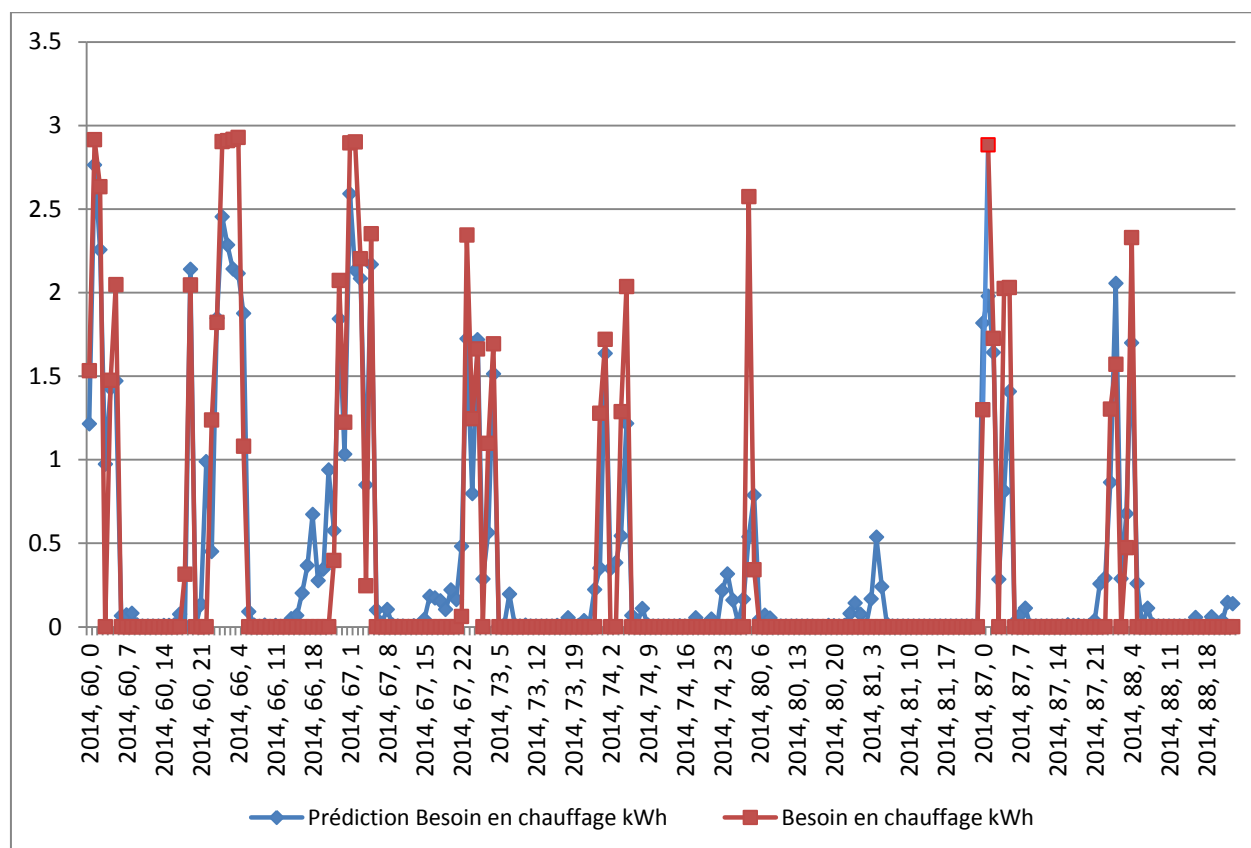


Figure 68 : Besoin en chauffage en kWh et la prédiction en sortie du GBT sur nos données de test pour les jours du week-end

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

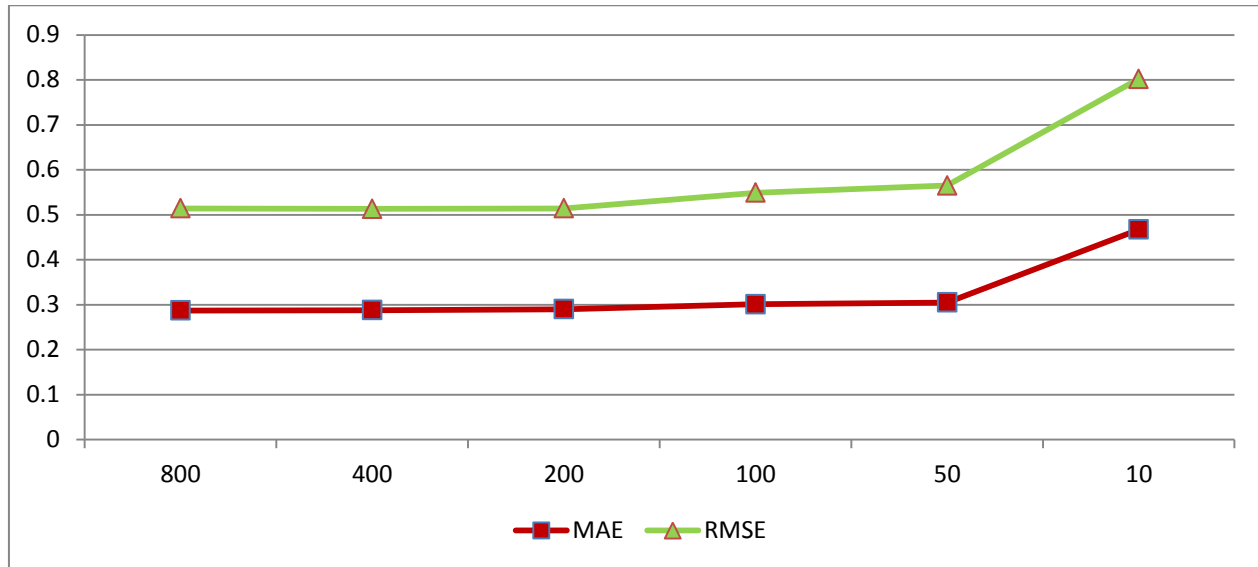


Figure 69 : Evolution de la précision de la prédiction en fonction du nombre de variables d'entrée pour la prédiction du besoin en chauffage

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).

