

Définition des modèles de consommation énergétique d'un bâtiment et dimensionnement d'un système de stockage d'énergie

Ing. N. DE MEEÛS D'ARGENTEUIL – Ing. J. PELLETIER –
Dr. F. DE BRUYNE
ECAM – Bruxelles

Cet article a pour objectif de présenter les résultats d'analyses de données et la construction d'un algorithme de Machine Learning permettant la prédiction de la consommation électrique d'un bâtiment. L'objectif futur de cette recherche est d'optimiser, à l'aide de ces prédictions, l'autoconsommation d'une installation photovoltaïque équipée de systèmes de stockage.

Mots-clefs : algorithmes prévisionnels, machine learning, photovoltaïque, stockage énergétique, peak-shaving.

The main goal of this article is to present the results of data analysis and the construction of a Machine Learning algorithm allowing the prediction of the electricity consumption of a building. The future objective of this research is to optimize, using these predictions, the self-consumption of a photovoltaic installation equipped with storage systems.

Keywords : forecasting algorithm, machine learning, photovoltaic, energy storage, peak-shaving.

1. Introduction

Face à une population mondiale croissante, l'augmentation de consommation d'électricité et d'émissions de gaz à effet de serre vient quotidiennement défier le domaine de l'énergie. Avec l'expansion des énergies renouvelables et la décentralisation de la production, il devient impératif d'apprendre à utiliser l'énergie de manière plus efficace, plus fiable et avec moins de gaspillage pour contrer leur intermittence.

Cet objectif est l'un des premiers de l'entreprise GreenWatch, au sein de laquelle a été mené ce projet de fin d'études. L'entreprise, ayant installé des panneaux PV¹ au mois de février 2021, désire maximiser l'autoconsommation et l'autoproduction de son installation en s'équipant d'un système de stockage par batteries. L'intérêt futur de ce projet est de développer des stratégies de contrôle pour ces batteries afin de réduire les flux énergétiques vers le réseau et d'améliorer l'autoconsommation de l'électricité produite par l'installation PV.

L'implémentation de stratégies de contrôle nécessite l'utilisation de Machine Learning pour anticiper la production des panneaux PV ainsi que la consommation d'électricité. L'apport principal de ce travail de fin d'études réside dans l'analyse des données historiques de consommation du bâtiment de GreenWatch et dans la construction de modèles prédictifs pour la consommation.

Dans un premier temps, cet article détaille et illustre les termes d'autoconsommation, d'autoproduction et de "peak-shaving". Ces définitions permettent d'expliquer l'intérêt de la construction de modèles prédictifs pour la consommation et la production d'électricité. Dans un second temps, cet article détaille la construction de l'algorithme de Machine Learning selon un protocole en 4 étapes : analyse et filtrage de données, sélection des structures de modèle, phase de test (3 ensembles de données d'entraînement) et validation des modèles.

2. Optimisation d'une installation photovoltaïque

L'utilisation de batteries permet aux petits producteurs équipés de panneaux PV de prendre une indépendance partielle vis-à-vis du réseau électrique et d'augmenter la rentabilité de leur installation PV. En effet, le stockage se présente comme une bonne solution pour améliorer la sécurité et la stabilité du réseau électrique. Dans cette section, l'intérêt de l'optimisation d'un système de batteries via l'utilisation de prédic-

¹ PV : photovoltaïques

tion de consommation et production est expliqué via les notions d'autoconsommation, d'autoproduction et de "peak shaving". Ensuite, quelques informations supplémentaires sur la problématique et le contexte du projet sont apportées.

2.1. Autoconsommation et autoproduction

Dans le cas de l'énergie solaire, l'autoconsommation énergétique est un facteur utilisé pour définir la portion d'énergie consommée sur place par rapport à l'énergie totale produite par les panneaux photovoltaïques du bâtiment. Le taux d'autoconsommation est un paramètre que les installateurs de panneaux et de batteries solaires tentent souvent de maximiser. L'objectif est d'être capable de consommer un maximum de la production locale. L'autoproduction énergétique est définie comme la part de la consommation totale qui est fournie par l'électricité produite en interne. On parle également du taux d'autoproduction ou d'autosuffisance et il représente le niveau d'autonomie d'un bâtiment par rapport au réseau électrique.

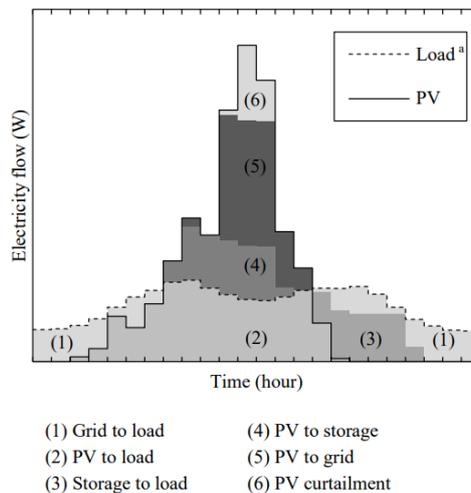


Figure 1 : Flux de production et de consommation sur une journée (illustration de l'autoconsommation et l'autoproduction)

Les deux notions expliquées ci-dessus peuvent également s'exprimer sous la forme d'une fraction en utilisant les différentes zones grisées sur la Figure 1. Celle-ci présente la répartition des flux de production et de consommation sur une journée.

