Innovative Power Solutions [13]

Socomec, Usine 3
Entreprise d'accueil : 11 route de Strasbourg
67230 Huttenheim

Stage 6 mois

# Analyse de données

**Consommation des systèmes auxiliaires**

*Encadrant :*
M.VIUDES Raphaël

EHRET Justine

*Superviseur Académique :*
M.PRUD'HOMME
Christophe

UFR de mathématique et d'informatique
Université de Strasbourg [12]

Master 2, Calcul Scientifique et Mathématiques de l'Innovation
Etablissement : Université de Strasbourg

Août, 2025

# Remerciements

Avant toute chose, je tiens à remercier chaleureusement **Raphaël VIUDES**, qui m'a accompagné tout au long de ce stage avec bienveillance et rigueur. Sa pédagogie, sa disponibilité et sa capacité à transmettre son savoir ont grandement facilité mon intégration et ma progression. Grâce à son encadrement, j'ai pu évoluer dans un environnement stimulant, propice à l'apprentissage et à la prise d'initiative.

Merci à **Eric PLUMERE** pour ses conseils précieux sur la récupération et l'interprétation des données, qui m'ont permis de mieux comprendre le système dans son ensemble.

Un grand merci à **Jason THORRIGNAC** et **Lisa GUGNOT**, qui m'ont apporté un soutien précieux en électronique de puissance dès les premiers jours du stage. Leurs explications sur les notions de base m'ont permis de démarrer sur des bases solides.

Je suis également reconnaissante à **Arthur GLOTIN** pour ses explications sur le fonctionnement mécanique du système. Sa disponibilité, notamment lors de la visite à Nordhouse, m'a permis de mieux visualiser les équipements sur le terrain, au-delà des simples données à traiter.

Merci à **Louis PARIS** pour sa patience et sa réactivité face aux problèmes que je lui remontais concernant les données. Ses explications complémentaires sur le fonctionnement global du système m'ont été d'une grande aide tout au long du stage.

Un grand merci aussi aux autres stagiaires - **Valentine, Emma, Louis, Baptiste et Mathieu** - pour tous les bons moments passés ensemble pendant le stage. Partager cette expérience avec vous a vraiment rendu le quotidien plus sympa.

# Table des matières

## Résumé

Ma mission de stage consiste à caractériser l'influence de différents paramètres sur la consommation des systèmes auxiliaires dans les systèmes de stockage d'énergie. En tant que data analyst/scientist, je travaille principalement sur l'analyse de données issues de ces systèmes afin d'identifier les facteurs qui impactent leur consommation énergétique. Cela implique de manipuler des jeux de données complexes, de développer des modèles d'analyse, et de produire des visualisations pour interpréter les résultats.

L'objectif est de mieux comprendre le comportement des systèmes auxiliaires et d'optimiser leur fonctionnement. C'est une mission à la fois technique et analytique, qui me permet de mettre en pratique mes compétences en mathématiques, en informatique et en traitement de données, tout en contribuant à des enjeux concrets liés à l'efficacité énergétique.



## Executive Summary

My internship's core mission is to characterize the influence of various parameters on the auxiliary systems' power consumption within our energy storage solutions. As a Data Analyst/Scientist, my primary work revolves around analyzing data from these systems to pinpoint the factors that impact their energy use. This involves manipulating complex datasets, developing analytical models, and producing data visualizations to interpret my findings.

The ultimate goal is to gain a deeper understanding of the auxiliary systems' behavior and to optimize their performance. This is a mission that is both technical and analytical, allowing me to apply my skills in mathematics, computer science, and data processing while contributing to concrete challenges in energy efficiency.

# 1 Introduction

## 1.1 Présentation de *Socomec*

*Socomec* est une entreprise industrielle française fondée en 1922 par Joseph SIAT à Benfeld, en Alsace. À l'origine spécialisée dans les équipements électromagnétiques (sonnettes, coupe-circuits, interrupteurs), elle a progressivement évolué vers des solutions de gestion, de conversion et de stockage de l'énergie électrique basse tension. Aujourd'hui, *Socomec* est un groupe international indépendant, présent dans plus de 80 pays, avec 12 sites de production et plus de 4400 collaborateurs. En 2024, l'entreprise a réalisé un chiffre d'affaires de 924 millions d'euros, dont 8% du chiffre d'affaires a été réinvesti en recherche et développement.

FIGURE 1 – Joseph SIAT, fondateur de *Socomec* [10]

Depuis sa création, *Socomec* est restée une entreprise familiale, aujourd'hui dirigée par Ivan STEYERT, arrière-petit fils du fondateur. Elle se distingue par une forte culture d'innovation, un engagement pour la transition énergétique, et une attention particulière portée à la qualité de ses produits et au bien-être de ses collaborateurs.

FIGURE 2 – Ivan STEYERT, actuel PDG de *Socomec* [8]

**Activités et organisation**

*Socomec* structure ses activités autour de cinq pôles d'expertise :

- **La coupure de l'arc électrique :** Pour contrôler l'énergie et protéger les personnes et les biens. Interrupteurs-sectionneurs, inverseurs de sources, fusibles.
- **La mesure de l'énergie :** Pour améliorer la performance énergétique et la surveillance des installations. Compteurs, capteurs, analyseurs de réseau.
- **La conversion d'énergie :** Pour assurer la disponibilité et la continuité d'une énergie de haute qualité. Onduleurs, convertisseurs de stockage, redresseurs industriels.
- **Systèmes de stockage d'énergie :** Pour décorréler la production de l'énergie de sa consommation.
- **Les services experts :** Audit, conseil et maintenance pour garantir une énergie disponible sûre et efficace.

Ces solutions sont conçues pour répondre aux exigences des secteurs critiques tels que les centres de données, la santé, l'industrie, les infrastructures, les bâtiments tertiaires et les énergies renouvelables.

FIGURE 3 – Usines *Socomec* : Siège social et Usine 2 (Benfeld, FRANCE)[22, 23]

FIGURE 4 – Usine 3, *Socomec* (Huttenheim, FRANCE) [24]

## 1.2 La Business Unit ESS (Energy Storage Solutions)

Face à la montée en puissance des énergies renouvelables et aux enjeux de stabilité des réseaux, *Socomec* a crée en 2017 la **BU ESS**, dédiée aux solutions de stockage d'énergie. Cette unité développe des systèmes modulaires comme le **SUNSYS HES-L** et **HES-XXL**, capables de stocker de 150 kWh à plus de 1,5 MWh d'énergie.


FIGURE 5 – **SUNSYS HEL-L** *Socomec*

Les systèmes ESS permettent :
— L'optimisation et l'autoconsommation.
— La gestion des pointes de consommation.
— Le soutien au réseau (grid support).
— L'alimentation de secours (off-grid).
— L'intégration de la mobilité électrique.

### Organisation du pôle ESS

Le pôle ESS est structuré autour de plusieurs sous-divisions, comme illustré dans l'organigramme suivant :

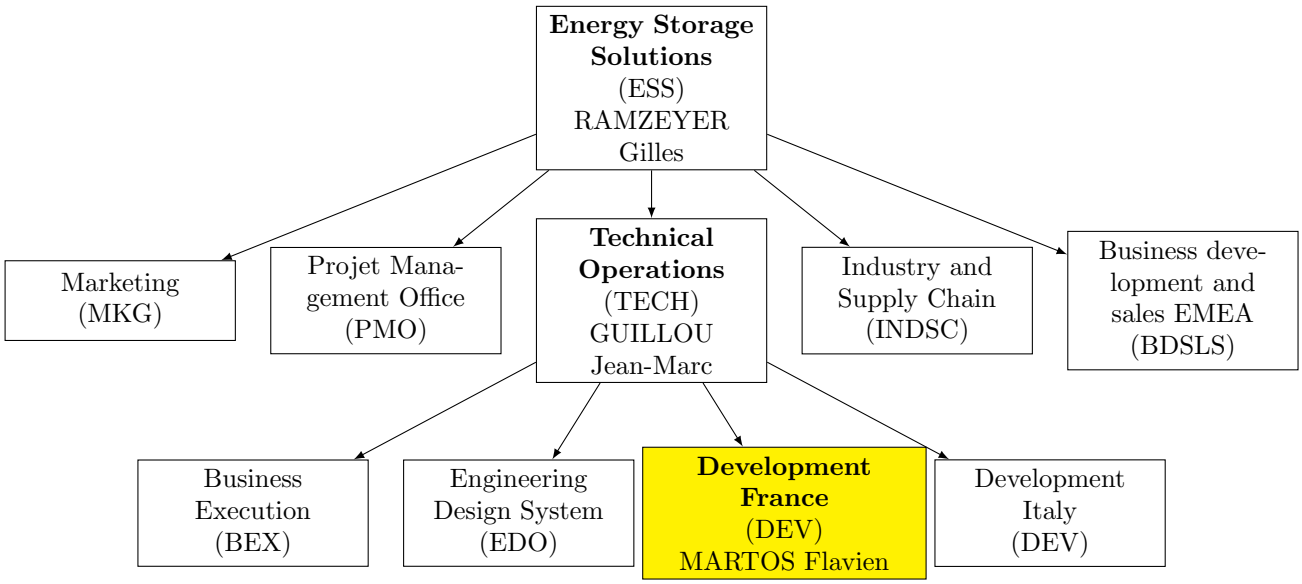


FIGURE 6 – Organigramme du pôle Energy Storage Solutions (ESS)

Durant mon stage, j'ai été intégré au sein de l'équipe **ESS TECH DEV**, plus précisément dans la branche française du développement technique. Cette équipe est responsable de la conception, du prototypage, des tests et de l'industrialisation des solutions de stockage.
L'organisation hiérarchique de mon intégration est représentée ci-dessous :

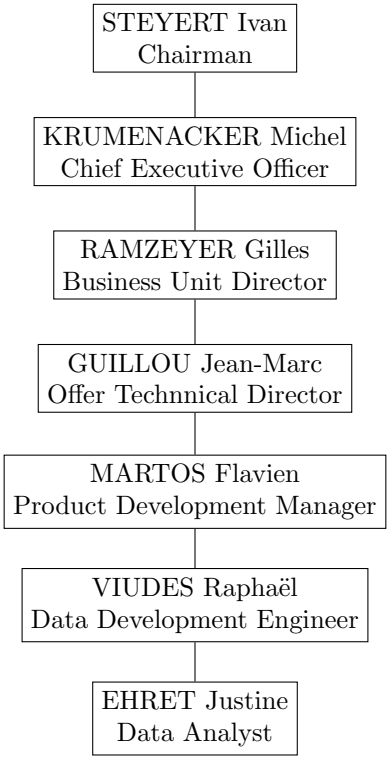


FIGURE 7 – Organigramme hiérarchique de mon intégration

## 1.3 Objectifs du stage

Dans les systèmes de stockage d'énergie, comme les batteries utilisées dans des environnements professionnels (écoles, hôtels, bâtiments tertiaires, etc.), la gestion de la consommation des auxiliaires est un levier essentiel pour améliorer l'efficacité énergétique globale. Ces auxiliaires regroupent notamment les systèmes de gestion de la batterie (BMS), les dispositifs de refroidissement, les capteurs, ainsi que d'autres composants nécessaires au bon fonctionnement et à la sécurité du système.

L'objectif principal de ce stage est **d'identifier et d'analyser les paramètres influençant la consommation énergétique des systèmes auxiliaires**, en particulier ceux liés à la régulation thermique des batteries. Cette démarche vise à proposer des pistes concrètes d'optimisation énergétique.

Pour atteindre cet objectif global, plusieurs étapes ont été définies :

— Se familiariser avec l'environnement de travail, les outils de développement, les plateformes de supervision et les bases de données disponibles.
— Comprendre l'architecture des systèmes de stockage et le rôle des auxiliaires dans leur fonctionnement.
— Collecter, structurer et analyser les données de consommation des auxi-

liaires.
- Identifier les facteurs influents sur la consommation énergétique (température ambiante, cycles de charge/décharge, configuration du système, etc.).
- Proposer des indicateurs pertinents permettant de quantifier et hiérarchiser l'impact de chaque paramètre.
- Mettre en évidence les leviers d'amélioration de l'efficacité énergétique des auxiliaires.
- Élaborer des recommandations ou des pistes d'optimisation, pouvant aller jusqu'à des propositions d'évolution matérielle ou logicielle.
- Présenter les résultats de manière claire, synthétique et exploitable par l'équipe technique.

Ce stage s'inscrit donc dans une démarche à la fois analytique et opérationnelle, avec pour finalité de contribuer à la performance énergétique des systèmes de stockage d'énergie.