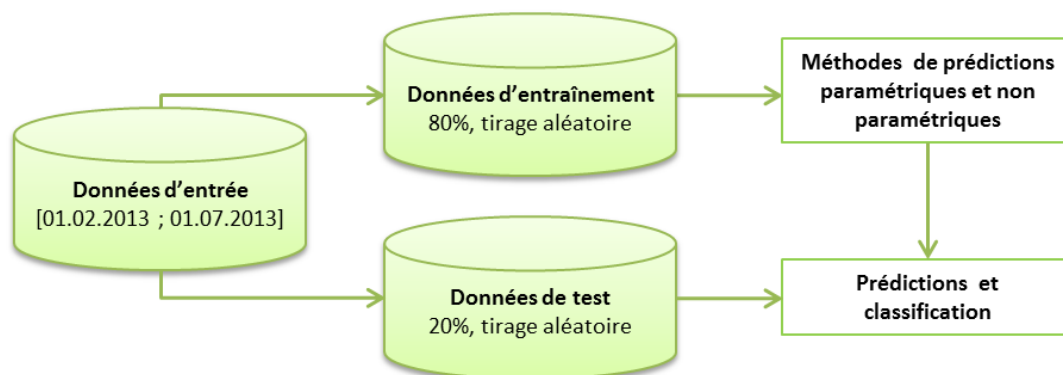


Figure 70 : Répartition des données pour la prédiction du besoin en eau chaude sanitaire

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

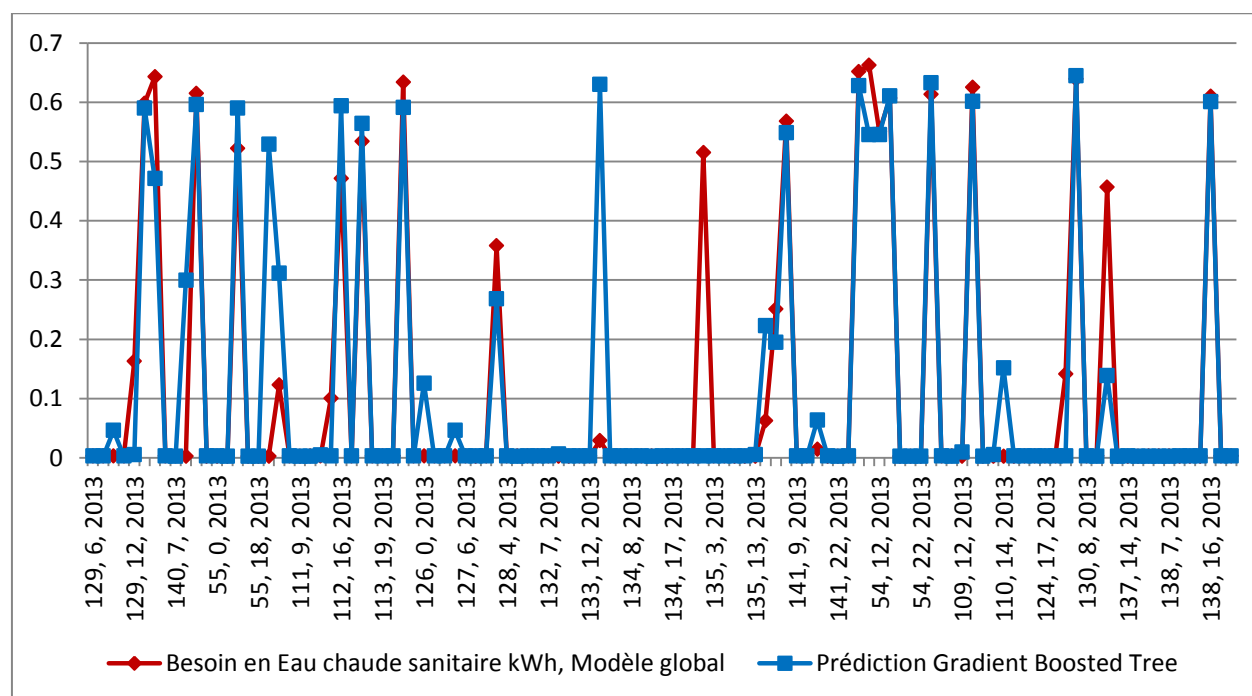
 Tableau 31 : Résultats de la prédiction par modèle mathématique pour une prédiction à $t + 1$ heure pour les jours de la semaine


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

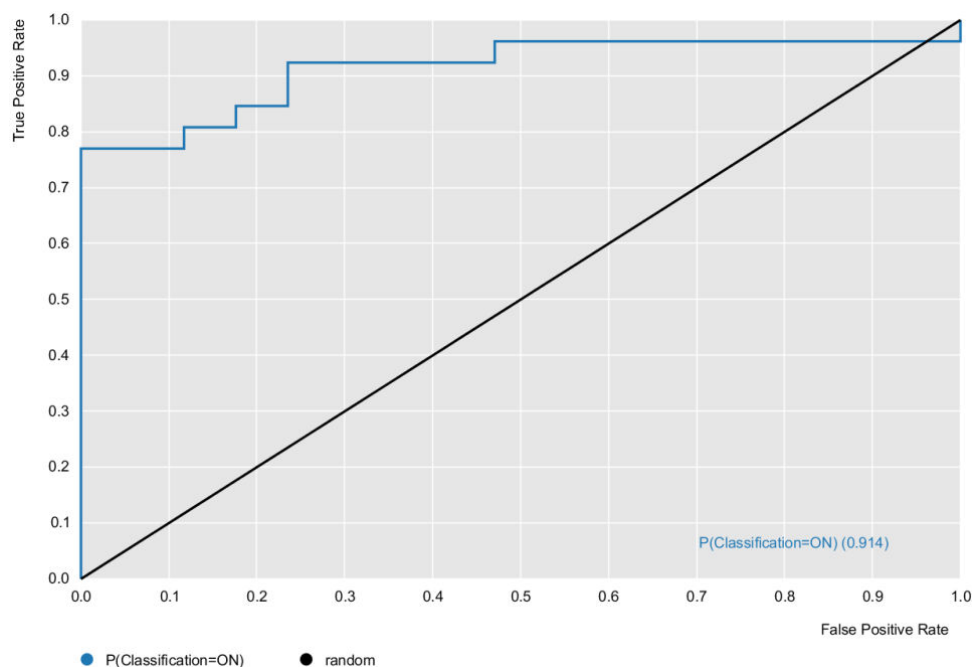


Figure 72 : Courbe ROC qui montre la performance du modèle à classer correctement l'allumage de l'eau chaude sanitaire pour les jours de la semaine

