

Analyse de données

Consommation des systèmes auxiliaires

Justine EHRET
Encadrant : Raphaël VIUDES

Août 2025

Sommaire

- 1 Présentation de l'entreprise
- 2 Contexte et mission du stage
- 3 Présentation des données
 - Visualisation
 - Puissance active, réactive et apparente
- 4 Analyse des données
 - Contrôle de qualité des données
 - Résultats de qualité des données
- 5 Hiérarchisation de l'influence des paramètres
 - Modèle
 - Régression linéaire
 - Arbres de décision
 - Réseaux de neurones
- 6 Optimisation
- 7 Conclusion et Perspective
- 8 Annexe

Sommaire

- 1 Présentation de l'entreprise
- 2 Contexte et mission du stage
- 3 Présentation des données
 - Visualisation
 - Puissance active, réactive et apparente
- 4 Analyse des données
 - Contrôle de qualité des données
 - Résultats de qualité des données
- 5 Hiérarchisation de l'influence des paramètres
 - Modèle
 - Régression linéaire
 - Arbres de décision
 - Réseaux de neurones
- 6 Optimisation
- 7 Conclusion et Perspective
- 8 Annexe

Présentation de l'entreprise

Socomec

- Fondée en 1922
- Initialement : équipements électromagnétiques (sonnettes, coupe-circuits, interrupteurs)
- Évolution vers la gestion, conversion et stockage de l'énergie électrique basse tension
- Entreprise familiale



Figure: Joseph SIAT, fondateur de Socomec [2]



Figure: Ivan STEYERT, actuel PDG de Socomec [1]

Présentation de l'entreprise

Socomec

Chiffres clés

- Présence dans plus de 80 pays
- 12 sites de production
- plus de 4400 collaborateurs
- CA 2024 : 924 M€
- 8 % du CA réinvesti en R&D

Domaine d'expertise



POWER
SWITCHING



POWER
MONITORING



POWER
CONVERSION



ENERGY
STORAGE



EXPERT
SERVICES



- Créée en 2017
- SUNSYS HES-L et HES-XXL, systèmes modulaires, capables de stocker de 150 kWh à plus de 1,5 MWh d'énergie



Figure: **SUNSYS HES-L** *Socomec*

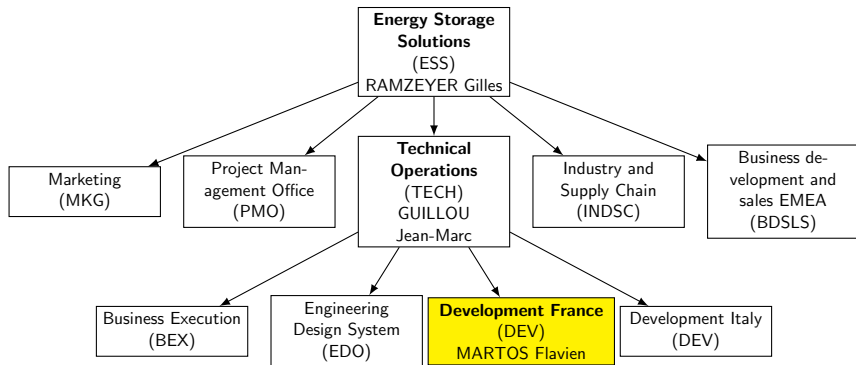


Figure: Organigramme du pôle Energy Storage Solutions (ESS)

Intégrée au pôle Développement France, encadrée par Raphaël VIUDES (Ingénieur développement de données).

Sommaire

- 1 Présentation de l'entreprise
- 2 Contexte et mission du stage
- 3 Présentation des données
 - Visualisation
 - Puissance active, réactive et apparente
- 4 Analyse des données
 - Contrôle de qualité des données
 - Résultats de qualité des données
- 5 Hiérarchisation de l'influence des paramètres
 - Modèle
 - Régression linéaire
 - Arbres de décision
 - Réseaux de neurones
- 6 Optimisation
- 7 Conclusion et Perspective
- 8 Annexe

Contexte

Dans les systèmes de stockage d'énergie, les auxiliaires sont indispensables au fonctionnement et à la sécurité. Leur consommation a un impact direct sur l'efficacité énergétique globale.