**Diagramme de Gantt**

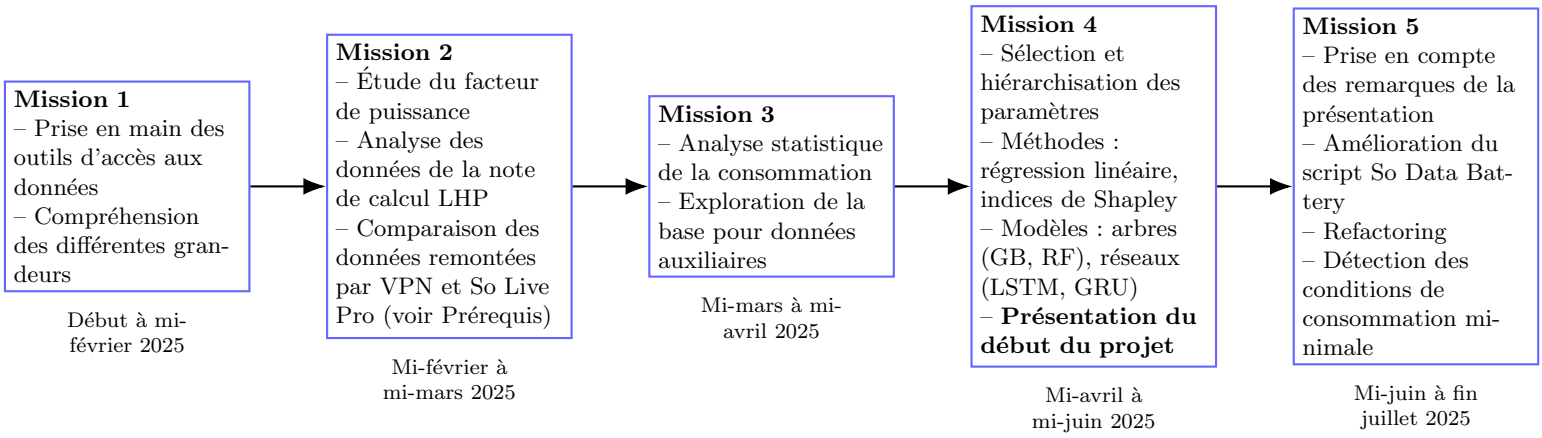


FIGURE 8 – Chronologie des missions du projet

# 2 Prérequis

## 2.1 Présentation des produits HES L et HES XXL

Les solutions HES L et HES XXL de Socomec sont des systèmes de gestion d'énergie modulaires destinés à la supervision et à l'optimisation des installations électriques. Leur architecture repose sur une structure modulaire, permettant d'adapter le système à différents types de bâtiments ou configurations techniques, tout en offrant une flexibilité pour la collecte et l'analyse des données.


FIGURE 9 – **SUNSYS HEL-L** *Socomec*. De gauche à droite : 2 armoires batterie (B-CAB) et 1 armoire de convertiseur AC-DC (C-CAB).

Les batteries ne stockent que du courant continu (DC), tandis que le réseau et les équipements utilisent du courant alternatif (AC). L'armoire de conversion (C-CAB) transforme l'AC en DC pour le stockage, et le DC en AC pour la restitution, permettant de gérer le flux d'énergie entre le réseau et les batteries.

## 2.2 Composants principaux

### 2.2.1 Les digiwares

Les digiwares constituent le cœur du système, centralisant les mesures et les informations relatives à la consommation énergétique. Ils assurent la communication entre les différents modules et garantissent la cohérence des données collectées.

### 2.2.2 Les passerelles et Solive Pro

Les passerelles permettent d'interfacer les modules HES avec le logiciel Solive Pro, qui offre un outil complet de supervision et d'analyse. Solive Pro rend possible le suivi en temps réel et l'export de données pour des traitements statistiques et des analyses approfondies.

### 2.2.3 Connexion à distance et collecte de données

Grâce à la connectivité IoT, ces systèmes peuvent être consultés à distance, ce qui facilite la collecte de données via un VPN sécurisé. Cette fonctionnalité permet de suivre les performances et d'exploiter les informations des systèmes HES L et HES XXL sans intervention physique sur site, garantissant à la fois sécurité et efficacité dans la supervision énergétique.

### 2.2.4 Architecture du système : PMS et M70

Le système HES repose sur une architecture hiérarchisée qui permet de superviser et de gérer efficacement les batteries et systèmes auxiliaires. Au niveau le plus fin, le PMS (Power Management System) surveille et contrôle cellule de batterie, collectant des informations essentielles telles que la tension, le courant, la température ou le nombre de cycles. Chaque PMS possède un numéro de série unique, ce qui permet de l'identifier dans la base de données. Les données issues des PMS sont ensuite centralisées par le M70, un module de supervision qui agrège les informations de plusieurs PMS et les transmet à la plateforme de suivi, comme SoLive Pro, ou via un accès VPN. La correspondance entre le PMS et le M70 est assurée grâce à la gateway, qui permet de relier les numéros de série et d'assurer la cohérence des données collectées. Cette architecture, complétée par la connectivité IoT, offre la possibilité de superviser à distance les systèmes, de récupérer des données détaillées sur le nombre de racks ou la puissance disponible, et de réaliser des analyses et optimisations à partir des informations consolidées.

## 2.3 Auxiliaire

Les systèmes auxiliaires sont composés de plusieurs éléments essentiels :

- Dans la B-CAB (Battery CABinet) :
  - Control Box
  - Chiller : il est utilisé pour refroidir l'air ou les liquides dans le système, il est lui-même composé de :
    - Pompes : assurent la circulation des fluides nécessaires au fonctionnement du système.
    - Ventilateur : permet de maintenir une circulation d'air adéquate pour le refroidissement et la ventilation.
    - Chauffage : responsable de la production de chaleur pour maintenir une température optimale.
- Dans le C-CAB (Convertissor Cabinet) :
  - Ventilateur : comme dans la B CAB, il assure une bonne circulation de l'air.
  - Électroniques
  - Chauffage : également présent pour fournir de la chaleur si nécessaire.

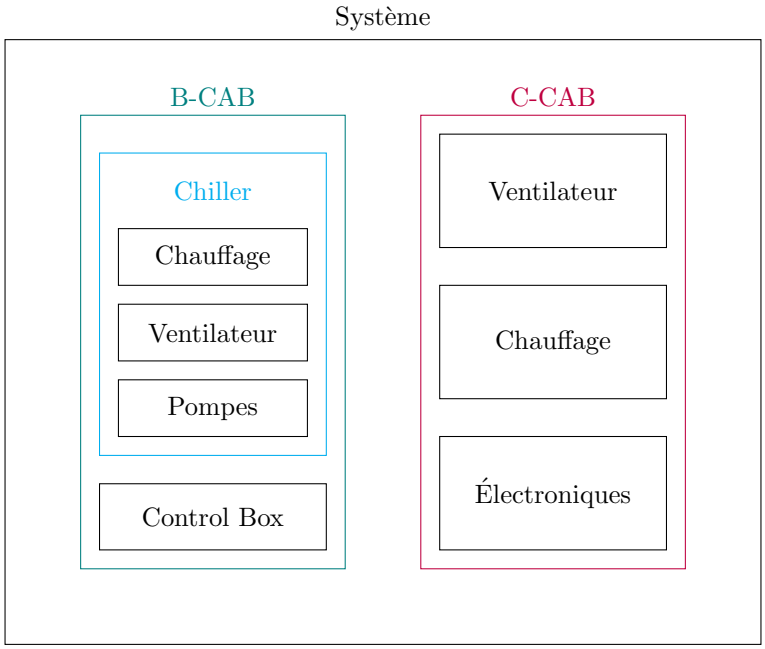


FIGURE 10 – Schéma d'un système avec les différents auxiliaires

Au minimum les auxiliaires consomment 100 W/B CAB et au maximum 2290 W/B CAB + 2850 W/C CAB (voir Mesures sur plateforme).

# 3 Analyse des puissances des auxiliaires, SoLive Pro

## 3.1 Présentation des données

### 3.1.1 Puissance active, réactive et apparente

- **Puissance active ($P$) :** La puissance active consommée par les auxiliaires représente l'énergie réellement utilisée pour faire fonctionner ces composants. Elle est mesurée en watts (W) et doit être minimisée pour maximiser l'efficacité globale du système.
- **Puissance réactive ($Q$) :** Les auxiliaires peuvent également consommer de la puissance réactive, mesurée en voltampères réactifs (VAr). Cette puissance ne produit pas de travail utile mais est nécessaire pour maintenir les champs magnétiques dans les composants inductifs, comme les transformateurs et les moteurs. Cependant, une puissance réactive excessive peut entraîner des pertes d'énergie et des problèmes électriques. Elle peut provoquer des échanges d'énergie entre les machines, perturbant ainsi le réseau électrique et réduisant son efficacité globale.

— **Puissance apparente ($S$) :** La puissance apparente, mesurée en volt-ampères (VA), est la somme vectorielle de la puissance active et réactive. Elle donne une indication de la charge totale supportée par le système, mais seule, elle ne fournit pas d'information sur l'efficacité énergétique.

Lien entre les grandeurs :

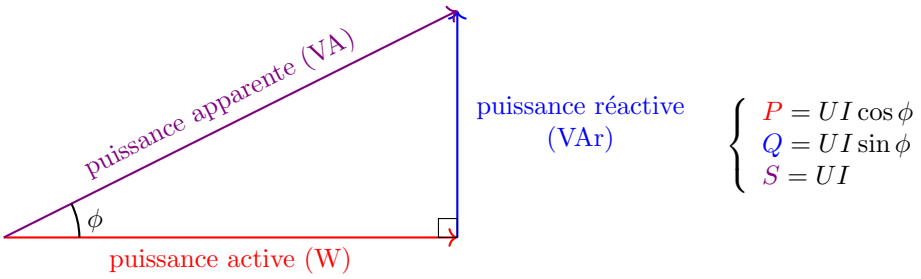


FIGURE 11 – Triangle des puissances, $S^2 = P^2 + Q^2$

$\phi$ est le déphasage entre la tension $U$ et l'intensité $I$. Ce déphasage est crucial pour les auxiliaires car il influence directement la puissance active et réactive consommée par ces composants. En régime sinusoïdal, la tension et l'intensité peuvent être déphasées (leurs valeurs maximales ne sont pas atteintes en même temps). Pour les auxiliaires, un déphasage important peut indiquer une consommation élevée de puissance réactive, ce qui peut entraîner des pertes d'énergie et réduire l'efficacité globale du système.

### Facteur de puissance

Le **facteur de puissance**, noté $\lambda$, est compris entre $[0\,;1]$ et est adimensionnel. Il est défini comme le rapport de la puissance active sur la puissance apparente : $\lambda = \frac{P}{S}$.
Le facteur de puissance est directement lié au déphasage $\phi$ entre la tension $U$ et l'intensité $I$. En effet, $\lambda = \cos\phi$. Un facteur de puissance proche de 1 indique que la puissance apparente est principalement composée de puissance active, ce qui signifie que l'énergie fournie est utilisée efficacement par les auxiliaires, sans pertes inutiles dues à la puissance réactive.
Pour les auxiliaires, un facteur de puissance élevé est essentiel pour minimiser les pertes d'énergie et maximiser l'efficacité du système. Un déphasage important peut indiquer une consommation élevée de puissance réactive, ce qui peut entraîner des pertes d'énergie et réduire l'efficacité globale du système. Par conséquent, il est crucial de maintenir un facteur de puissance supérieur à 0.9 pour assurer un fonctionnement optimal des auxiliaires et garantir une utilisation efficace de l'énergie.

**Taux de distorsion harmonique**

En plus de la puissance active, réactive et apparente, il est important de considérer le **taux de distorsion harmonique (THD)**. Le THD est une mesure de la linéarité du traitement du signal dans un appareil. Il compare le signal de sortie à un signal d'entrée parfaitement sinusoïdal. La distorsion harmonique se produit lorsque le système déforme cette sinusoïde, créant des harmoniques, c'est-à-dire des sinusoïdes de fréquences multiples de la fréquence fondamentale. Le THD est exprimé en pourcentage et représente le rapport entre la valeur efficace des harmoniques et celle de la fréquence fondamentale.

En ce qui concerne les auxiliaires de batteries, la formule $S^2 = P^2 + Q^2 + D^2$ est utilisée pour représenter la puissance apparente (S), la puissance active (P), la puissance réactive (Q) et la puissance de distorsion (D). Cette relation permet de comprendre comment les différentes composantes de la puissance interagissent dans un système électrique.