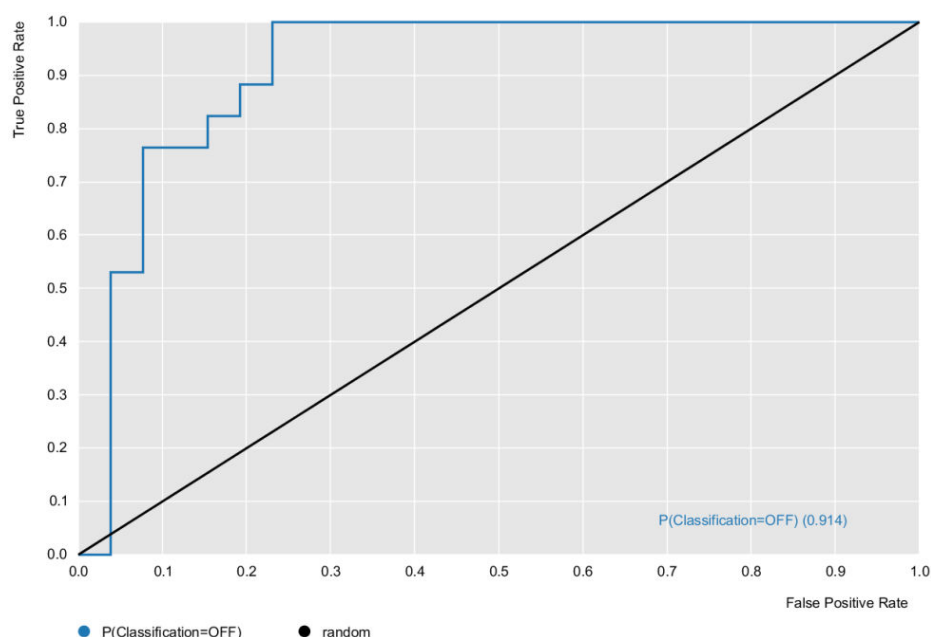


Figure 73 : Courbe ROC qui montre la performance du modèle à classer correctement l'arrêt de l'eau chaude sanitaire pour les jours du week-end

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

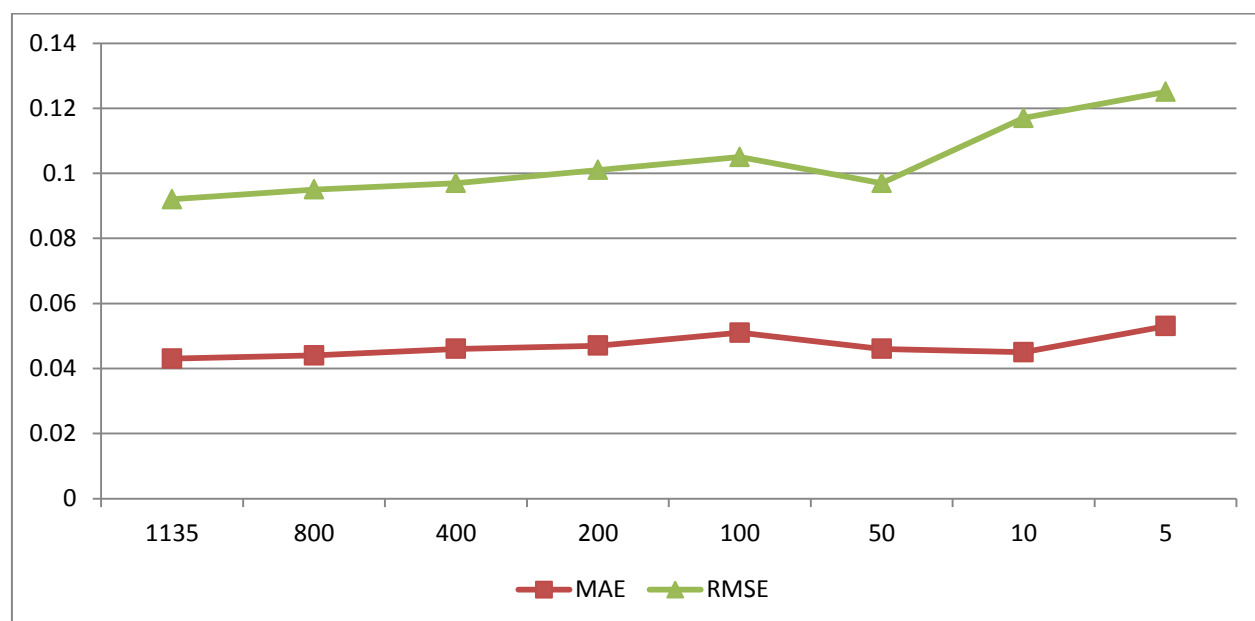


Figure 74 : Evolution de la précision de la prédiction en fonction du nombre de variables d'entrée pour la prédiction du besoin en eau chaude sanitaire