Selon la Figure 1, il est possible d'exprimer les taux d'autoconsommation et d'autoproduction respectivement via les équations (1) et (2).

$$\text{Autoconsommation} = \frac{(2) + (3)}{(2) + (4) + (5) + (6)} \quad (1)$$

$$\text{Autoproduction} = \frac{(2) + (3)}{(1) + (2) + (3)} \quad (2)$$

Il est commun d'avoir recours à des algorithmes prévisionnels de consommation et de production. En effet, dans le but de maximiser les deux paramètres cités plus haut, la réalisation de prédiction peut permettre d'instaurer des stratégies de contrôle de batterie. À titre d'exemple, il est possible de maximiser l'autoconsommation en optimisant la charge de la batterie. Lorsqu'une situation de surproduction importante est prédite pour le jour suivant, l'optimisation consiste à vider au maximum la batterie la journée et pendant la nuit, pour être capable de stocker un maximum de ce qui est surproduit lors de la journée suivante. Sur la Figure 1, cela revient à agrandir l'aire de la zone (4), minimisant l'aire des zones (5) et (6), pour ensuite augmenter la zone (3).

2.2. Peak-shaving

Lorsqu'un bâtiment est équipé d'une installation PV, il rencontre par alternance des périodes de surproduction et de consommation nette. Dans le cas d'une surproduction, l'énergie des panneaux est injectée sur le réseau électrique. Alors que pour une période de surconsommation, l'énergie est prélevée depuis le réseau. Lors de cas extrêmes de surproduction ou surconsommation, cela engendre des pics d'import et d'export importants qui coûtent cher au prosumer (néologisme issu des mots producteur et consommateur). L'utilisation d'une batterie permet de limiter ces coûts par application d'un "peak shaving" ou d'un lissage des pics. Un exemple est illustré en Figure 2.

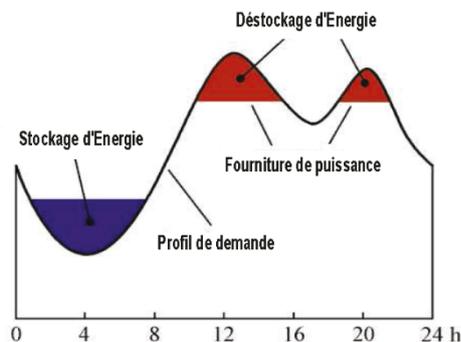


Figure 2 : Peak shaving ou lissage des pics

Pour limiter un export d'électricité trop important (situation de surproduction), il est possible de définir un seuil de surproduction à partir duquel l'excédent de production serait stocké dans la batterie. À l'inverse, dans une situation de surconsommation, l'utilisation de la batterie permettrait de limiter la quantité importée du réseau.

À nouveau, l'utilisation de modèles prédictifs se montre très utile pour réaliser ce genre d'optimisation. Par anticipation, il serait possible de garder la batterie vide pour être capable de stocker suffisamment pour limiter l'export au moment critique de surproduction. Charger la batterie au maximum permettrait également un lissage de la demande en période de surconsommation. L'insertion d'un système de batterie et d'une méthode de contrôle assurant une maximisation des taux d'autoproduction et d'autoconsommation ainsi que des anticipations pour un écrêtage sont les éléments permettant l'optimisation d'une installation photovoltaïque.

2.3. Problématique et contexte

L'entreprise GreenWatch, créée en 2009, est active dans trois secteurs principaux. Elle travaille dans les équipements pour le monitoring photovoltaïque et éolien, dans le contrôle énergétique (balisage, retail, etc.) et dans la recherche dans le domaine des énergies renouvelables. Elle s'est alliée sur le plan commercial avec l'entreprise Steel SA pour former Memoco. Lors du mois de février 2021, des panneaux photovoltaïques ont été installés sur le toit de l'entreprise Steel, située à Louvain-la-Neuve.

L'installation étant appliquée dans le cadre d'un bâtiment d'entreprise, les périodes de production auront lieu au même moment que les périodes de consommation. Malheureusement, cela n'est pas un cas optimal pour une installation de stockage d'énergie. Certains serveurs tournent également 24h/24, ce qui signifie que l'énergie stockée servira en grande partie à combler les besoins nocturnes. Enfin, le climat belge n'étant pas le plus adapté pour les panneaux photovoltaïques, l'utilisation de l'élément de stockage se montrera utile pour les mois d'été et malheureusement beaucoup moins pour l'hiver.

3. Création de modèles prédictifs

La construction de l'algorithme prévisionnel a été réalisée selon une démarche en 4 étapes. En premier lieu, les données brutes quart-horaire provenant de l'historique du compteur électrique du bâtiment ont été filtrées et nettoyées. Différents patterns de consommation ont été ciblés à travers plusieurs analyses. En parallèle, une brève analyse sur des estimations de données de production a également été effectuée. Ensuite, une sélection de modèles prédictifs a été réalisée et les entrées pertinentes de ces modèles ont été sélectionnées. En troisième lieu, une phase de test a été menée sur deux types de modèles en utilisant pour chacun d'eux trois ensembles de données

différents. Enfin, une procédure de validation a été appliquée afin de comparer les performances des modèles étudiés.

3.1. Analyses de données de consommation et de production

Dans cette section, des analyses théoriques approfondies permettent d'identifier des patterns utiles à la prédiction dans les données historiques de consommation et de production.

Analyses des données de consommation

La première analyse s'est portée sur l'évolution de la consommation au cours des années 2012-2019. Ceci a permis de démontrer la présence d'un pattern relativement régulier, marqué par certaines différences en fonction des saisons. Un regard plus approfondi sur les données de 2019 a été posé de sorte à en retirer davantage d'informations. Les données de cette année-là ont été triées par quart d'heure et séparées ensuite en deux ensembles pour distinguer les différences dues aux saisons.

Une analyse statistique a montré que l'écart-type moyen et maximal des données hivernales était nettement plus élevé que celui des données d'été. Ceci était d'autant plus marqué pour les heures de jour (6h00 – 18h00).