## 3.2 Traitement des données

Un fichier correspond à l'ensemble des données mesurées et enregistrées sur une plage de temps donnée. Chaque fichier est analysé dans son intégralité pour vérifier la conformité des données selon les règles suivantes :

### 3.2.1 Puissance

Chaque type de puissance (active, réactive ou apparente) est considéré comme non conforme s'il ne respecte pas certains critères.
Les fichiers sont jugés non conformes si :

- les puissances sont constantes,
- ils présentent à la fois des valeurs positives et négatives, avec plus d'un changement de signe, ce qui est généralement le signe d'un mauvais câblage de phase.

En revanche, les fichiers sont considérés comme conformes si :

- les puissances sont toujours négatives,
- ou si elles sont toujours positives, auquel cas on prend l'opposé.

### 3.2.2 Température

Les fichiers sont considérés comme non conformes si la température est constante et si elle n'est pas dans l'intervalle $[-200; 200]$°C.
Tous les autres cas sont jugés conformes.

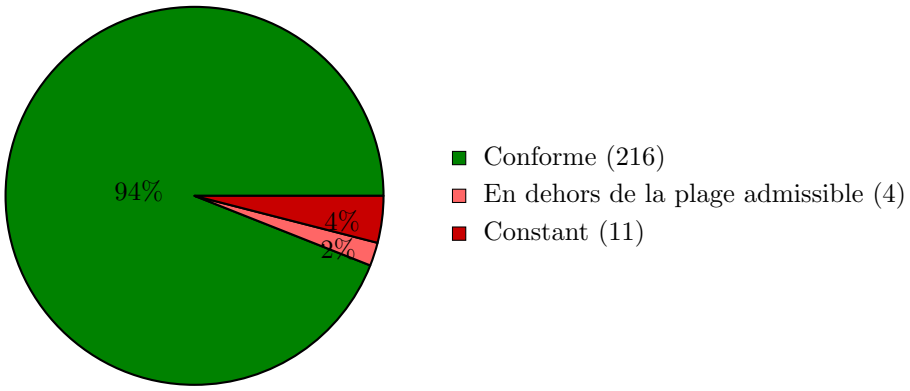


FIGURE 12 – Répartition des fichiers de température

### 3.2.3 Humidité

Les fichiers sont considérés comme non conformes si l'humidité est constante et si elle n'est pas dans l'intervalle $[0; 100]$°C.
Tous les autres cas sont jugés conformes.

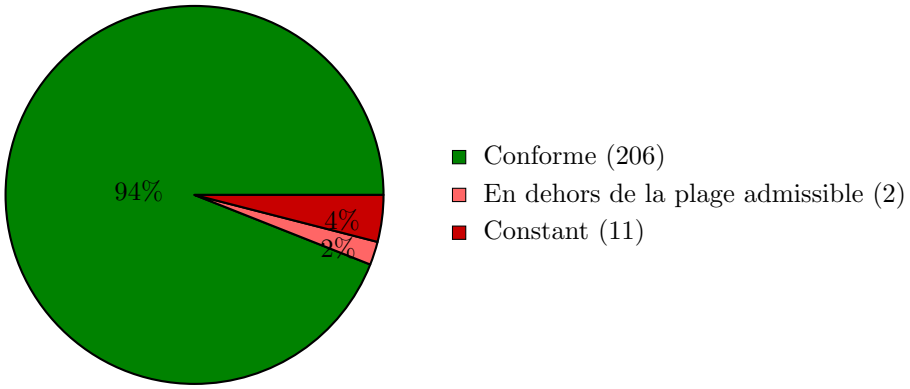


FIGURE 13 – Répartition des fichiers d'humidité extérieure

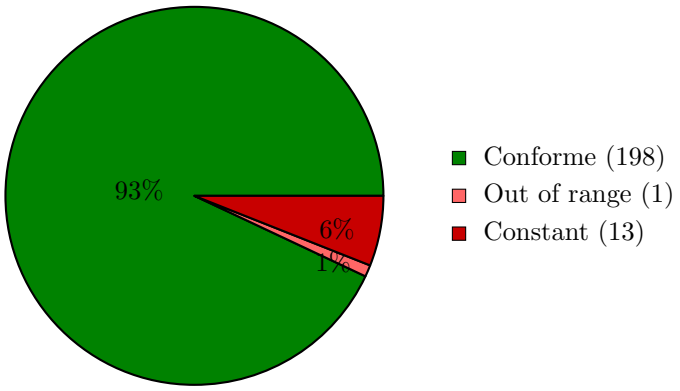


FIGURE 14 – Répartition des fichiers d'humidité intérieure

#### 3.2.4 Puissance ESS

Les fichiers concernant la puissance de l'ESS (Energy Storage System) sont considérés comme non conformes si les valeurs sont constantes tout au long de la période où l'ont a décidé de remonter les données.
Il est à noter qu'une valeur positive indique que le système décharge de l'énergie (il en fournit), tandis qu'une valeur négative signifie que le système charge de l'énergie (il en absorbe).

#### 3.2.5 Nombre de cycle par jour

Un cycle correspond à une charge et une décharge complètes du système de stockage d'énergie. Les données sur le nombre de cycles par jour sont considérées comme non conformes si les valeurs sont constantes. Tous les autres fichiers sont jugés conformes.
Pour obtenir la valeur du nombre de cycles par jour, on additionne toutes les valeurs de cycle remontées sur une période d'une journée.

### 3.3 Premiers pas

Les résultats obtenus sont ceux obtenus depuis le notebook `power_factor_firt_step.py` avec une remontée de données à partir de 1er juillet 2024 jusqu'au 11 juillet 2025.

#### 3.3.1 Validation de la relation $S^2 = P^2 + Q^2$

Pour valider l'exactitude des données remontées, nous avons d'abord effectué une analyse de régression pour vérifier la relation théorique entre la puissance apparente ($S$), la puissance active ($P$) et la puissance réactive ($Q$), exprimée par la formule $S^2 = P^2 + Q^2$. Afin de quantifier la qualité de cette relation sur l'ensemble de notre jeu de données, nous avons calculé deux indicateurs de performance pour chaque régression : le **coefficient de détermination ($R^2$)** et la **racine carrée de l'erreur quadratique moyenne ($RMSE$)**.

Les résultats agrégés de ces coefficients sont les suivants :

**$R^2$**

- Moyenne : 0,9946
- Médiane : 1,0000
- Minimum : -0,8770
- Maximum : 1,0000

**$RMSE$, en W**

- Moyenne : 47,9626
- Médiane : 0,4089
- Minimum : 0,0157
- Maximum : 3327,4452

Rappel : Au minimum les auxiliaires consomment 100 W/B CAB et au maximum 2290 W/B CAB + 2850 W/C CAB (voir Mesures sur plateforme).

L'analyse de ces métriques révèle que si la médiane du $R^2$ est proche de 1 et celle du $RMSE$ est très faible, certains fichiers présentent des coefficients médiocres (un $R^2$ négatif et un $RMSE$ très élevé). Ces anomalies suggèrent que la formule standard $S^2 = P^2 + Q^2$ pourrait ne pas être totalement adéquate pour tous les cas.

Nous émettons l'hypothèse que l'écart observé est dû à la **distorsion harmonique totale (THD)**, qui introduit une composante de puissance de distorsion ($D$) dans le calcul de la puissance apparente, modifiant ainsi la relation en $S^2 = P^2 + Q^2 + D^2$. Cependant, en l'absence de données de THD remontées par la plateforme **SoLive Pro**, cette explication reste une hypothèse que nous n'avons pas pu vérifier directement.

### 3.3.2 Facteur de puissance

Le facteur de puissance est un indicateur crucial de l'efficacité énergétique d'une installation électrique, avec une valeur idéale de 1. Un facteur de puissance est généralement considéré comme acceptable s'il est supérieur à 0,9. Les résultats de notre analyse des données, portant sur 131 fichiers, permettent de caractériser la performance des systèmes étudiés selon trois critères distincts.

**Le facteur de puissance moyen**

L'analyse du facteur de puissance moyen a révélé que la majorité des systèmes n'atteignent pas l'objectif de 0,9 sur une base annuelle.

- Seuls 25 fichiers sur 131 présentent un facteur de puissance moyen supérieur à 0,9, ce qui représente 19,1 % des cas.
- Inversement, la vaste majorité, soit 80,9 % des fichiers, opère en dessous de ce seuil recommandé.

**Stabilité du facteur de puissance**

Pour évaluer la constance de la performance, nous avons examiné le nombre de batteries dont 95 % des mesures de facteur de puissance sont supérieures ou égales à 0,9. Ce critère met en évidence la fiabilité des systèmes à maintenir un bon facteur de puissance dans le temps. Les résultats sont encore plus frappants :

- Seulement 6 fichiers sur 131, soit 4,6 %, satisfont à cette condition de stabilité.

— Cela signifie que la grande majorité des systèmes, soit 95,4 %, connaissent des variations significatives qui les empêchent de maintenir un facteur de puissance élevé de manière constante.

**Répartition par quantiles**

Pour une analyse plus granulaire de la distribution des performances, un histogramme a été généré pour répartir les systèmes par quantiles de leur facteur de puissance. Cette visualisation permet d'identifier la concentration des systèmes dans différentes tranches de performance, ce qui est essentiel pour cibler les efforts d'optimisation.

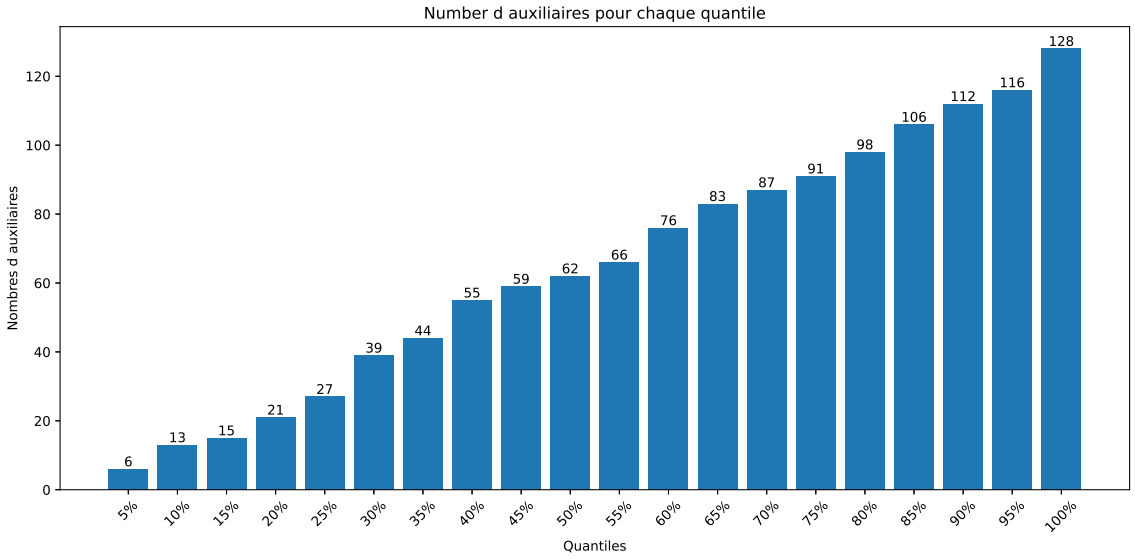


FIGURE 15 – Histogramme représentant le nombre d'auxiliaires par quantile de $0.05k$ où $k \in [\![1; 20]\!]$ avec seuil de 0,9 (fait sur Python)

Chaque barre de l'histogramme indique le nombre d'auxiliaires dont le facteur de puissance se situe dans le quantile correspondant, permettant de visualiser la concentration des performances et d'identifier les systèmes nécessitant une attention prioritaire.

## 3.4 Conclusion

Les premières analyses menées sur les données de la plateforme **SoLive Pro** révèlent deux observations majeures concernant la qualité de l'énergie des systèmes étudiés.

Tout d'abord, l'étude du **facteur de puissance** met en évidence une performance globalement faible. L'analyse montre qu'une minorité de systèmes,

soit seulement 19,1 %, respecte le seuil de 0,9 en moyenne annuelle. Cette tendance est confirmée par la forte concentration des systèmes dans les quantiles inférieurs, et une absence quasi-totale dans les tranches de performance élevées. De plus, la stabilité du facteur de puissance est très faible, puisque 95,4 % des fichiers ne parviennent pas à maintenir un facteur de puissance supérieur à 0,9 pour 95 % de leurs mesures.

En parallèle, la validation de la relation fondamentale du triangle des puissances ($S^2 = P^2 + Q^2$) a révélé des anomalies significatives pour certains fichiers, comme en témoignent des coefficients $R^2$ négatifs et des valeurs $RMSE$ élevées. Ces écarts suggèrent fortement la présence d'une composante de puissance de distorsion, issue de la **distorsion harmonique totale (THD)**, qui modifierait la relation en $S^2 = P^2 + Q^2 + D^2$.