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

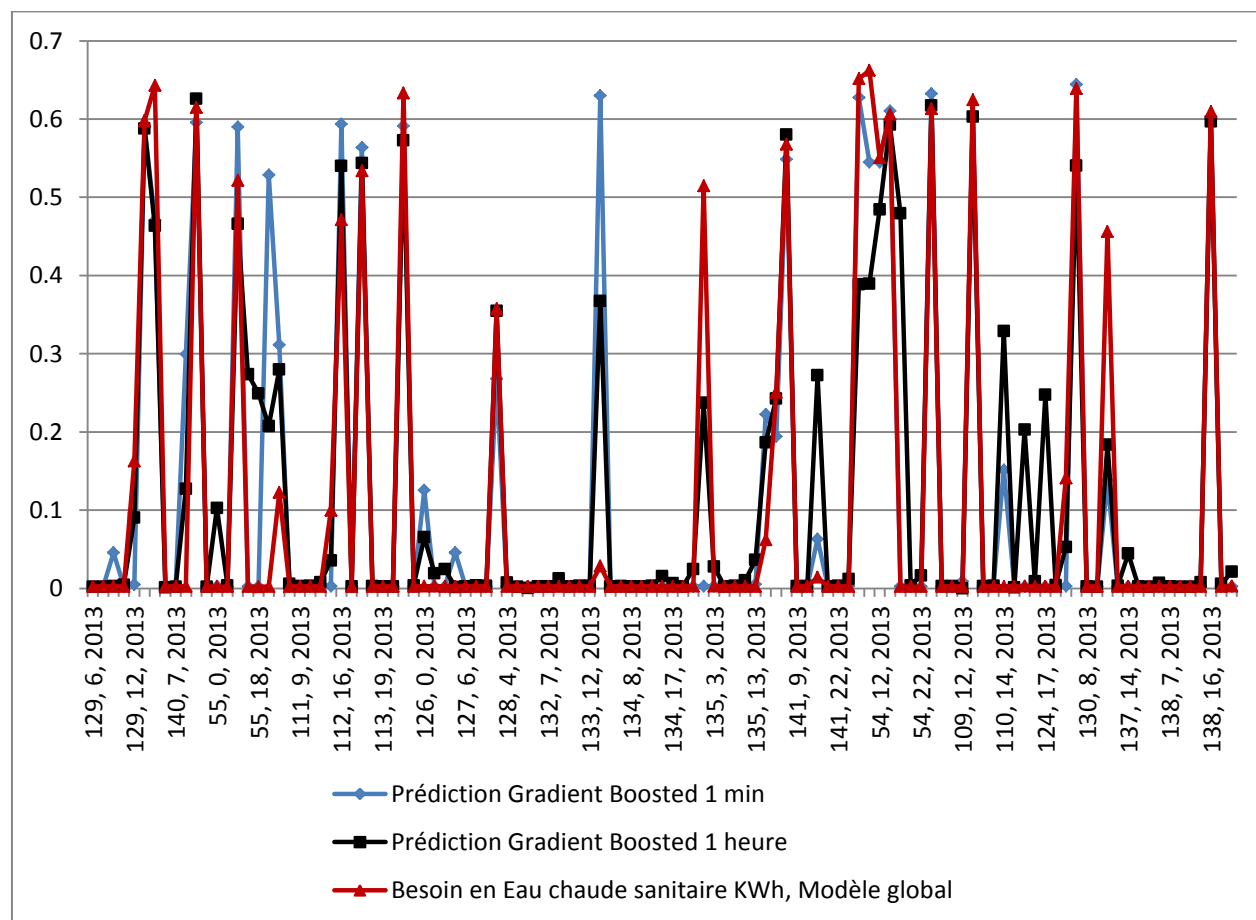


Figure 75 : Prédiction du besoin en eau chaude sanitaire en fonction d'une collecte des données à la minute ou à l'heure

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.

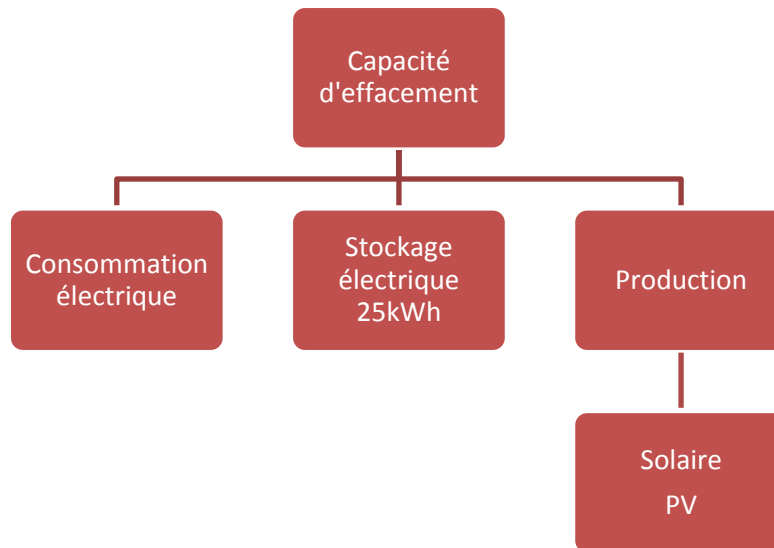


Figure 76 : Caractérisation de la capacité d'effacement qui nécessite la prédiction de la consommation électrique pour chaque entreprise connectée au système d'information et de la production solaire

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.

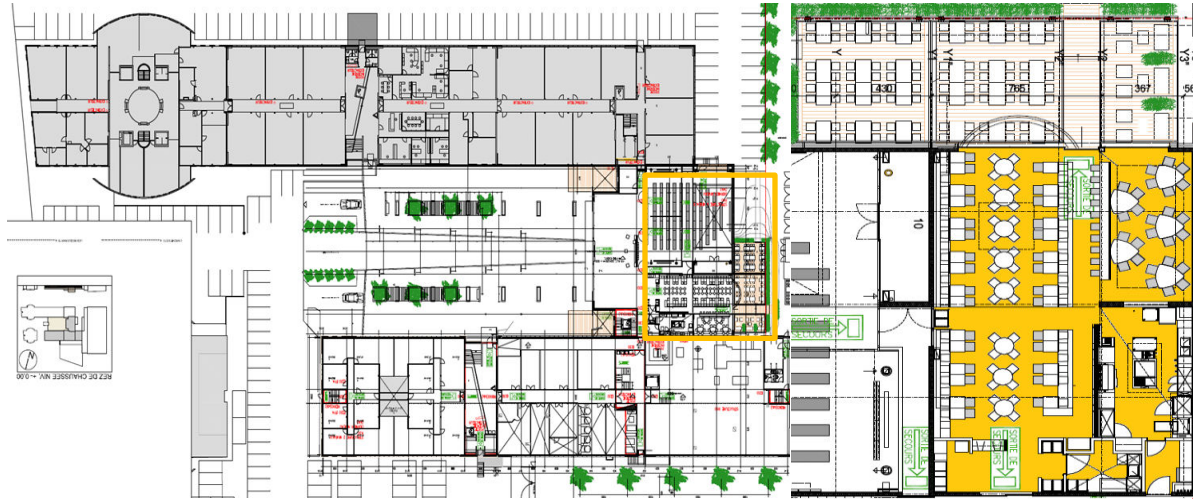


Figure 77: Plan d'architecte du rez-de-chaussée du TechnoPôle à gauche, avec en orange la surface dédiée au restaurant d'entreprise

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.