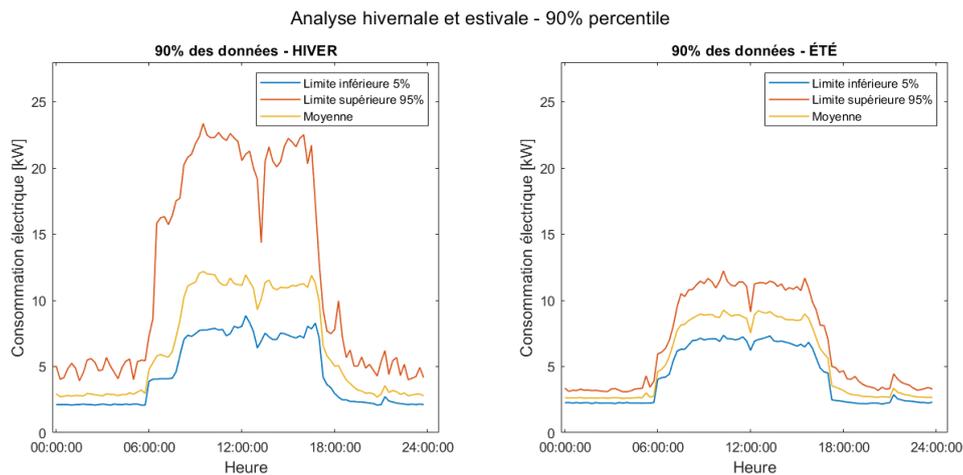


Figure 3 : Analyse percentile des données hivernales et estivales de 2019

La Figure 3 présente une analyse percentile 90 des données hivernales et estivales de 2019. Grâce à cette analyse, la présence d'une distribution non-normale a été détectée dans les données d'hiver.

Afin de comprendre l'origine de cette moyenne non-centrée, une analyse a été portée sur la différence entre les semaines 6 et 13 de 2019. Ceci a été illustré en Figure 4. On remarque que les deux profils sont très différents. En effet, pour la semaine 6, on observe plusieurs pics de consommation pour le lundi, mardi, mercredi et vendredi. Quant à la semaine 13, le pattern de consommation sur les 5 jours de la semaine est plutôt régulier et aucune charge importante ne semble avoir été mise en route.

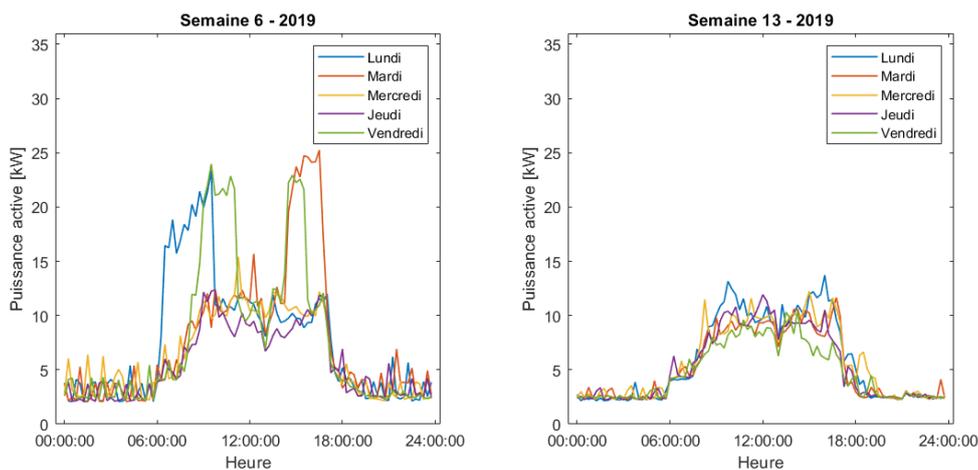


Figure 4 : Différence entre semaine 6 et semaine 13 - 2019

Pour confirmer l'hypothèse d'une charge lourde mise en route en hiver, un tableau listant les principales charges électriques du bâtiment a été dressé. La présence de deux chauffages électriques de 6/12 kW a expliqué et confirmé la corrélation entre l'allumage des chauffages et les pics observables lors des jours d'hiver. Différents patterns classiques ont également pu être classifiés (jour de semaine, congés, week-ends, etc.) et cela a permis de retirer davantage d'informations sur les données.

Analyses des données de production

Avant de s'intéresser aux systèmes de stockage, il a été nécessaire d'estimer la production de l'installation de panneaux photovoltaïques. Puisque ces derniers n'ont été que récemment installés, aucune donnée historique de production n'avaient été enregistrée. Pour le projet, GreenWatch s'est basée sur les données de production fournies par ELIA². C'est donc à partir de ces données que les premières analyses ont été établies. Dans cette section, plusieurs analyses sur les données historiques de production de 2019 seront présentées. Comme pour la consommation, l'objectif était de

² Données issues du site Internet du groupe Elia : <https://www.elia.be/en/grid-data/power-generation/solar-pv-power-generation-data>

pouvoir estimer la production journalière moyenne et définir un pattern moyen pour l'été et l'hiver.