En somme, ces résultats initiaux indiquent un problème de qualité de l'énergie récurrent, dont la cause principale serait probablement la distorsion harmonique. Des investigations complémentaires seraient nécessaires, si les données sont disponibles, pour confirmer cette hypothèse et identifier précisément les sources de ces anomalies.

# 4 Analyse de la consommation des auxiliaires, SoLive Pro

La moyenne $\bar{x}$ d'un échantillon est considérée comme représentative si pour sa médiane $\tilde{x}$ et son écart-type $\sigma$, on a $\bar{x} \in [\tilde{x} - \sigma, \tilde{x} + \sigma]$.

En effectuant cette analyse sur les fichiers de facteur de puissance, on obtient que 100% des auxiliaires ont une moyenne représentative. Pour l'analyse des consommations, on prendra donc le facteur de puissance moyen.

On a :

— la puissance active instantanée $P$,

— et la consommation d'énergie active est définie par : $E(\Delta t) = \int_{t_1}^{t_2} P(t)dt$.

Les valeurs de puissance active sont discrètes (mesurées à des instants spécifiques), on peut alors approcher l'intégrale qui définit le consommation d'énergie par une somme (méthode des rectangles) : $E \approx \sum_{i=1}^{n} P(t_i)\Delta t_i$, avec $\Delta t_i$ l'intervalle de temps entre deux mesures successives.

Pour analyser les données de consommation des auxiliaires, les facteurs de puissance moyens ont été classés par ordre croissant et divisés en 4 groupes. De même, les consommations ont été calculées pour différentes périodes : 1 heure, 6 heures, 1 jour ou 1 semaine, puis pour chaque période séparées en 4 groupes. Par la suite, le nombre d'auxiliaires a été comptabilisé par quartile de facteur de puissance dans chaque quartile de consommation.

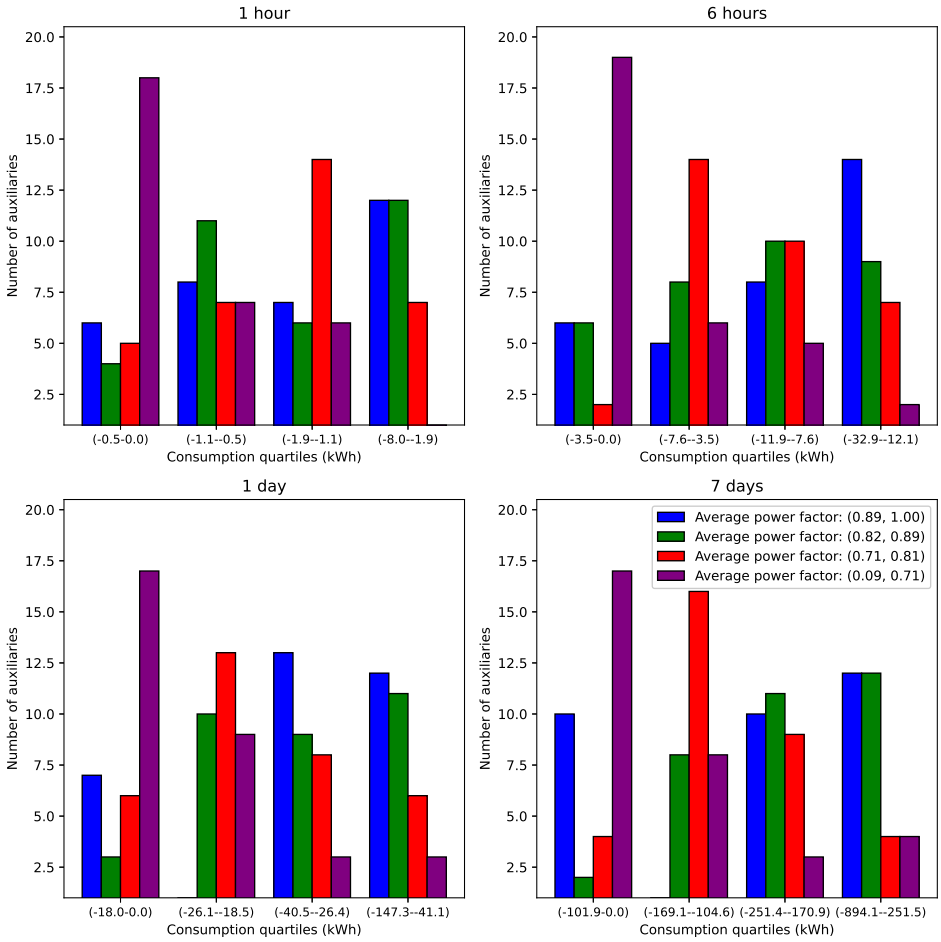


FIGURE 16 – Distribution des Quartiles des Facteurs de Puissance selon les Quartiles de Consommation pour différentes durées (fait sur Python)

L'histogramme croise la répartition des systèmes par quartiles de leur facteur de puissance moyen avec leur répartition par quartiles de consommation, sur différentes périodes.
La principale observation est la forte corrélation négative entre le facteur de puissance et la consommation d'énergie. On peut en déduire que :

- Les systèmes ayant le meilleur facteur de puissance se trouvent majoritairement dans les quartiles de consommation les plus faibles. Cela suggère que ces auxiliaires sont efficaces.
- Inversement, les systèmes avec le plus faible facteur de puissance sont principalement concentrés dans les quartiles de forte consommation.

Cette conjecture indique que les problèmes de facteur de puissance ne sont pas uniformément répartis, mais sont plutôt concentrés sur les systèmes qui

consomment le plus. En conséquence, les efforts d'optimisation devraient être priorisés sur ces auxiliaires pour obtenir le plus grand impact sur l'efficacité énergétique et la réduction des pertes.

Remarque : À la suite de ces observations, le facteur de puissance ne sera plus un axe d'analyse prioritaire pour la suite de ce projet/stage.

## 4.1 Consommation par groupe technique

Tous les systèmes de stockage d'énergie peuvent être personnalisés dans la limite du possible. *Socomec* propose une gamme standard ainsi que des solutions spécifiques. Il est donc intéressant de réaliser une analyse de consommation par groupe de caractéristiques, notamment en fonction du nombre de batteries et de la puissance disponible d'un convertisseur.

On a deux bases de données pour nos informations :

— `System Configuration Service`, `getAllSystems` : qui va nous donner sur le système son numéro de série, la puissance disponible au niveau du système de conversion bidirectionnelle AC/DC (PCS) en kW et le nombre de B-CAB.

— `IoT API Gateway` : qui va nous remonter les données des variables et aussi faire le lien entre le numéro de série du PMS et du M70 (voir Prérequis).

Le numéro de série du système prélevé dans `getAllSystems` est celui du PMS. Pour faire le lien entre le numéro du série de PMS et celui du M70, on va utilisé la gatewayId. La gatewayId correspond au nom utilisé dans So Live Pro pour l'identifiant unique d'une passerelle au sein d'un réseau informatique. Dans notre contexte, il sert à relier les valeurs du PMS et du M70 au sein d'un même système de stockage d'énergie. Un fichier `links.csv` a été créé pour faire ce lien.

Au niveau des statistiques, sur les 267 sites répertoriés, il y en a :

— 242, soit 90,6 % qui ont les deux numéros de série,

— 20, soit 7,5 % qui ont uniquement un numéro de série PMS,

— 5, soit 1,9 % avec uniquement un numéro de série M70.

### 4.1.1 Nombre de B-CAB (numberOfRacks)

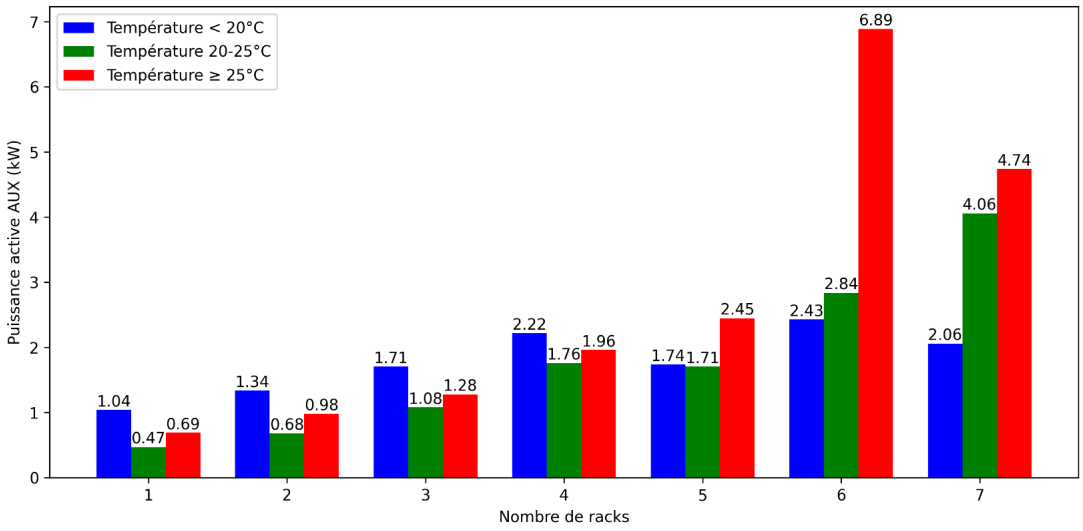


FIGURE 17 – Histogramme représentant la puissance moyenne suivant des conditions de températures

L'analyse de l'histogramme ci-dessus révèle une corrélation entre la consommation des auxiliaires, la température extérieure et la taille du système. Pour les systèmes comportant jusqu'à 4 B-CAB, on observe une tendance inverse : la puissance active moyenne des auxiliaires est plus élevée lorsque la température est basse (¡20°C). Cependant, pour les systèmes plus importants (6 et 7 B-CAB), la consommation semble augmenter significativement avec la température. Cette inversion de tendance pourrait s'expliquer par un manque de données suffisantes pour les systèmes les plus gros, ce qui pourrait biaiser la moyenne et ne pas refléter un comportement réel. Le pic de consommation de 6.89 kW pour le système à 6 B-CAB sous haute température pourrait ainsi être le résultat d'un nombre restreint de mesures dans ces conditions.

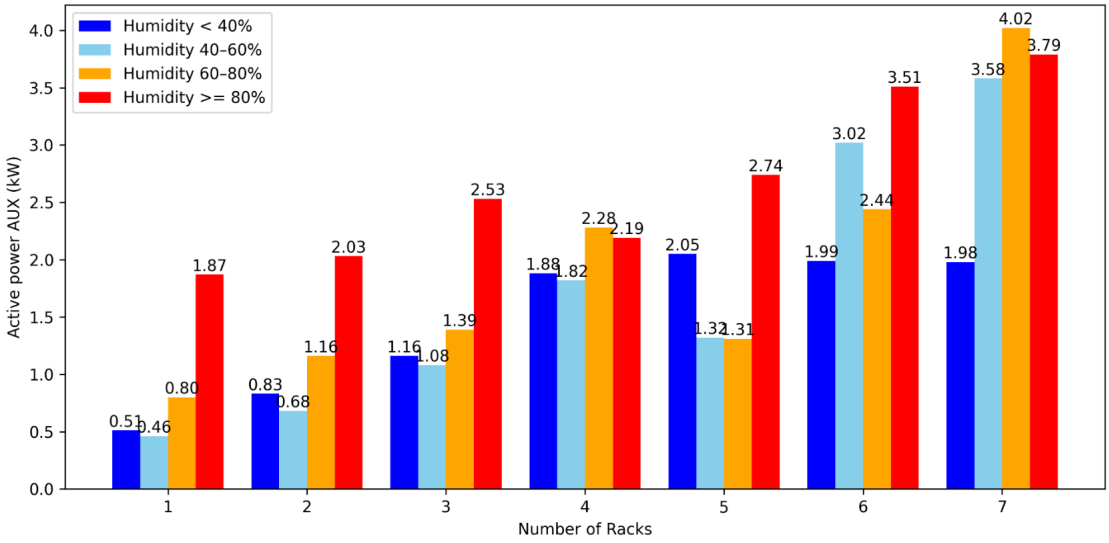


FIGURE 18 – Histogramme représentant la puissance moyenne suivant des conditions d'humidité

L'histogramme ci-dessus met en évidence une relation directe entre l'humidité extérieure et la puissance active des auxiliaires, quelle que soit la taille du système. La consommation tend à augmenter de manière générale à mesure que le taux d'humidité s'élève. On remarque notamment une forte augmentation pour les systèmes exposés à une humidité très élevée ($\geq 80\%$), confirmant l'impact significatif de ce facteur. Par exemple, pour les systèmes à 6 et 7 B-CAB, la puissance consommée sous forte humidité est nettement supérieure à celle mesurée dans d'autres conditions.