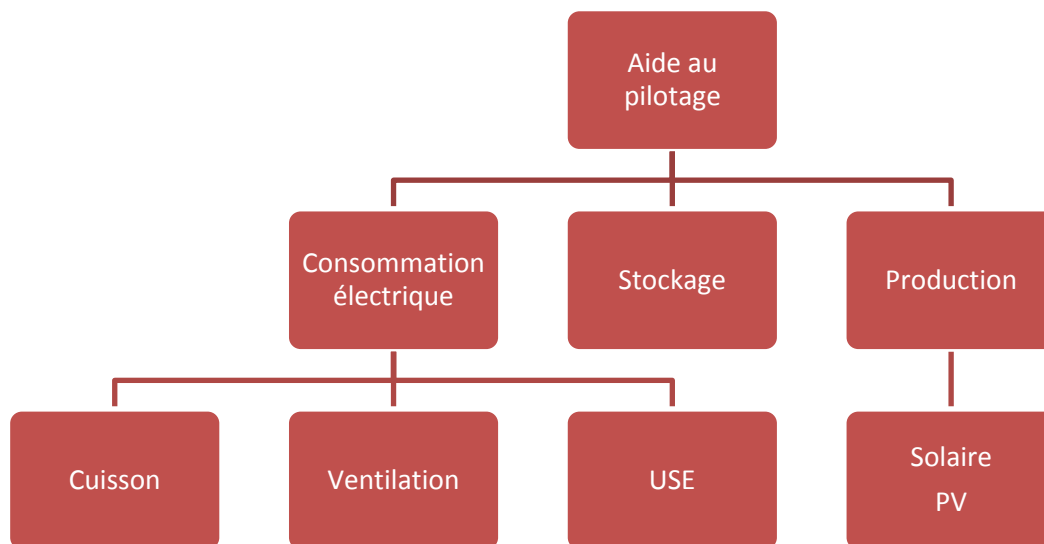


Figure 78 : Caractérisation de la capacité d'effacement qui nécessitent la prédiction du chauffage, de l'eau chaude sanitaire, de la consommation électrique globale et de la production solaire

Le deuxième cas d'étude est l'institut entrepreneuriat 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 :

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

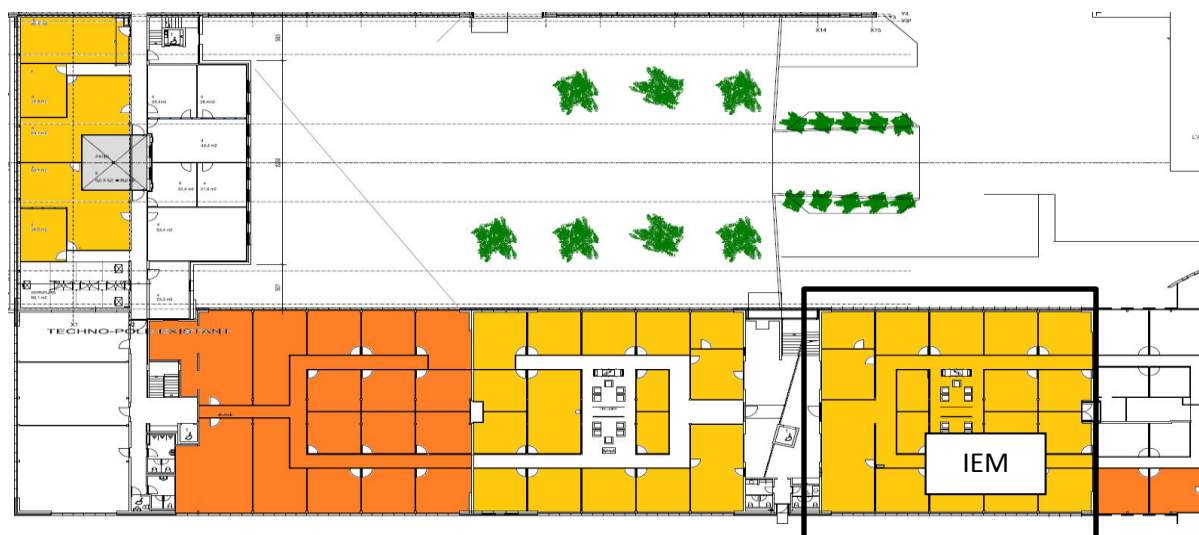


Figure 79 . Représentation des sous-systèmes d'étude dans le secteur d'activité bureaux

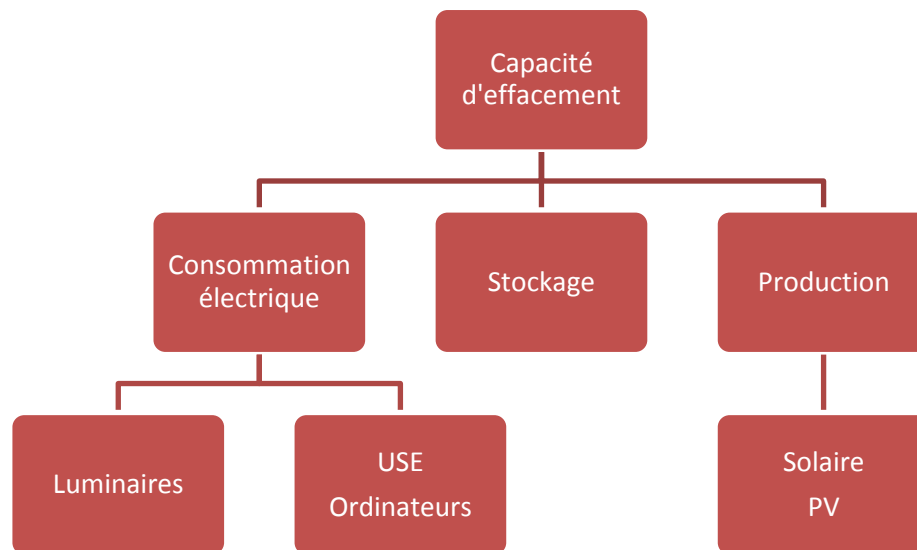


Figure 80 : Caractérisation de la capacité d'effacement qui nécessite la prédiction du chauffage, de l'eau chaude sanitaire, de la consommation électrique globale et de la production solaire

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 , j_s 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 :