L'objectif étant d'estimer la quantité d'énergie stockable dans la journée, une mise en graphique a été réalisée pour repérer un pattern de production moyen pour les

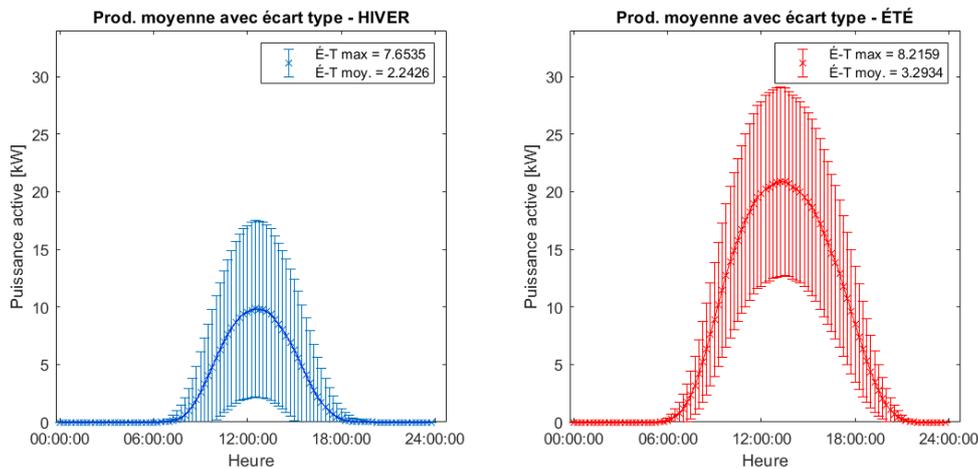


Figure 5 : Moyennes et écarts-types sur le quart d'heure - Production - HIVER ÉTÉ 2019

deux saisons. Comme pour les données de consommation, la moyenne et l'écart-type ont été calculés sur chaque quart d'heure de la journée (hiver/été). La Figure 5 montre le résultat obtenu.

Ce graphique permet de visualiser et de quantifier la nette différence de production entre les deux saisons. En effet, le pattern de production estival moyen est plus conséquent que le pattern hivernal. Le pic de production de la courbe moyenne estivale est à 20,8 kW contre 10 kW pour l'hiver. On remarque également que les écarts-types maximaux de production sont plus élevés que les écarts-types maximaux rencontrés dans les données de consommation. Cela signifie que les patterns sont assez volatiles, ce qui a du sens : il suffit d'un mauvais jour d'ensoleillement pour que la production journalière chute fortement et cela, indépendamment des saisons.

Afin de visualiser plus clairement les patterns de semaine et de les comparer aux patterns de consommation analysés précédemment, un graphique comparatif de la semaine 6 et 13 a été tracé. La Figure 6 montre le résultat. Il est à souligner que ces analyses sont réalisées dans le but de faciliter le dimensionnement d'un système de stockage. Or, si on compare la semaine 6 de la Figure 6 avec la semaine 6 de la Figure 4, on en conclut que ce système de stockage ne sera malheureusement pas/peu utilisable lors de la saison hivernale.

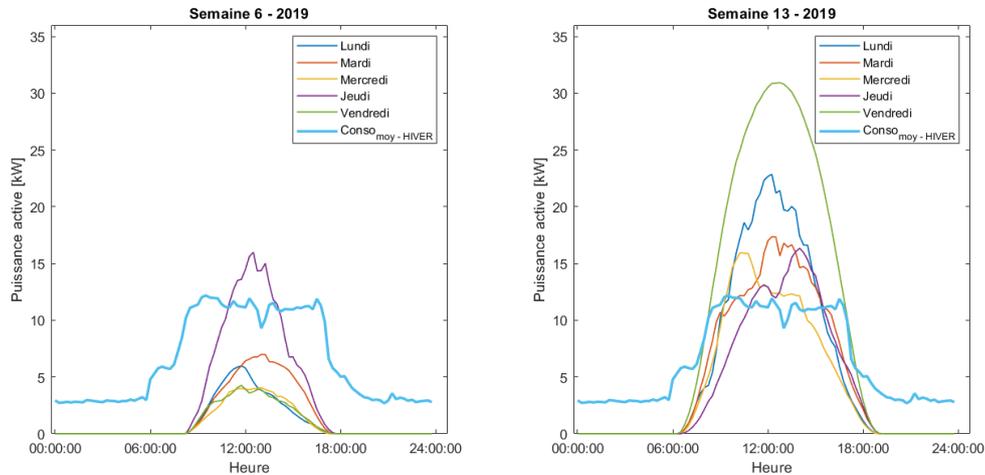


Figure 6 : Différence entre semaine 6 et semaine 13 - Production 2019

3.2. Choix des modèles théoriques

La création d'un algorithme prédictif se montrant utile pour l'optimisation future du système de stockage, il a fallu définir un horizon de prédiction. En première approche, un modèle à horizon de 24 h a été pensé. Les entrées statiques et dynamiques jugées pertinentes pour la prédiction de consommation sont listées ci-dessous :

Prédicteurs externes (statiques)

- Heure de la journée
- Jour de la semaine
- Week-end
- Jours fériés
- Mois de l'année
- Température extérieure

Prédicteurs internes (dynamiques)

- Consommation du jour précédent
- Consommation de la semaine précédente

Suite à la sélection de données, la structure des modèles devient connue. Il s'agit d'un système à 8 entrées, composé de p points. Sur base de cette structure ont été essayés deux types de modèle très distincts. Le premier modèle est une régression linéaire et le deuxième un réseau de neurones. Ce dernier est un DNN³ de type feed-forward⁴ avec 20 couches cachées. La méthode de Levenberg-Marquardt a été utilisée pour la minimisation de la fonction de coût et l'adaptation des valeurs des pondérations et des biais du modèle.