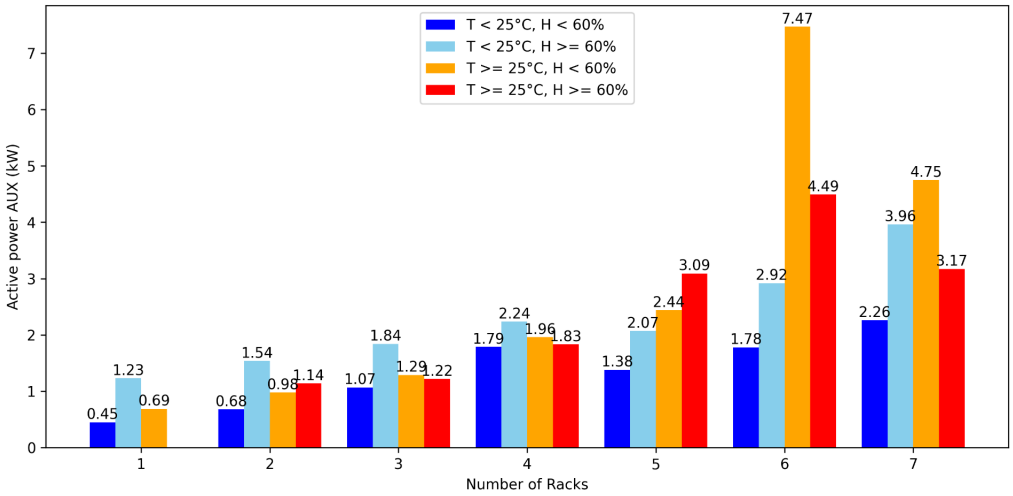


FIGURE 19 – Histogramme représentant la puissance moyenne suivant des conditions de température et d'humidité

En croisant les données de température et d'humidité, une tendance plus claire émerge. Les conditions de température élevée et d'humidité faible ($T \geq 25°C$, $H < 60\%$) entraînent les consommations les plus fortes pour les systèmes les plus importants (6 et 7 B-CAB), avec un pic à 7,47 kW pour le système à 6 B-CAB. Ces résultats confirment l'importance de ces deux paramètres combinés sur la performance du système. En revanche, les consommations les plus faibles sont généralement observées dans des conditions de basse température et de faible humidité (T¡25°C, H¡60%), et ce, pour la quasi-totalité des configurations.

### 4.1.2 Conclusion

L'analyse de la consommation par groupe technique, basée sur la configuration des systèmes, a permis de dégager plusieurs tendances majeures. L'étude a révélé l'importance des facteurs environnementaux, en particulier la température et l'humidité, sur la puissance active consommée par les auxiliaires.

Les systèmes plus petits (jusqu'à 4 B-CAB) montrent une consommation accrue par temps froid, tandis que les plus grands systèmes (6 et 7 B-CAB) semblent consommer davantage sous des températures élevées. Cette apparente inversion de tendance est cependant à interpréter avec prudence, étant probablement influencée par un nombre de données plus restreint pour les configurations les plus importantes. L'humidité a, quant à elle, un impact direct sur la consommation, qui augmente de manière générale avec le taux d'humidité, quel que soit le nombre de B-CAB.

La combinaison de ces deux facteurs confirme l'impact significatif des conditions environnementales, avec des pics de consommation observés sous des conditions de température élevée et de faible humidité pour les systèmes les plus

grands. Ces observations soulignent la complexité de modéliser la consommation des auxiliaires et l'intérêt d'une segmentation fine des données pour une analyse plus robuste.

# 5 Hiérarchisation des paramètres influençant la consommation énergétique des auxiliaires

Les variables d'entrée qui ont été retenues sont : la température, l'humidité, le nombre de cycle par jour, la puissance de l'ESS et le nombre de racks (le nombre de B CAB). Nous nous intéressons à leur influence sur la variable de sortie qui est la puissance active des auxiliaires.

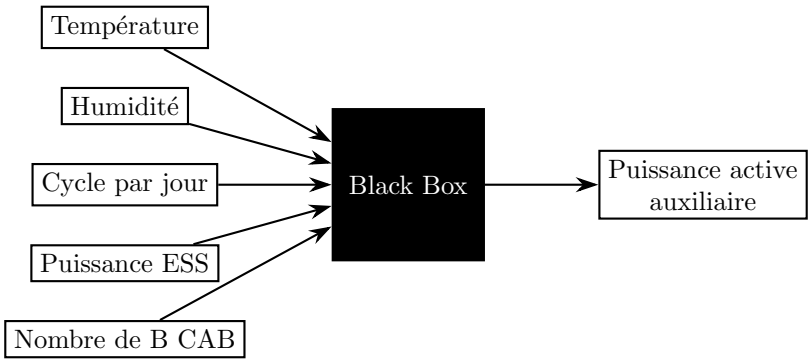


FIGURE 20 – Schéma de la modélisation de la puissance auxiliaire

Les données remontées par le VPN et celle de SoLive PRO n'ayant pas une différence significative sur un intervalle de 15 minutes, on regardera les données de SoLive PRO. Pour rappel, les données de SoLive PRO sont remontées toutes les 10 à 15 minutes : la puissance est moyennée sur cette intervalle et la valeur remontée est alors la moyenne de toutes les puissances mesurées, de même pour le température et l'humidité.

## 5.1 Création du fichier

Pour qu'un système soit retenu il faut que les valeurs remontées soient valides, qu'avec la gateway on ait réussi à faire le lien entre le numéro de série du PMS et du M70 et que le numéro de série du PMS soit dans la base de données pour pouvoir récupérer le nombre de racks. Ainsi, avec des données remontées du 07 avril 2025 (minuit) au 22 avril 2025 (entre 08h et midi), on obtient 101 fichiers qui seront fusionnés dans le fichier `.csv` final.
Le fichier final contiendra ces colonnes : temperature, humidity, cyclePerDay, powerESS, numberOfRacks. Or un des problèmes est que les valeurs des variables issues du PMS sont remontées toutes les 10 minutes alors que celles du M70 toutes les 15 minutes. Pour palier à ça on moyenne donc toutes les valeurs

comprises entre deux dates successives pour faire correspondre les dates. A noter que pour le nombre de cycles par jour, on additionne toutes les valeurs obtenues au cours d'une journée pour avoir le nombre de cycles de charge/décharge et à toute les dates on aura la valeur journalière du nombre de charge/décharge du jour en question.

## 5.2 Régression linéaire

Afin de modéliser la puissance active auxiliaire, nous avons choisi de commencer par explorer l'efficacité de plusieurs modèles de régression linéaire régularisée. Ces modèles, à savoir Ridge, Lasso et Elastic Net, ont été sélectionnés pour leur capacité à gérer la multicolinéarité potentielle entre les variables d'entrée et à prévenir le surajustement (overfitting) via la validation croisée. De plus, les normalisations min-max (MM) et centrée-réduire (S) ont été appliquées pour évaluer leur impact sur la performance des modèles.

— Ridge Cross-Validation :

$$\min_{\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \alpha \left( \sum_{j=1}^{p} \beta_j^2 \right)$$

— Lasso Cross-Validation :

$$\min_{\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \alpha \left( \sum_{j=1}^{p} |\beta| \right)$$

— Elastic Net Cross-Validation :

$$\min_{\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \alpha \left( \frac{1 - l_{1\text{ ratio}}}{2} \sum_{j=1}^{p} \beta_j^2 + l_{1\text{ ratio}} \sum_{j=1}^{p} |\beta_j| \right)$$

| Régression | Ridge Cross-Validation | | | Lasso Cross-Validation | | | Elastic Net Cross-Validation | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Brutes** | **MM** | **S** | **Brutes** | **MM** | **S** | **Brutes** | **MM** | **S** |
| **Alpha** | 1000 | 1000 | 1000 | $10^{-7}$ | $10^{-5}$ | $10^{-7}$ | $111.10^{6}$ | $7{,}42.10^{-6}$ | 0.648 |
| $l_1$ **ratio** | | | | | | | 0.1 | 0.1 | 0.1 |
| **MSE** | $12.10^{6}$ | $1{,}22.10^{-5}$ | 0.948 | $12.10^{6}$ | $1{,}22.10^{-5}$ | 0.948 | $12.10^{6}$ | $1{,}22.10^{-5}$ | 0.949 |
| $\mathbf{R}^2$ | 0.052 | 0.049 | 0.052 | 0.052 | 0.048 | 0.052 | $3{,}16.10^{-4}$ | 0.049 | 0.051 |
| **Classement** | 1. C<br>2. R<br>3. T<br>4. H<br>5. P | 1. H<br>2. R<br>3. C<br>4. T<br>5. P | 1. R<br>2. H<br>3. T<br>4. C<br>5. P | 1. R<br>2. H<br>3. T<br>4. C<br>5. P | 1. C<br>2. R<br>3. T<br>4. H<br>5. P | 1. R<br>2. H<br>3. T<br>4. C<br>5. P | 1. P<br>2. THCR | 1. H<br>2. R<br>3. T<br>4. C<br>5. P | 1. R<br>2. H<br>3. C<br>4. P<br>5. T |

FIGURE 21 – Tableau récapitulatif des résultats des régressions avec validation croisée.
MM = normalisation min-max, S = normalisation centrer-réduire.
C = Cycle par jour, R = nombre de Racks, T = Température, H = Humidité et P = Puissance ESS.

Cependant, les résultats présentés dans le tableau indiquent clairement que ces régressions n'ont pas réussi à capturer une relation significative entre les entrées (température, humidité, cycles, etc.) et la puissance auxiliaire. Les coefficients de détermination ($R^2$) sont tous proches de zéro, voire négatifs, ce qui démontre que les modèles expliquent moins de 5% de la variance des données. Parallèlement, les valeurs élevées du $MSE$ confirment l'incapacité de ces modèles à prédire avec précision la puissance auxiliaire. Ces résultats sont donc inexploitables et nous amènent à la conclusion que la modélisation de la puissance auxiliaire ne peut se faire de manière fiable par une simple régression linéaire, suggérant une relation non linéaire ou l'absence de données pertinentes pour ce type d'approche.

## 5.3 Indices de Shapley

Suite à l'échec des régressions linéaires, nous avons poursuivi notre analyse avec les indices de Shapley. Cette méthode nous permettra de quantifier précisément la contribution de chaque variable à la prédiction de la puissance auxiliaire, sans présupposer une relation linéaire. Les calculs seront d'abord effectués sur des **arbres de décision**, connus pour leur interprétabilité, avant d'être étendus aux **réseaux de neurones**, afin d'obtenir des résultats plus robustes et de valider nos observations initiales.

Les indices de Shapley sont empruntés de la théorie des jeux. Ils mesurent la contribution moyenne marginale d'une variable dans toutes les combinaisons possibles. Souvent utilisé quand le modèle est complexe (même une boîte noire), avec interactions, et qu'on cherche une explication locale ou globale.
Remarque : Le module Shapley utilisé dans le notebook donne le contribution moyenne en valeur absolue.

### 5.3.1 Arbres de décisions

Les arbres de décision sont des modèles d'apprentissage supervisé qui prennent des décisions en suivant une structure arborescente, où chaque nœud représente une condition sur une variable, menant à des branches qui aboutissent à des prédictions. Bien qu'intuitifs et faciles à interpréter, ils peuvent être instables et sujets au surapprentissage. Pour pallier ces limites, des méthodes d'ensemble comme **Random Forest** et **Gradient Boosting** ont été développées.

Remarque : Les arbres de décision ont une meilleure performance sur des données brutes. Cette performance s'explique par le fait que les arbres de décision n'opèrent pas sur les valeurs elles-mêmes, mais sur des seuils de comparaison. Ils créent des règles binaires (par exemple, "si `température > 20°C`", alors...) pour scinder les données. Contrairement à des modèles comme la régression linéaire ou les réseaux de neurones qui sont sensibles à l'échelle des variables, l'échelle des données brutes n'a donc pas d'impact sur le processus de prise de décision de l'arbre. En conséquence, il n'est généralement pas nécessaire de normaliser ou de standardiser les données pour qu'un arbre de décision fonctionne efficacement.

Les résultats présentés ici sont ceux des systèmes L (à noter que pour la hiérarchisation il n'y avait pas de différences avec les données uniquement des systèmes L et ceux de tous les systèmes). De plus, les données ont été remontées du 07 avril 2025 au 22 avril 2025.