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

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

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

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

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

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

$$\text{Classes}_c]$$

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

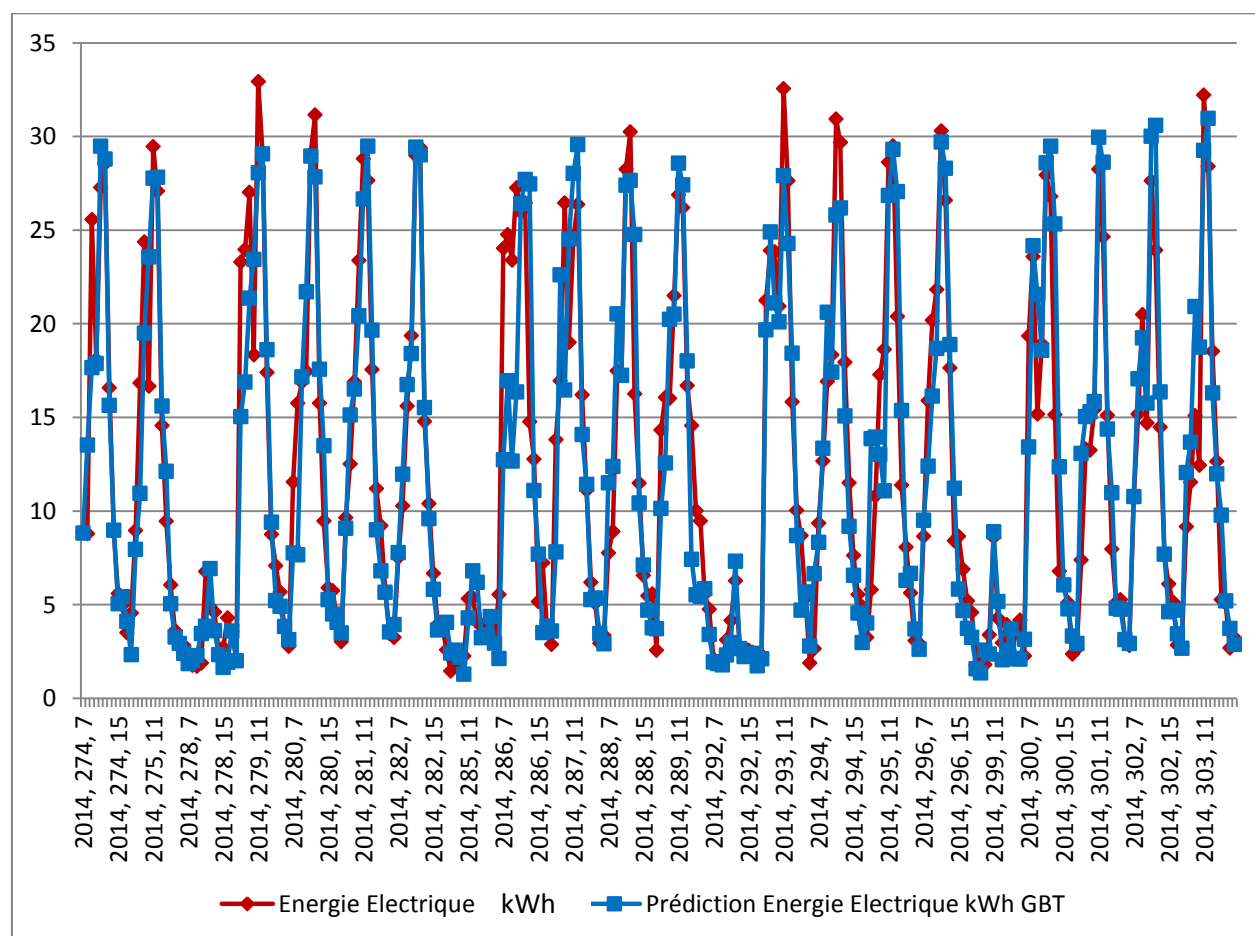


Figure 81 : Consommation électrique globale et la prédiction en sortie du gradient boosted tree en Kwh dans le secteur tertiaire, restauration (Année, jour de l'année, heure)

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	Déviations 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 classer 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 classer correctement, 8 d'entre elles l'ont été.

Modèles	Taux de classification	Déviations 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 :