³ DNN : Deep Neural Network (Réseau de neurones profond)

⁴ Feedforward : propagation avant

Régression linéaire

Dans le cas où le modèle sélectionné est un modèle de régression linéaire, il est possible d'écrire le système sous la forme matricielle suivante :

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(p)} \end{bmatrix} = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_8^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_8^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(p)} & x_2^{(p)} & \dots & x_8^{(p)} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_9 \end{bmatrix} + \begin{bmatrix} \varepsilon^{(1)} \\ \varepsilon^{(2)} \\ \vdots \\ \varepsilon^{(p)} \end{bmatrix} \quad (3)$$

Une colonne de 1 a été ajoutée à la matrice d'entrée pour ajouter un biais. L'ajout d'un biais peut être considéré comme analogue au rôle d'une constante dans une fonction linéaire, dans laquelle une droite est translatée par la valeur du terme indépendant.

L'équation (3) peut également s'écrire sous la forme simplifiée suivante :

$$Y = X.W + \xi \quad (4)$$

Avec :

- Y : la matrice de sorties
- X : la matrice d'entrées
- W : le vecteur de paramètres
- ξ : la matrice d'erreur

L'objectif est d'être capable de déterminer le vecteur de paramètre optimal W_{opt} permettant de minimiser la fonction de coût suivante :

$$E(W) = \frac{1}{2} (Y - X.W)^T (Y - X.W) = \frac{1}{2} \xi^T \xi \quad (5)$$

Le vecteur de paramètre optimal W_{opt} s'obtient par voie analytique :

$$W_{opt} = (X^T X)^{-1} X^T Y \quad (6)$$

Réseau de neurones

Le modèle de réseau de neurones étant plus complexe, son fonctionnement n'est pas expliqué dans le cadre de ce travail⁵. Sa structure simplifiée est tout de même présentée par l'équation (7) ci-dessous.

$$\begin{bmatrix} a_0^{(1)} \\ a_1^{(1)} \\ \vdots \\ a_n^{(1)} \end{bmatrix} = \sigma \left(\begin{bmatrix} w_{0.0} & w_{0.1} & w_{0.2} & \cdots & w_{0.n} \\ w_{1.0} & w_{1.1} & w_{1.2} & \cdots & w_{1.n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{k.0} & w_{k.1} & w_{k.2} & \cdots & w_{k.n} \end{bmatrix} \begin{bmatrix} a_0^{(0)} \\ a_1^{(0)} \\ \vdots \\ a_n^{(0)} \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} \right) \quad (7)$$

Voici les variables rentrant en jeu :

- Les valeurs associées aux neurones de la première couche : $a_0^{(0)}, a_1^{(0)}, \dots, a_n^{(0)}$
- Les coefficients de « poids » : $w_{0.0}$
- Les facteurs de biais : b_0, b_1, \dots, b_n
- La fonction d'activation (sigmoïde) : σ

Le fonctionnement d'un réseau de neurones se base également sur la minimisation d'une fonction de coût. La méthode de minimisation de Levenberg-Marquardt a été utilisée dans le cadre de ce travail. Il s'agit d'un procédé fonctionnant de manière itérative visant à adapter les paramètres du réseaux de neurones dans le but de réduire l'écart entre les valeurs de prédictions et les réelles valeurs. Ce procédé d'adaptation se réalise sur base de données historiques d'entraînement et la validation du modèle s'étudie sur base de données de validation.

3.3. Résultats

Afin de tester l'efficacité d'un algorithme, les données historiques sont triées en deux ensembles distincts : un ensemble de données d'entraînement et un ensemble de données de tests. Cette technique de séparation de donnée se nomme « validation non-croisée ». L'ensemble de données d'entraînement est utilisé pour « entraîner » le modèle prédictif. Ensuite, l'ensemble des données de tests est utilisé pour valider la précision et la fiabilité du modèle. Cette section est composée de trois parties. Chacune d'entre elles présentent les résultats obtenus avec l'utilisation de set de données différents pour les deux modèles étudiés : régression linéaire et réseau de neurones.

⁵ Davantage de détails ont été apportés dans le travail de fin d'études [1].

Ensemble de données totales

Dans un premier temps, l'idée d'utiliser un maximum de données d'entraînement a été mise en application. Les données ayant été lissées ont été séparées manuellement de la façon suivante :

- Set d'entraînement : 2013 - 2018
- Set de test/validation : 2019

Les résultats obtenus pour le modèle de régression linéaire sont illustrés en Figure 7. La partie supérieure du graphique présente la superposition des données de validation avec les données de prédiction. Un pourcentage de 41,97 % d'ajustement est mentionné en légende de ce graphique.

Régression linéaire (set d'entrainement 2013-2018)