**Gradient Boosting**
Le Gradient Boosting est une méthode d'ensemble qui construit des modèles faibles (souvent des arbres de décision peu profonds) de manière séquentielle. Chaque nouveau modèle corrige les erreurs du précédent en se concentrant sur les exemples mal prédits. Contrairement à Random Forest, les arbres ne sont pas indépendants mais construits les uns après les autres. Les modèle final est une somme pondérée de tous les arbres, ce qui permet une grande flexibilité et une haute précision. Il est très performant mais peut être plus sensible aux paramètres et au sur-apprentissage.

Une validation croisée a d'abord été menée sur l'ensemble des systèmes pour déterminer le nombre optimal d'estimateurs, qui a été fixé à 1500.

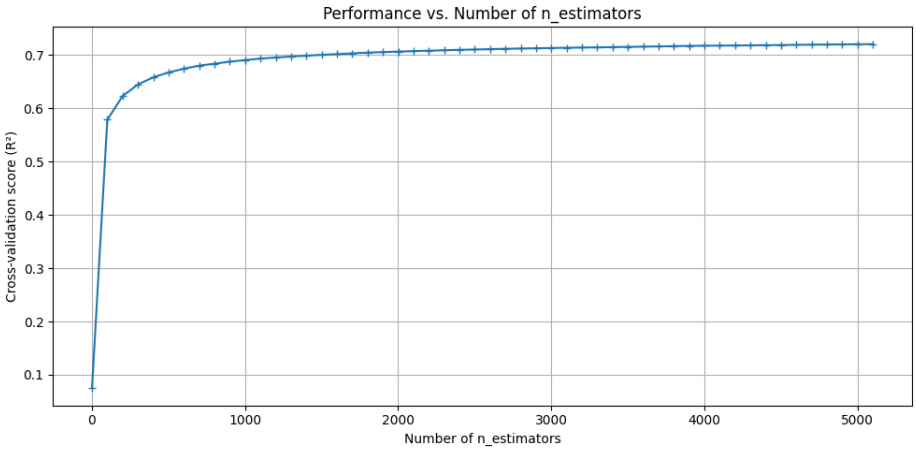


FIGURE 22 – Cross-Validation sur le nombre d'estimateurs pour Grdient Boosting sur les sytèmes L (fait sur Python)

Avec ces hyperparamètres, le modèle a obtenu un $RMSE$ de 666 et un $R^2$ de 0,71, ce qui représente une amélioration notable par rapport aux régressions initiales. En se basant sur ces résultats, les indices de Shapley ont été calculés pour évaluer l'influence de chaque variable sur la prédiction de la puissance auxiliaire. Le classement des variables par ordre d'importance (avec les valeurs absolues des coefficients de Shapley entre parenthèses) est le suivant :

1. Humidité (548)
2. Nombre de B-CAB (278)
3. Nombre de Cycle par jour (180)
4. Puissance ESS (145)
5. Température (135)

Une analyse plus fine a été réalisée en partitionnant les données par le nombre de B-CAB. Le tableau de performance qui suit résume les résultats pour différentes configurations :

| Nombre de B-CAB | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Ranking | 1. H | 1. H | 1. H | 1. H | 1. P | 1. H |
| | 2. T | 2. T | 2. C | 2. C | 2. T | 2. T |
| | 3. P | 3. P | 3. T | 3. P | 3. H | 3. C |
| | 4. C | 4. C | 4. P | 4. T | 4. C | 4. P |
| **RMSE** (en W) | 289 | 411 | 819 | 962 | 681 | 841 |
| $\mathbf{R}^2$ | 0,84 | 0,76 | 0,60 | 0,32 | 0,56 | 0,73 |
| **Number of rows** | 26847 | 39773 | 40553 | 1463 | 2943 | 14720 |

FIGURE 23 – Tableau de synthèse des performances des modèles de régression Gradient Boosting par nombre de racks. H = Humidité, T = Température, P = Puissance ESS, C = Cycle par jour.

Comme pour les régressions linéaires, la performance des modèles de Gradient Boosting varie considérablement en fonction du nombre de B-CAB. Les modèles pour les systèmes à 1, 2 et 6 B-CAB affichent de bons scores ($R^2$ supérieurs à 0,6), ce qui permet une prédiction de la puissance auxiliaire relativement fiable. En revanche, les performances pour les systèmes à 3 et 5 B-CAB sont médiocres ($R^2$ de 0,32 et 0,56), ce qui rend le modèle inexploitable pour ces configurations.

Des validations croisées spécifiques ont été réalisées pour les configurations à 4 et 5 B-CAB afin d'optimiser les hyperparamètres du modèle.

1. Pour les systèmes à 4 B-CAB, le nombre optimal d'estimateurs est de 80, ce qui a permis d'obtenir un $R^2$ de 0,48 et un $RMSE$ de 840 W. Bien que ces résultats restent insuffisants pour une prédiction fiable, le classement des variables a changé, la Puissance ESS devenant la plus influente.
2. Pour les systèmes à 5 B-CAB, le nombre optimal d'estimateurs est de 500. Le modèle a obtenu un $R^2$ de 0,63 et un $RMSE$ de 637 W, confirmant une performance acceptable, mais avec un nouveau classement où l'humidité et la température sont les variables les plus importantes.

Ces résultats confirment que les systèmes sont très hétérogènes, et qu'un modèle de prédiction unique n'est pas adapté pour tous les cas. L'importance des variables et la performance du modèle sont intrinsèquement liées à la configuration physique du système.

**Random Forest**

La Random Forest est un algorithme d'apprentissage automatique basé sur un ensemble d'arbres de décision. Elle construit plusieurs arbres sur des sous-échantillons aléatoires des données, puis combine leurs prédictions pour améliorer la précision et réduire le sur-apprentissage. Chaque arbre vote, et la majorité

(en classification) ou la moyenne (en régression) détermine la production finale. Elle est robuste aux données bruitées et gère bien les variables non linéaires et corrélées.
C'est un modèle souvent utilisé comme référence pour sa performance et sa simplicité d'utilisation. Cependant dans notre cas d'utilisation, le programme s'est avéré très long à tourné c'est pourquoi nous l'avons donc utilisé sur un échantillon de données :

- Nombre d'estimateurs (`n_estimators`) : Fixé à 100.
- Taille de l'échantillon (`sample_size`) : Égale au minimum de 10 000 ou la taille de l'ensemble d'entraînement (`len(X_train)`).

Ces paramètres ont été appliqués pour chaque sous-ensemble de données, partitionné par le nombre de racks, afin de maintenir une cohérence dans l'approche méthodologique. Le choix d'une taille d'échantillon limitée (10 000) permet de réduire le temps de calcul tout en assurant une performance représentative pour les ensembles de données les plus volumineux.

Voici le tableau de synthèse des performances des modèles Random Forest par nombre de racks :

| **Nombre de Racks** | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|
| **Classement** | 1. H<br>2. T<br>3. P<br>4. C | 1. H<br>2. T<br>3. C<br>4. P | 1. H<br>2. C<br>3. P<br>4. T | 1. P<br>2. T<br>3. C<br>4. H | 1. H<br>2. T<br>3. C<br>4. P | 1. H<br>2. C<br>3. T<br>4. P |
| **RMSE** (en W) | 265 | 377 | 799 | 835 | 573 | 790 |
| $\mathbf{R^2}$ | 0,86 | 0,80 | 0,62 | 0,48 | 0,70 | 0,76 |
| **Nombre de lignes** | 26847 | 39773 | 40553 | 1463 | 2943 | 14720 |

FIGURE 24 – Tableau de synthèse des performances des modèles de régression Random Forest par nombre de racks. H = Humidité, T = Température, P = Puissance ESS, C = Cycle par jour.

Comme pour les Gradient Boosting, l'analyse a été menée en segmentant les données par le nombre de racks. Les résultats, résumés dans le tableau ci-dessous, montrent que le modèle Random Forest est particulièrement performant sur certaines configurations, notamment celles à 1, 2, 3, 5 et 6 racks (avec des $R^2$ supérieurs à 0,60). Cependant, la performance est moins bonne pour les configurations à 4 racks, ce qui renforce l'idée de l'hétérogénéité des systèmes.

**Conclusion :** L'analyse menée avec les modèles **Random Forest** et **Gradient Boosting** a confirmé l'impossibilité de créer un modèle de prédiction

unique et universel. Les performances varient considérablement d'une configuration à l'autre, avec des résultats solides pour certains systèmes et des échecs pour d'autres, ce qui souligne la forte hétérogénéité des données.
Malgré ces variations, une conclusion transversale se dégage : l'humidité apparaît de manière constante comme la variable la plus influente dans la majorité des configurations de racks. Cette observation confirme son rôle prépondérant dans la prédiction de la puissance auxiliaire et la met en évidence comme un paramètre clé à considérer pour les futures études, même si des modèles spécifiques à chaque configuration de système doivent être développés pour obtenir des prédictions fiables.

### 5.3.2 Réseaux de neurones

Un réseau de neurones récurrents (RNN = Recurrent Neural Network) est un réseau de neurones qui traite des données dans l'ordre, en mémorisant les informations précédentes grâce à une boucle interne. Cela le rend utile pour des tâches comme : la traduction automatique, la reconnaissance vocale, l'analyse des sentiments ou la prédictions des séries temporelles. Mais les RNN classiques ont un problème majeur : ils ont du mal à retenir des informations sur le long terme à cause du phénomène de gradient évanescent (le gradient devient trop petit pour apprendre efficacement).

Remarque : Contrairement aux arbres de décision, qui ne sont pas sensibles à l'échelle des données, les réseaux de neurones sont très influencés par les valeurs d'entrée. Il est donc nécessaire de les normaliser : dans le cadre de ce projet on aura testé deux normalisations : centrée-réduite et min-max.

Les données utilisées seront les mêmes que celles des arbres de décisions, aussi uniquement sur les systèmes L.
Pour chacun des modèles, nous avons effectué une validation croisée sur le nombre d'époques et affiché une courbe montrant l'évolution de l'erreur d'apprentissage et de l'erreur de validation.

**LSTM**
Un LSTM est une amélioration du RNN qui résout ce problème. Il introduit une structure interne plus complexe, avec : une cellule mémoire qui garde l'information sur le long terme et 3 portes qui contrôlent le flux d'information. Les 3 portes sont :

1. porte d'oubli : décide quoi oublier
2. porte d'entrée : décide quoi ajouter à la mémoire
3. porte de sortie : décide quoi transmettre à la sortie

Grâce à ce mécanisme, un LSTM peut retenir des informations importantes pendant de longues séquences ce qui le rend très performant pour des tâches comme : la génération du texte, la modélisation du langage ou la prévision de séries temporelles complexes.

Les résultats après l'entraînement du modèle était les suivants :
— Données centrées réduites : la meilleure époque est de 200, avec un coefficient $R^2$ de 0,88 et un $RMSE$ de 427 W.
— Données min-max : la meilleure époque est aussi de 200, avec un coefficient $R^2$ de 0,70 et un $RMSE$ de 670 W.

**GRU**

GRU est aussi une solution au problème exposé au début de la partie. Il introduit des portes qui contrôlent le flux d'information. Les 2 portes principales sont :

1. porte de mise à jour : décide de combien d'informations passées doivent être conservées
2. porte de réinitialisation : décide combien d'informations passées doivent être oubliées

Ces mécanismes permettent au GRU de mémoriser des séquences longues, s'entraîner plus rapidement que les LSTM (car il y a moins de paramètres) et d'être plus simples à implémenter. Les deux portes permettent au GRU de contrôler le flux d'information sans avoir besoin d'une cellule mémoire séparée comme dans le LSTM.

Les résultats après l'entraînement du modèle était les suivants :
— Données centrées réduites : la meilleure époque est de 200, avec un coefficient $R^2$ de 0,88 et un $RMSE$ de 428 W.
— Données min-max : la meilleure époque est aussi de 200, avec un coefficient $R^2$ de 0,71 et un $RMSE$ de 657 W.

Pour le calcul des indices de Shapley, les résultats était les mêmes que pour les modèles arbres de décision pour tous les systèmes L confondus : l'humidité se démarque nettement suivie du nombre de B-CAB et enfin la température, le nombre de cycle et la puissance de l'ESS qui sont proches.

# 6 Optimisation

Afin de trouver les conditions pour lesquelles la consommation des auxiliaires est optimale, il a fallu résoudre un problème d'optimisation sachant que notre fonction est une boîte noire mais que nous l'avons modélisée par un réseau de neurones. Cinq méthodes ont été retenues afin d'explorer les différents résultats que nous pouvons avoir avec.