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

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

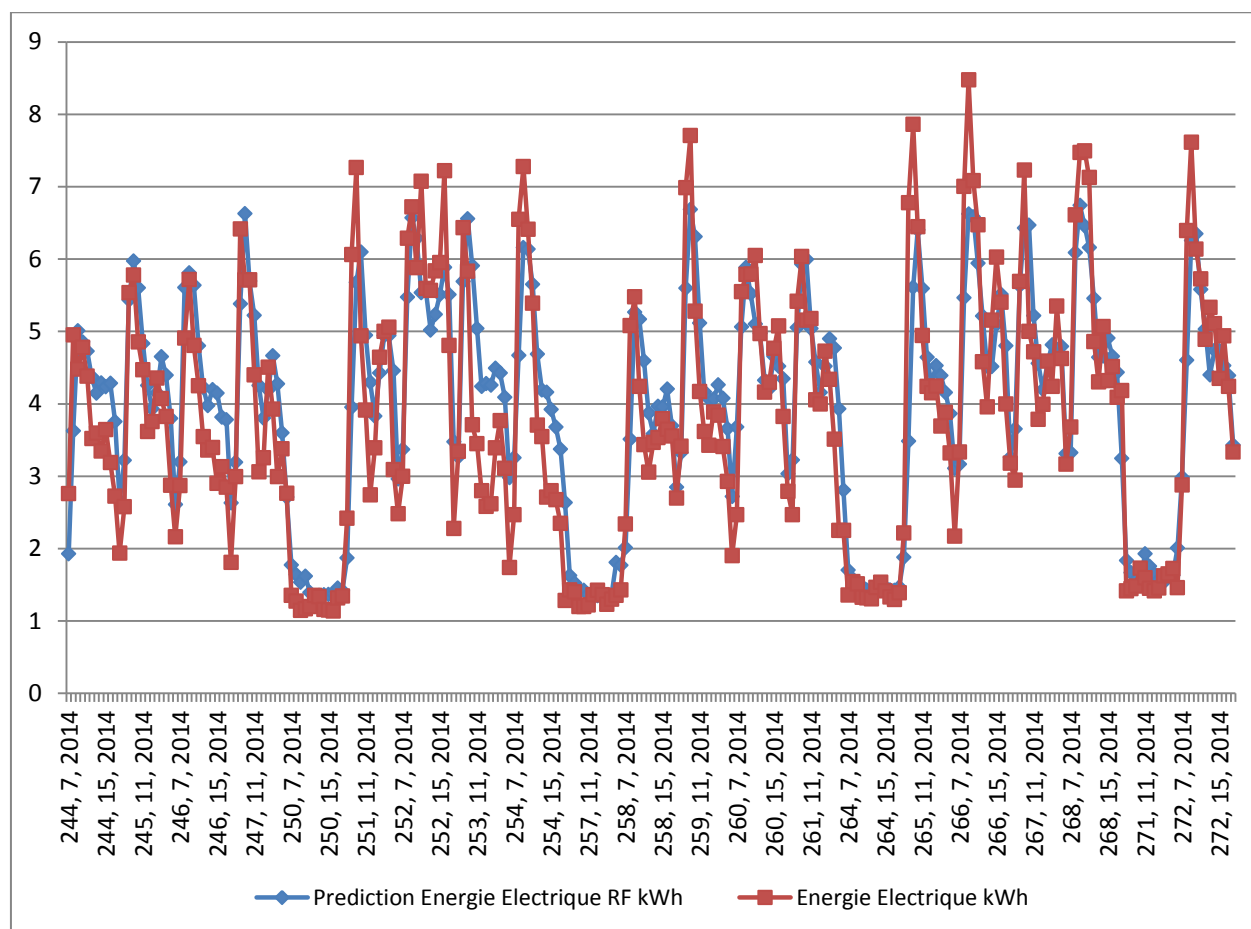


Figure 82 : Consommation électrique globale et la prédiction en sortie du Random Forest Tree en kWh dans le secteur tertiaire, secteur d'activité bureaux

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	Déviati on 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 :

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

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.

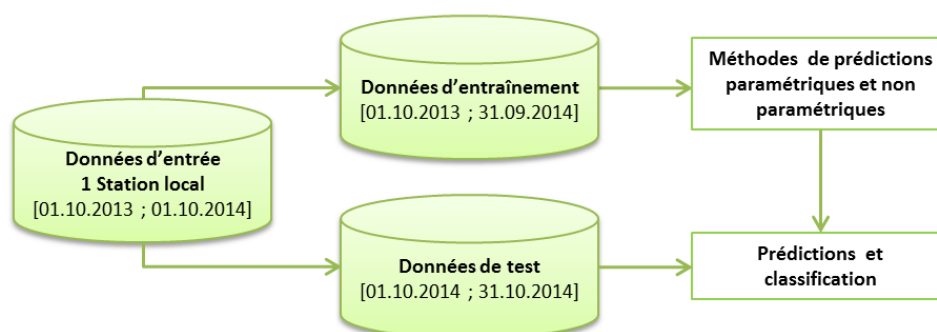


Figure 83 : Répartition des données pour la prédiction de la production décentralisée solaire

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

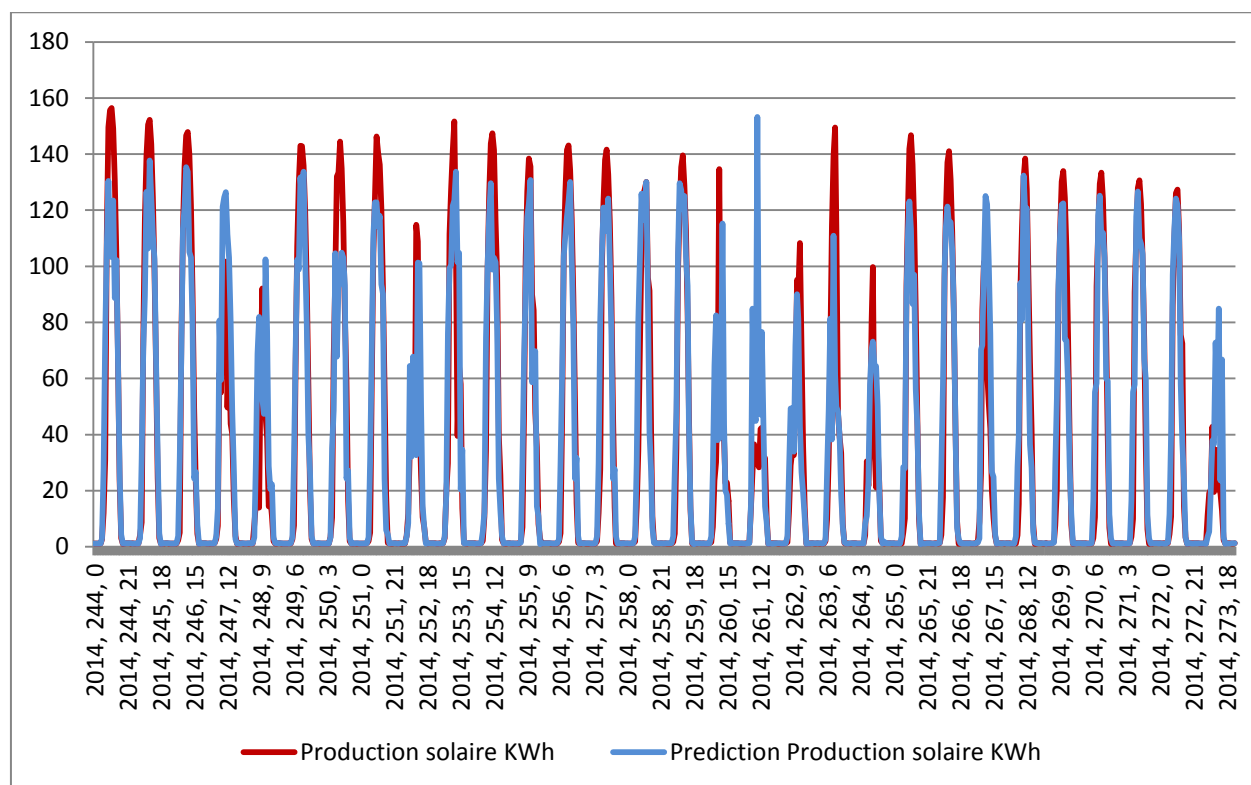


Figure 84 : Production solaire et la prédiction en sortie du GBT pour le mois de septembre avec la station de Sion distante de 15 km du site de production