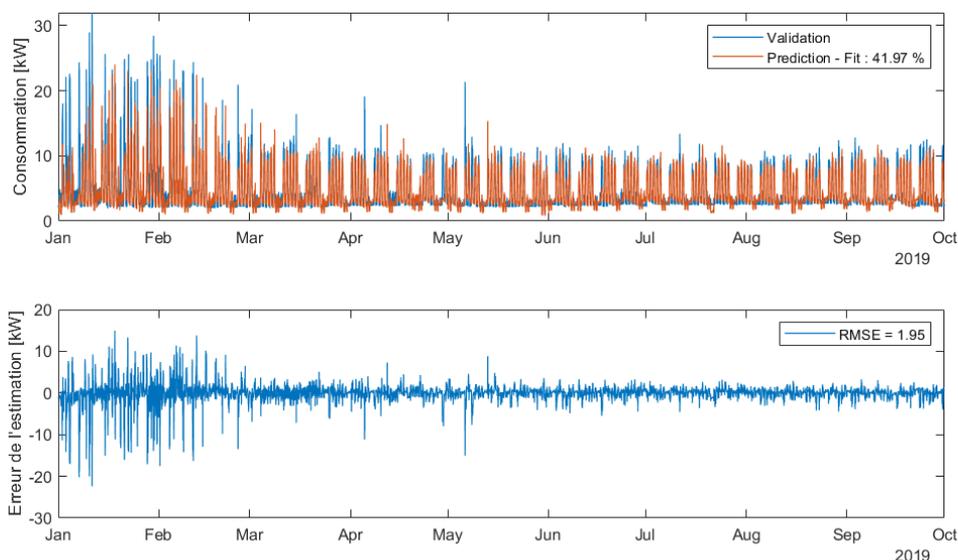


Figure 7 : Régression linéaire - résultats obtenus pour les données de 2019

Le graphique du bas de la Figure 7 met en évidence l'erreur d'estimation en kW, également appelée "résidu". Le RMSE est de 1,950. Le résidu correspond à la différence entre les prévisions et les données de consommation historiques réelles de 2019. On observe premièrement que les erreurs les plus importantes se trouvent en janvier et février, les mois où le chauffage électrique est mis en route. Puisque la prédiction sous-estime la consommation à ces moments particuliers, l'erreur de l'estimation est majoritairement négative. Il est évident que l'allumage de cette charge

n'est pas pris en compte dans les prédicteurs et qu'il est normal qu'il ne soit pas prévisible.

Une analyse d'autocorrélation des résidus a été effectuée pour évaluer la performance du modèle. Elle est présentée en Figure 8. Les résultats de cette analyse se montrent assez bons et permettent de valider partiellement le modèle. Il y a en effet peu de corrélation entre les résidus, mis à part pour le délai de 7 et de 14 jours. Cela signifie que le modèle réalise une erreur systématique toutes les 24 heures.

Pour cet ensemble de données, les résultats obtenus via réseau de neurones étaient légèrement supérieurs. Le pourcentage d'ajustement était de 52,32 % et le RMSE était de 1,602.

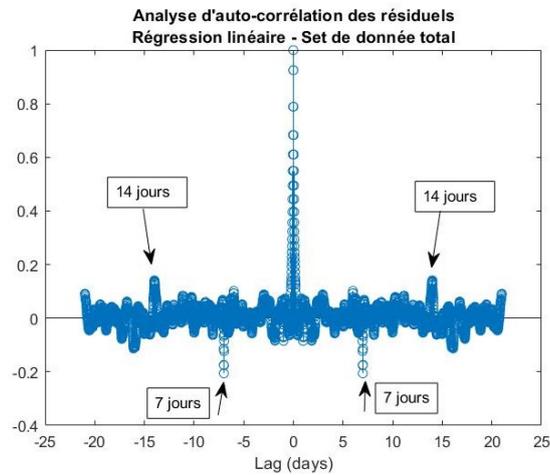


Figure 8 : Autocorrélation des résidus (Régression linéaire – total)

La configuration n'étant pas optimale et les résultats n'étant pas concluant, il a été décidé de se concentrer uniquement sur les données estivales. En effet, suite à une analyse des estimations de données de production photovoltaïque, il a été jugé moins utile d'obtenir de bons résultats de prédiction pour l'hiver. La production PV en Belgique étant faible à cette période-là, il est inutile de tenter d'optimiser l'utilisation de la batterie puisque la production des panneaux PV dépasse rarement la consommation. En hiver, l'électricité produite est donc directement consommée depuis les panneaux, plutôt que d'être stockée.

Ensemble de données estivales

Les pics de consommation apparaissant en hiver n'étant pas évident à prédire, il est jugé plus intéressant de changer de set de données d'entraînement et d'utiliser des

données d'été pour identifier un modèle de prédiction plus robuste. Les données suivantes ont été utilisées :

- Set d'entraînement : 1 juin au 1 octobre 2017
- Set de test/validation : 1 juin au 1 octobre 2019

Les résultats obtenus avec ces données par le modèle de régression linéaire sont présentés en Figure 9. Le pourcentage d'ajustement est de 65,81 % et le RMSE est de 0,887. On observe qu'à nouveau, le modèle semble assez performant puisque les patterns de prédiction sont semblables aux patterns des données de validation. On observe cependant une forte déviation lors du dimanche 21 juillet 2019 (cfr Figure 9). En effet, ce jour étant particulier (jour férié et weekend), le modèle réalise une forte erreur à cet endroit-là. C'est assez cohérent puisqu'aucune entrée distincte n'a été ajoutée aux modèles pour spécifier des cas particuliers de la sorte. De plus, ceci n'a pas été rencontré dans les données d'entraînement. Une amélioration du modèle avec une méthode de sélection d'entrées pertinentes fera l'objet d'un futur travail.