1. **Nelder-Mead :** La méthode de Nelder-Mead est une technique d'optimisation sans dérivés qui fonctionne en manipulant un simplex, une figure géométrique formée de $n+1$ points dans un espace de dimension $n$ ($n \in N^*$). A chaque itération, le simplex est modifié à l'aide d'opérations

comme la réflexion, la contraction ou la réduction, afin de se rapprocher d'un minimum local. Remarque : Même si elle est simple à mettre en oeuvre, elle peut échouer sur des fonctions très irrégulières ou dans des espaces de grande dimension.

2. **CMA-ES** (Covariance Matrix Adaptation Evolution Strategy) : CMA-ES est une méthode d'optimisation stochastique inspirée des algorithmes évolutionnaires, conçue pour résoudre des problèmes complexes, non convexes ou bruits. Elle génère des candidats selon une distribution gaussienne dont la moyenne, l'écart-type et la matrice de covariance sont continuellement ajustés pour guider efficacement la recherche vers les zones prometteuses. Très robuste, elle est particulièrement utile quand la dérivée de la fonction n'est pas disponible.
3. **Descente de gradient avec contraintes :** La descente de gradient avec contraintes consiste à minimiser une fonction tout en respectant des contraintes (égalités ou inégalités). ¨Plusieurs techniques peuvent être utilisées : la projection sur le domaine admissible, les multiplicateurs de Lagrange, ou les méthodes de barrières. À chaque étape, le gradient guide la dimension de descente, mais les contraintes influencent le chemin en forçant la solution à rester dans un espace valide.
4. **Simulated Annealing** (Recuit simulé) : Le retrait simulé est une méthode probabiliste d'optimisation qui s'inspire du processus de refroidissement des matériaux en métallurgie. Elle explore l'espace des solutions en acceptant parfois des mouvements vers des solutions moins bonnes, avec une probabilité qui diminue au fil du temps (température décroissante). Cette stratégie permet d'échapper aux minimaux locaux et favorise la convergence vers un optimum global, bien que lentement.
5. **Optimisation bayésienne :** L'optimisation bayésienne est une stratégie adaptée aux fonctions coûteuses à évaluer. Elle construit un modèle probabiliste (souvent un processus gaussien) pour approximer la fonction objectif, puis utilise une fonction d'acquisition pour choisir intelligemment les prochains points à tester. Elle équilibre l'exploration (chercher dans des zones incertaines) et l'exploration (approfondir les zones prometteuses), ce qui la rend très efficace avec peu d'évaluations.

Les tests effectués avec ces différentes méthodes sur le réseau de neurones n'ont pas donné de résultats concluants, révélant une instabilité significative. Les observations se sont résumées à deux cas principaux :

— Convergence vers les bornes du domaine : Les algorithmes tendaient à converger vers les conditions extrêmes définies par les limites du problème, ce qui pouvait indiquer une sensibilité excessive aux frontières ou une incapacité à trouver un optimum interne.
— Forte dépendance à la condition initiale : Pour les méthodes nécessitant un point de départ, de légers changements dans les conditions initiales conduisaient à des résultats d'optimisation radicalement différents. Cette instabilité a soulevé des doutes quant à la fiabilité des solutions trouvées.

Pour pallier ces difficultés, nous avons exploré plusieurs pistes. D'une part, un effort a été fait pour améliorer la qualité de l'entraînement du réseau de neurones spécifiquement sur les systèmes L, mais cette tentative n'a pas été concluante. D'autre part, une approche de segmentation du modèle a été envisagée : créer un réseau de neurones distinct pour chaque catégorie de B-CAB. L'objectif était que ce nouveau modèle, plus spécialisé, surpasse en performance le réseau initial sur l'ensemble des systèmes. Malheureusement, le temps d'entraînement très long et l'émergence d'autres priorités n'ont pas permis de finaliser cette approche.

# 7 Mesures sur plateforme

Sur la plateforme de test de Nordhouse, un chiller **Pfannenberg**, destiné à un éventuel ajout aux systèmes **SUNSYS HES-XXL**, a été démonté pour réaliser des tests et mesures, mais n'a finalement pas été retenu. Ces tests visaient à évaluer les performances et l'efficacité énergétique du chiller dans différentes conditions de fonctionnement.

## 7.1 Relevé des paramètres

Les puissances actives nominales relevées sur le chiller mais aussi sur les auxiliaires de la C-CAB sont répertoriés dans le tableau ci-dessous.

| Auxiliaires | |
|---|---|
| B CAB : Chiller | C CAB |
| Ventilateurs : $3 \times (140 - 180)$ W | Ventillateurs : $2 \times (170)$ W |
| Chauffage : 2000 W | Chauffage : $3 \times (950)$ W |
| Compresseur : 1050 W | |
| Pompe : 190 W | |
| Electronique : 100 W | |

Le chiller possède quatre modes de fonctionnement :

— **Standby** : Ce mode est utilisé lorsque le chiller est en attente et que le système ne nécessite ni de refroidissement ni chauffage. Seule l'électronique fonctionne.
Dans ce mode de fonctionnement, les auxiliaires restent actifs et consomment 100 W/B CAB.

— **Self circulating** : Ce mode permet de faire circuler le fluide sans activer le compresseur ou les ventilateurs, permettant de maintenir une circulation interne. Les composants utilisés sont ici la pompe et l'électronique.
Les auxiliaires consomment 290 W/B CAB.

— **Refroidissement** : Ce mode est activé pour abaisser la température du système à un niveau souhaité. Pendant cette phase, le compresseur, les ventilateurs, la pompe et l'électronique fonctionnent.
Les auxiliaires consomment 1880 W/B CAB et 340 W/C CAB.
— **Chauffage** : Ce mode est utilisé pour augmenter la température du système à un niveau souhaité mais aussi pour faire baisser le taux d'humidité si celui-ci est trop élevé. Le chauffage, la pompe et l'électronique sont utilisés.
Les auxiliaires consomment 2290 W/B CAB et 2850 W/C CAB.

Remarques :

1. En mode de refroidissement, le système peut fonctionner en régulation : les composants ne tournent pas nécessairement à pleine puissance. En revanche, en mode chauffage, une fois activé, le système fonctionne toujours à puissance maximale (il est soit en marche, soit à l'arrêt, sans mode intermédiaire).
2. Si l'humidité est trop élevée, le système passe en mode chauffage pour l'éliminer. Dans le cas où le système relève un température trop élevée et une humidité trop élevée aussi, il commence par chauffer pour évacuer l'humidité, puis passe en mode refroidissement.

Pour contrôler l'état (température/humidité) du système la logique est la suivante :

1. **Détection de la tension** : Un capteur de tension intégré dans le module batterie mesure la tension des cellules.
2. **Gestion de la batterie (BMS)** : Le système de gestion de la batterie (Battery Management System) analyse les données du capteur de tension et d'autres paramètres pour assure une gestion optimale de la batterie.
3. **Consigne de température du coolant** : La température souhaitée du fluide de refroidissement est définie en fonction des besoins du système.
4. **Activation du système associé** : En fonction de la consigne de température, le mode approprié (refroidissement ou chauffage) est activé pour atteindre la température cible.
5. **Atteinte de la consigne** : Une fois la température cible atteinte, le système ajuste son fonctionnement pour maintenir cette température de manière stable et efficace.

Lors de la phase de test, il était prévu de relever la consommation globale du chiller, ainsi que celle de ses composants principaux : les ventilateurs, la pompe et le compresseur. Ces données devaient permettre d'évaluer l'efficacité énergétique du système et d'identifier des pistes d'amélioration pour optimiser ses performances.

Cependant, il n'a finalement pas été possible de collecter ces mesures comme prévu. Par conséquent, l'analyse de l'efficacité énergétique et l'optimisation du chiller n'ont pas pu être menées à terme.

# 8 Missions annexes

## 8.1 Rajout du dictionnaire pour les données remontées avec le VPN

Le Power Management System (PMS) constitue le cœur décisionnel du système. Il est physiquement implanté dans l'Automation Box, située au sein de la C-CAB, mais fonctionne essentiellement comme un logiciel. Le PMS traite et analyse les données du système afin de piloter ses opérations, notamment en déterminant les niveaux optimaux de charge et de décharge des batteries.

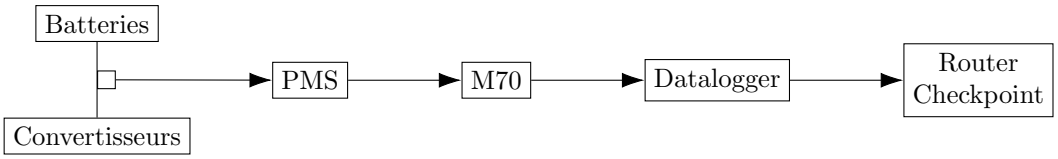


FIGURE 25 – Schéma de la remontée des données batterie et convertisseur jusqu'au Checkpoint

Au début de notre analyse, nous nous sommes posés la question de la source de données à privilégier : celles remontées par le VPN ou celles de So Live Pro. La distinction est cruciale : les données de So Live Pro sont des moyennes calculées sur des intervalles de 10 à 15 minutes, tandis que les données du VPN sont des valeurs "instantanées", actualisées toutes les 10 à 1000 millisecondes selon les sites.

Pour évaluer la pertinence de ces données en temps réel, nous avons mis en place un script permettant de se connecter directement aux sites via le VPN. Nous avons également enrichi ce script en ajoutant un dictionnaire pour extraire spécifiquement les données de puissance auxiliaire qui nous intéressaient.

Afin de comparer les deux sources, nous avons sélectionné de manière aléatoire cinq sites et avons effectué des remontées de données. L'objectif était de déterminer si les variations observées sur 10 à 15 minutes justifiaient l'utilisation des données VPN. Cependant, les résultats ont montré que les variations de puissance entre les données VPN n'étaient pas suffisamment significatives sur cet intervalle. Par conséquent, nous avons décidé de conserver les données de So Live Pro, qui sont plus adaptées à notre analyse sur le long terme.

## 8.2 Amélioration du script de remontée de données de So Data Battery

L'outil **CLI_SoDaBa** a été initialement développé pour répondre à un besoin fondamental : l'extraction et l'organisation des données de systèmes de stockage d'énergie (ESS) collectées par le datalogger et transmises à la plateforme So Data Battery. Dans sa version de base, le programme permettait déjà une

récupération efficace des données. Il les structurait dans des fichiers distincts, organisés selon le niveau de hiérarchie des équipements : le système global, les racks individuels ou les modules de batteries.

Mais des améliorations ont été faites pour rendre l'outil plus robuste et flexible :

- Gestion multi-systèmes et multi-périodes : L'une des limitations de la version initiale était l'obligation de traiter les systèmes un par un. La nouvelle version permet désormais de spécifier plusieurs numéros de série et de définir différentes plages de dates pour chaque système en une seule exécution, ce qui a considérablement optimisé le flux de travail lors des analyses de données à grande échelle.
- Options de filtrage par défaut : Pour simplifier et accélérer le processus d'extraction, des réponses prédéfinies (presets) ont été ajoutées pour les options de filtrage.
- Organisation et clarté de la structure des données : La gestion des fichiers de sortie a été repensée pour offrir une meilleure traçabilité. Les données extraites sont maintenant stockées dans une arborescence de dossiers claire et logique, nommée selon le format `serial_YYYY-MM-DD_YYYY-MM-DD`, ce qui facilite la navigation et l'archivage.
- Automatisation et sécurité des accès : La saisie manuelle des identifiants de connexion à chaque utilisation représentait un point de friction. Pour y remédier, un fichier `credentials.py` a été introduit, récupéré et adapté à partir d'autres programmes Socomec développés par mon référent Raphaël Viudes. Cette solution simplifie le processus d'authentification et renforce la sécurité des identifiants en recommandant leur inclusion dans le fichier `.gitignore`, évitant ainsi leur commit accidentel.

En synthèse, ces évolutions ont fait passer l'outil d'une solution fonctionnelle à un programme évolutif. L'accent a été mis sur l'amélioration de l'expérience utilisateur, la robustesse du traitement des données et la sécurisation des informations d'accès.

## 8.3 Analyse de données pour PMS V2

Dans la version 2 du Product Management System, l'installation d'un nouvel écran sur les systèmes est envisagée. Pour déterminer si cette installation est réalisable, une analyse des données des systèmes en fonctionnement depuis la création de la BU a été nécessaire.
On a donc remonté les données de tous les systèmes depuis janvier 2017 et fait 3 analyses :

1. Isoler tous les intervalles de temps où la température extérieure a été supérieure ou égale à 36°C. Sur ces intervalles, il faudra ensuite calculer le nombre total d'occurrences ainsi que la durée moyenne, minimale, maximale et médiane.
2. Faire de même avec la température intérieure mais avec un seuil à 49°C.