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

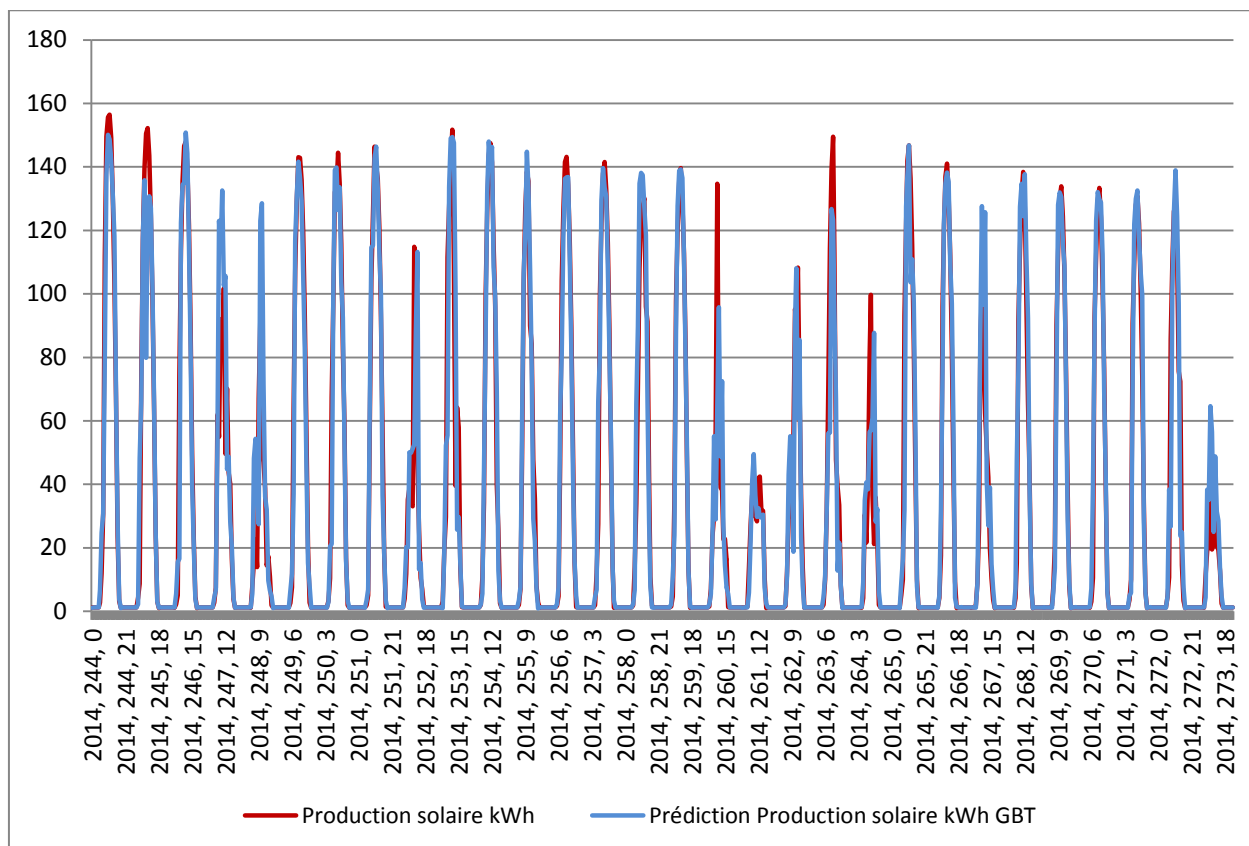


Figure 85 : Production solaire et la prédiction en sortie du gradient boosted tree pour le mois de septembre avec la station de Sion distante de 15 km du site de production

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.

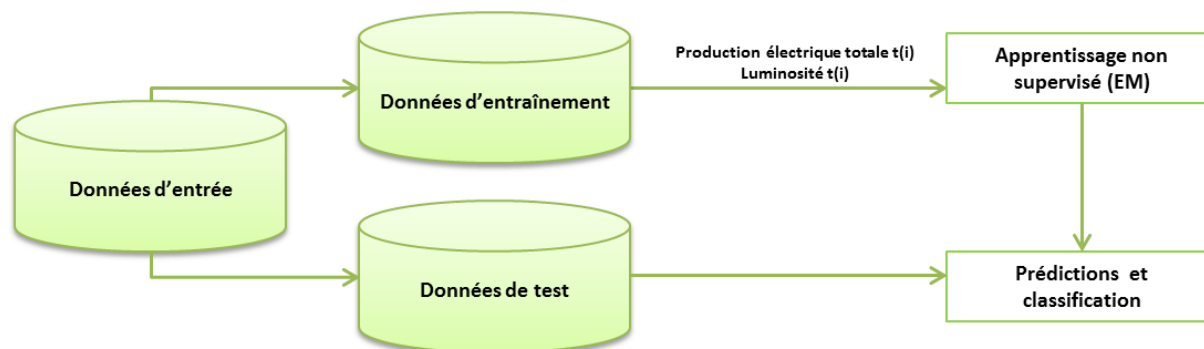


Figure 86 : Méthodes de classification des données pour la production décentralisée solaire

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	Déviation standard	Répartition %
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	Déviation 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 économe. Selon les coûts d'exploitation, le système HVAC est le système le plus économe, 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.

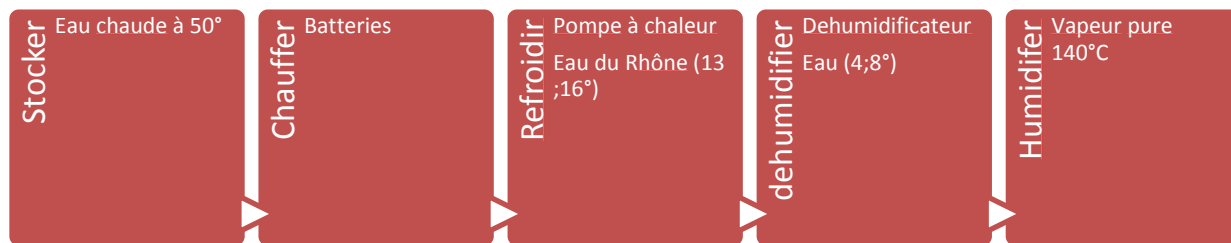


Figure 87 : Processus de régulation d'air sur le site industriel

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