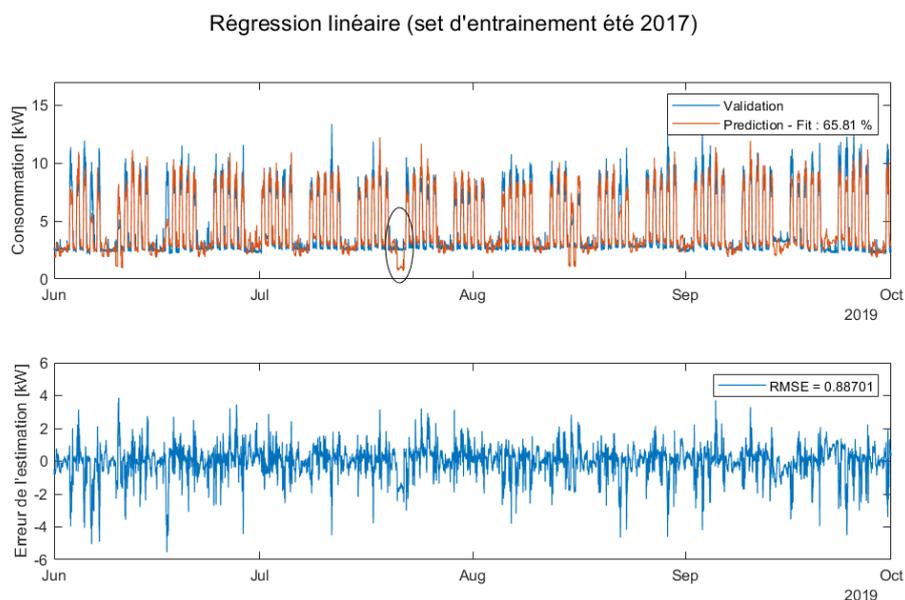


Figure 9 : Résultats régression linéaire - données estivales

Le graphique d'autocorrélation présenté en Figure 10 confirme que le modèle est plutôt performant. On a un pic de corrélation pour un "lag" de zéro et les autres valeurs sont assez petites pour les autres "lags". Il s'agit d'un corrélogramme se rapprochant d'un corrélogramme de bruit blanc.

Les essais identiques ont été effectués avec réseau de neurones et les mêmes commentaires en ressortent. Les résultats sont illustrés en Figure 11 et Figure 12. Le corrélogramme confirme que les résultats sont légèrement meilleurs. Le pourcentage d'ajustement est de 68,64 % et le RMSE est de 0,814.

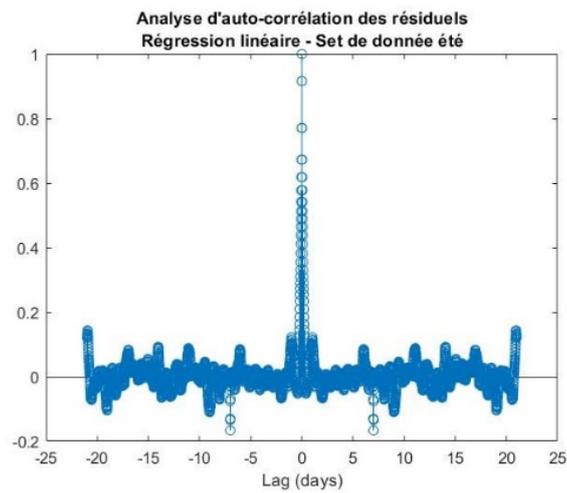


Figure 10 : Autocorrélation des résidus (Régression linéaire – été)

Réseau de neurones (set d'entrainement été 2017)

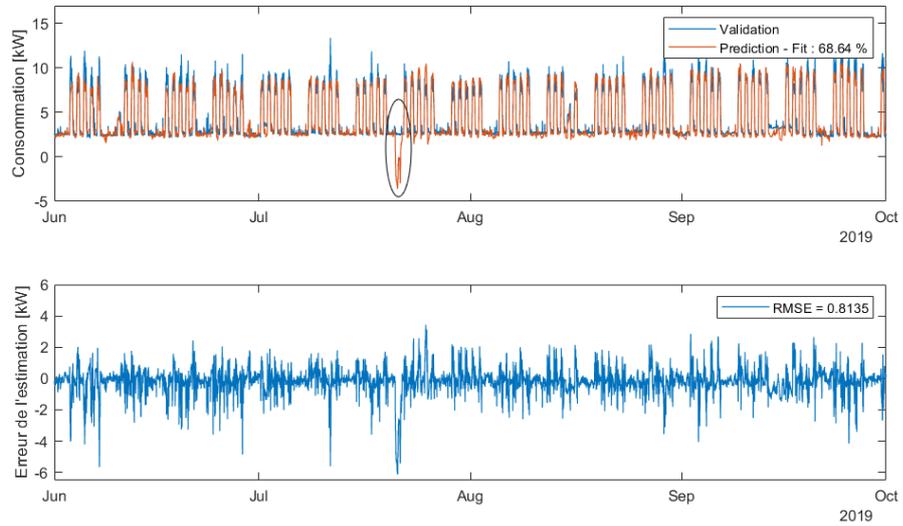


Figure 11 : Résultats réseau de neurones - données estivales

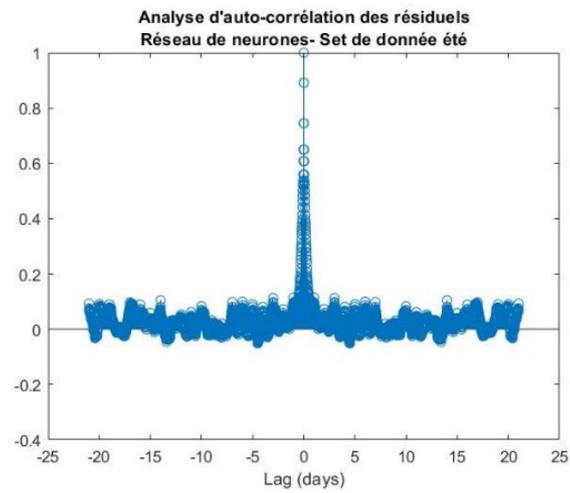


Figure 12 : Autocorrélation des résidus (Réseau de neurones - été)

Concaténation d'ensemble de données estivales

De sorte à résoudre ce problème rencontré pour les jours exceptionnels, une nouvelle sélection de donnée a été réalisée. Pour avoir plus d'information, il a été décidé de mettre une série d'ensemble de données d'été l'une à la suite des autres. Les entrées des modèles d'identification ont donc été mises à la chaîne. Cette concaténation permet d'avoir un ensemble de donnée d'entraînement plus conséquent et contient dès lors plus d'information.