3. Déterminer la proportion d'intervalles où la température extérieure était à plus de 36°C et où le système de séchage était actif.

Il est crucial de noter que pour toutes ces analyses, seules les températures comprises entre -70°C et 70°C seront prises en compte. Un "intervalle" est défini par une série de mesures consécutives avec un écart temporel inférieur à une heure. Dès qu'un écart dépasse une heure, l'intervalle est considéré comme clos. De plus, pour gérer les intervalles incomplets, la durée finale sera estimée en utilisant la moyenne des durées des intervalles courts ($< 1$ heure) qui la précèdent.

# 9 Conclusion et perspectives

Ce stage a été une expérience particulièrement enrichissante, me permettant de développer mes compétences en analyse de données, en modélisation statistique et en apprentissage automatique. J'ai pu appliquer des concepts théoriques à des données réelles et concrètes, tout en m'adaptant aux spécificités techniques et aux problématiques de l'entreprise.

Les analyses menées ont confirmé que la consommation des auxiliaires est un phénomène complexe et multifactoriel, fortement influencé par des paramètres environnementaux comme l'humidité ou la température, ainsi que par la configuration physique des systèmes (nombre de racks). L'échec des modèles de régression linéaire a souligné la nature non linéaire de cette relation. En revanche, les modèles basés sur les arbres de décision (Random Forest et Gradient Boosting) et les réseaux de neurones (GRU et LSTM) ont montré une capacité de prédiction nettement supérieure, avec des scores de performance acceptables sur certaines configurations de systèmes. Ces modèles ont également permis d'identifier l'humidité comme la variable la plus influente dans la majorité des cas.

Cependant, il est apparu clairement qu'un modèle de prédiction unique n'est pas adapté à l'ensemble des systèmes, en raison de leur forte hétérogénéité. Les performances et le classement des variables varient considérablement d'une configuration à l'autre. Une analyse approfondie n'a pas révélé de surapprentissage (overfitting) significatif, ce qui valide la robustesse des modèles retenus.

Pour l'avenir, une perspective intéressante serait de créer une application capable de prédire la consommation des auxiliaires. Pour cela, il serait impératif de développer des modèles spécifiques à chaque configuration de système pour garantir des prédictions fiables et précises. En parallèle, l'approche d'optimisation, bien que non concluante lors de ce stage en raison d'une instabilité des résultats, pourrait être retravaillée avec un modèle de prédiction plus robuste pour explorer les conditions d'opération optimales et ainsi identifier les leviers d'amélioration de l'efficacité énergétique.

# Annexes

## A Informations système

| Élément | Détail |
|---|---|
| Système d'exploitation | Windows 10 (10.0.19045) |
| Architecture | AMD64 |
| Processeur | Intel(R) Core(TM) Ultra 7 165H |
| Cœurs physiques | 16 |
| Threads logiques | 22 |
| Fréquence CPU | 1400.00 MHz |
| Mémoire RAM totale | 31.46 GB |
| Disque principal (C :) | 953.24 GB |

Table 1 – Informations système de la machine utilisée

## B Basse Tension

La **basse tension** désigne l'électricité dont la tension est inférieure ou égale à 1000 V en courant alternatif. Elle est utilisée dans la majorité des installations électriques domestiques, tertiaires et industrielles pour alimenter des équipements standards (éclairage, moteurs systèmes informatiques, etc.).
Dans le cadre de *Socomec*, la basse tension ne se limite pas à une simple distribution d'énergie : elle est au cœur de solutions intelligentes et sécurisées qui permettent de gérer, convertir, protéger et stocker cette énergie de manière fiable et performante.

## C Statistique : Moyenne, Médiane, Ecart-Type et Quantiles

Les paramètres statistiques permettent de résumer la distribution d'une variable quantitative.
Il y a deux types de paramètres :

- paramètres de position : moyenne, médiane, quantiles, etc.
- paramètres de dispersion : écart-type, etc.

## Moyenne

La **moyenne**, noté $\bar{x}$, est un paramètre statistique de position qui représente la valeur centrale d'une distribution. Elle est obtenue en additionnant toutes les valeurs d'un ensemble de données et en divisant le total par le nombre d'observations. La moyenne est très utilisée pour résumer un ensemble de données, mais elle peut être influencée par des valeurs extrêmes (appelées **outliers**).
La moyenne est définie par : $\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i$, où les $x_i$ sont les différentes valeurs prises par le jeu de données et $n$ est le nombre total d'observations.

## Médiane

La **médiane** est aussi un paramètre de position qui permet de situer la valeur centrale d'un ensemble de données. Contrairement à la moyenne, la médiane n'est pas influencée par les valeurs extrêmes. Elle divise l'ensemble des données en deux parties égales : 50% des valeurs sont inférieures à la médiane, et 50% sont supérieures.
Pour trouver la médiane, il faut d'abord trier les données par ordre croissant. Si le nombre de données est impair, la médiane est la valeur située au milieu de l'ensemble. Si le nombre est pair, la médiane est la moyenne des deux valeurs centrales.
La médiane est particulièrement utile lorsque les données sont asymétriques ou contiennent des valeurs aberrantes.

## Ecart-Type

L'écart-type est un paramètre de dispersion qui mesure la variabilité des données par rapport à la moyenne. Plus l'écart-type est élevé, plus les données sont dispersées, ce qui signifie que les valeurs sont éloignées de la moyenne. À l'inverse, un écart-type faible indique que les données sont concentrées autour de la moyenne. Il se calcule en prenant la racine carrée de la variance, qui est la moyenne des carrés des différences entre chaque donnée et la moyenne. L'écart-type est particulièrement utile pour évaluer la stabilité ou la prévisibilité d'un phénomène.
L'écart-type est défini par : $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$, où les $x_i$ sont les différentes valeurs prises par le jeu de données et $\bar{x}$ la moyenne de celui-ci.

## Quantiles

Les quantiles sont des valeurs qui divisent un ensemble de données en plusieurs intervalles égaux. Les quantiles d'ordre $p$ sont des valeurs qui partagent les données en $\frac{1}{p}$ parties égales. Par exemple, le premier quartile (Q1) est la valeur qui sépare les 25% des données les plus faibles, et le troisième quartile (Q3) sépare les 75% des données les plus faibles. Le médiane est en fait le deuxième quartile (Q2).

Outre ces quartiles, il existe aussi des quantiles d'ordre arbitraire, tels que le quantile d'ordre 0.05, qui correspond à la valeur en dessous de laquelle se trouvent 5% des données. Ces quantiles d'ordre $p$, où $p \in [0; 1]$, sont utilisés pour analyser les données de manière plus précise, en se concentrant sur des segments spécifiques de la distribution. De manière générale, les quantiles permettent de mieux comprendre la répartition des données, d'identifier les tendances et d'analyser les valeurs aberrantes, qu'elles soient faibles ou élevées. Ils sont également utilisés pour calculer l'intervalle interquartile, qui mesure l'écart entre Q3 et Q1 et donne une idée de la dispersion des données centrales.

# D L'erreur quadratique moyenne et l'erreur $R^2$

L'erreur quadratique moyenne (Mean Squared Error, $MSE$) et l'erreur $R^2$ sont deux mesures couramment utilisées pour évaluer la qualité d'un modèle de régression, mais elles mesurent des aspects différents de la performance.

## L'erreur quadratique moyenne

C'est une mesure de la différence moyenne entre les valeurs prédites par le modèle et les valeurs réelles observées. Calculée comme la moyenne des carrés des erreurs, c'est-à-dire la différence entre chaque valeur réelle et prédite, élevée au carré. Elle est exprimée en unités carrées de la variable de sortie. Plus la valeur du MSE est faible, mieux c'est, car cela signifie que les prédictions sont proches des valeurs réelles.
Définition : $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, où $\hat{y}_i$ sont les valeurs prédites par le modèle.

## L'erreur $R^2$

C'est une mesure de la proportion de la variance des données expliquée par le modèle. Il varie entre 0 et 1 (ou peut être négatif si le modèle est extrêmement mauvais). Un $R^2$ proche de 1 indique que le modèle explique bien la variance des données, tandis qu'un $R^2$ proche de 0 suggère que le modèle n'explique pas bien les variations. Un $R^2$ négatif peut apparaître si le modèle est moins précis que simplement prédire la moyenne des données.
Définition : $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$, où $\hat{y}_i$ sont les valeurs prédites par le modèle et $\bar{y}$ est la moyenne des valeurs observées $y_i$.

## Différences

$MSE$ mesure l'erreur absolue (l'écart quadratique moyen entre les prédictions et les vraies valeurs), tandis que $R^2$ mesure la proportion de la variance expliquée par le modèle. MSE est plus directe dans la mesure des erreurs, mais il est difficile à interpréter sans référence (car il est dans les unités carrées de la variable de sortie). En revanche, $R^2$ donne une idée plus intuitive de la qualité du modèle

en termes de proportion de la variance expliquée. $R^2$ peut parfois être trompeur, surtout si le modèle est trop complexe ou trop simple. Par exemple, il peut être élevé même pour un modèle qui surajuste les données.
En résumé, $MSE$ donne une mesure d'erreur brute, tandis que $R^2$ montre dans quelle mesure votre modèle explique la variabilité des données.

# Document References


[1] *Bayesian Optimization*. Wikipedia, consulté le 3 juillet 2025. URL : https://en.wikipedia.org/wiki/Bayesian_optimization.

[2] *Covariance Matrix Adaptation Evolution Strategy*. Wikipedia, consulté le 4 juillet 2025. URL : https://cma-es.github.io/.

[3] *Covariance Matrix Adaptation Evolution Strategy (CMA-ES)*. https://cma-es.github.io/. Consulté le 3 juillet 2025.

[4] *Elastic Net Regularization*. Wikipedia, consulté le 24 avril 2025. URL : https://en.wikipedia.org/wiki/Elastic_net_regularization.

[5] *Gated Recurrent Unit*. Wikipedia, consulté le 28 mai 2025. URL : https://en.wikipedia.org/wiki/Gated_recurrent_unit.

[6] *Gradient Boosting*. Wikipedia, consulté le 7 mai 2025. URL : https://en.wikipedia.org/wiki/Gradient_boosting.

[7] *Gradient Descent*. Wikipedia, consulté le 3 juillet 2025. URL : https://en.wikipedia.org/wiki/Gradient_descent.

[9] Farzad Jaliliantabar, Rizalman Mamat et Sudhakar Kumarasamy. "Prediction of lithium-ion battery temperature in different operating conditions equipped with passive battery thermal management system by artificial neural networks". In : *Materials Today : Proceedings* 48 (2022). Innovative Manufacturing, Mechatronics  Materials Forum 2021, p. 1796-1804. ISSN : 2214-7853. DOI : https://doi.org/10.1016/j.matpr.2021.09.026. URL : https://www.sciencedirect.com/science/article/pii/S2214785321058223.

[11] *Lasso (statistics)*. Wikipedia, consulté le 24 avril 2025. URL : https://en.wikipedia.org/wiki/Lasso_(statistics).

[14] *Long Short-Term Memory*. Wikipedia, consulté le 28 mai 2025. URL : https://en.wikipedia.org/wiki/Long_short-term_memory.

[15] *Nelder–Mead method*. Wikipedia, consulté le 3 juillet 2025. URL : https://en.wikipedia.org/wiki/Nelder%E2%80%93Mead_method.

[16] *NumPy Documentation*. Consultée à plusieurs reprises entre février et juillet 2025. URL : https://numpy.org/doc/stable/.

[17] *Random Forest*. Wikipedia, consulté le 7 mai 2025. URL : https://en.wikipedia.org/wiki/Random_forest.

[18] *Ridge Regression*. Wikipedia, consulté le 24 avril 2025. URL : https://en.wikipedia.org/wiki/Ridge_regression.

[19] *scikit-learn Documentation*. Consultée le 24 avril 2025. URL : https://scikit-learn.org/stable/documentation.html.

[20] *SHAP Documentation*. Consultée de 15 mai 2025. URL : https://shap.readthedocs.io/.

[21] *Simulated Annealing*. Wikipedia, consulté le 4 juillet 2025. URL : `https://en.wikipedia.org/wiki/Simulated_annealing`.

## Image References

[10] *Joseph SIAT*. Consulté le 17 juin 2025. URL.

[12] *Logo de l'UFR*. Consulté le 27 février 2025. URL.

[13] *Logo de Socomec*. Consulté le 20 février 2025. URL.

[22] *Socome siege social*. Consulté le 18 juin 2025. URL.

[23] *Socome usine 2*. Consulté le 18 juin 2025. URL.

[24] *Socome usine 3*. Consulté le 18 juin 2025. URL.