Pour cette configuration, les résultats obtenus par régression linéaire n'ayant pas apportés d'amélioration significative, seuls les résultats par réseau de neurones sont illustrés. La Figure 13 présente la comparaison entre les données réelles de validation et les données de prédiction obtenues via le réseau de neurones. Il s'agit ici de la configuration présentant les meilleurs résultats. Ce modèle a été confirmé via le calcul de l'erreur quadratique moyenne ($RMSE = 0,648$), le pourcentage d'ajustement ($Fit = 75,02\%$) ainsi que par une analyse d'autocorrélation des résiduels en Figure 14.

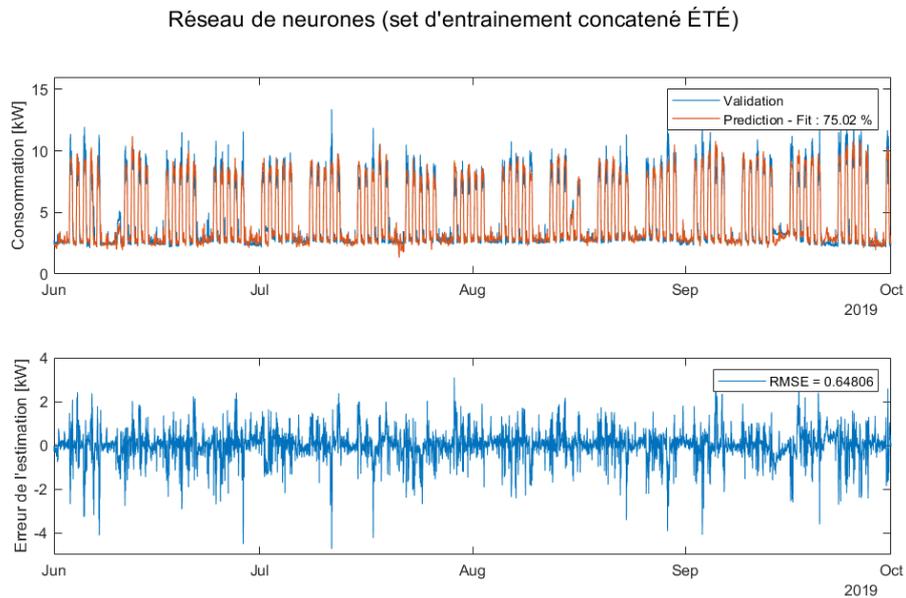


Figure 13 : Résultat réseau de neurones avec ensembles de données concaténés (ÉTÉ)

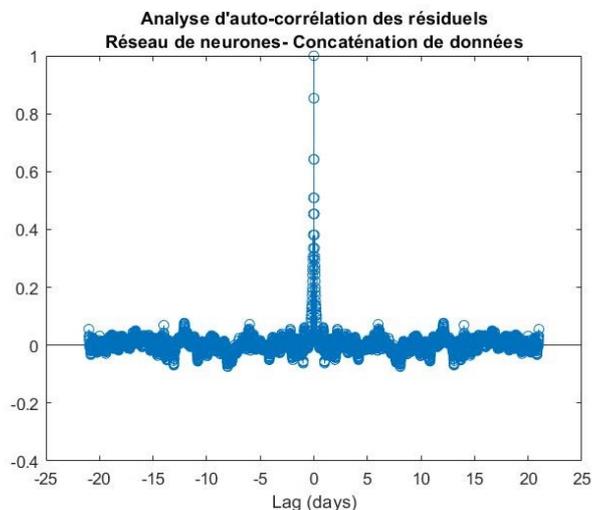


Figure 14 : Autocorrélation des résidus (Réseau de neurones - concaténation)

Ce modèle de réseau de neurones devra être utilisé en connaissance de cause, car certaines situations non-occasionnelles, telle que la présence d'un jour de weekend également férié, viennent parfois fausser les performances des prévisions.

4. Conclusion

Ce travail de fin d'études avait pour premier objectif de définir des modèles de consommation et production énergétique en se basant sur des données historiques. Cela a permis d'obtenir des patterns moyens saisonniers, de comprendre les modes de consommation de l'entreprise et de préparer la construction d'un modèle prédictif pour un horizon de prédiction de 24 h.

Les essais réalisés sur les modèles de prédiction ont été concluants. L'utilisation de différents ensembles de données a permis de comprendre l'importance d'utiliser des données suffisamment informatives pour obtenir une bonne prédiction. Le réseau de neurones (modèle non-linéaire) a permis d'obtenir de meilleurs résultats que le modèle de régression linéaire. Grâce aux non-linéarités et à la complexité de sa construction, le réseau de neurone est capable d'approximer les zones particulières des données de consommation. Certaines analyses supplémentaires auraient pu être réalisées pour la validation du modèle. En effet, le réseau de neurones fonctionnant de manière itérative pour minimiser la fonction de coût, est sujet au surapprentissage. Il aurait été intéressant de mettre en parallèle l'évolution de la fonction de coût

d'identification et de validation au cours des itérations. Ces modèles seront utilisés lorsque les stratégies de contrôle de batteries seront mises en place.

Dans le cadre de cette étude, les modèles ont été essayés pour un horizon de prédiction de 24 h et les prédictions finales se sont concentrées sur les périodes estivales. Dans une future étude, il serait également important de tester ces modèles sur des données hivernales et de réaliser quelques modifications pour éventuellement obtenir une prédiction à plus long terme. Une technique de sélection d'entrées pertinentes pourrait également être utilisée.

5. Source

- [1] DE MEEÛS D'ARGENTEUIL, N., *Définition des modèles de consommation énergétique d'un bâtiment et dimensionnement d'un système de stockage d'énergie*, Mémoire de master, Bruxelles, Belgique : ECAM, juin 2021.