Objectifs / Mission du stage

- ➊ Prise en main et compréhension des données
- ➋ Analyse des consommations
- ➌ Structuration et exploitation des données
- ➍ Définition d'indicateurs et modélisation
- ➎ Optimisation

Analyse réalisée sur des notebooks Python et déposée sur le GitLab de Socomec.

Sommaire

- 1 Présentation de l'entreprise
- 2 Contexte et mission du stage
- 3 **Présentation des données**
 - Visualisation
 - Puissance active, réactive et apparente
- 4 Analyse des données
 - Contrôle de qualité des données
 - Résultats de qualité des données
- 5 Hiérarchisation de l'influence des paramètres
 - Modèle
 - Régression linéaire
 - Arbres de décision
 - Réseaux de neurones
- 6 Optimisation
- 7 Conclusion et Perspective
- 8 Annexe

Présentation des données

Visualisation



Figure: **SUNSYS HES L**, Socomec

Min : 100 W/B CAB.

Max : 2290 W/B CAB + 2850 W/C CAB.

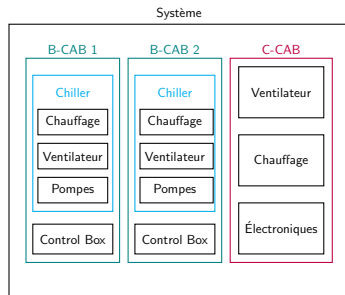


Figure: Schéma d'un système avec les différents auxiliaires

Présentation des données

Puissance active, réactive et apparente

- **Puissance active, P** : puissance utile au système (en W).
- **Puissance réactive, Q** : (en VAr) énergie qui crée des pertes et peut faire des problèmes électriques. Elle s'échange entre les machines et perturbent le réseau électrique.
- **Puissance apparente, S** : seule elle n'indique rien (en VA).

Lien entre les grandeurs : ϕ est le déphasage entre U et I

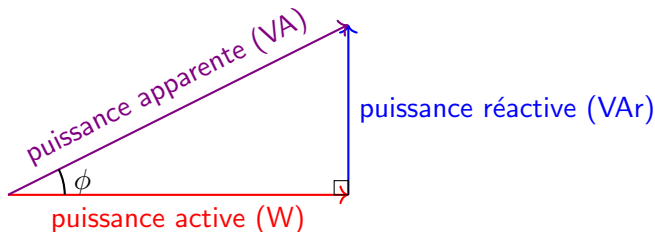


Figure: Triangle des puissances,

$$S^2 = P^2 + Q^2$$

Sommaire

- 1 Présentation de l'entreprise
- 2 Contexte et mission du stage
- 3 Présentation des données
 - Visualisation
 - Puissance active, réactive et apparente
- 4 Analyse des données
 - Contrôle de qualité des données
 - Résultats de qualité des données
- 5 Hiérarchisation de l'influence des paramètres
 - Modèle
 - Régression linéaire
 - Arbres de décision
 - Réseaux de neurones
- 6 Optimisation
- 7 Conclusion et Perspective
- 8 Annexe

Règles appliquées aux différents paramètres :

Paramètre	Conforme	Non conforme
Puissance (P,Q,S)	Signe constant (+ ou -)	Constante / signe variable
Température	$-200 \leq T \leq 200$ °C	Constante / hors intervalle
Humidité (int./ext.)	$0 \leq H \leq 100\%$	Constante / hors intervalle
Puissance ESS	+ décharge / - charge	Constante

Nombre de cycles par jour = nombre de cycles de charge/décharge par jour

Du 07 avril 2025 au 22 avril 2025 :

- Puissance active : 92% des sites avec des données utilisables
- Température : 94%
- Humidité : 94%
- puissance ESS : 95 %

Site considéré = données conformes + 2 numéros de série + présent dans la base de données

⇒ **101 sites considérés** pour le jeu de données final.

Sommaire

- 1 Présentation de l'entreprise
- 2 Contexte et mission du stage
- 3 Présentation des données
 - Visualisation
 - Puissance active, réactive et apparente
- 4 Analyse des données
 - Contrôle de qualité des données
 - Résultats de qualité des données
- 5 Hiérarchisation de l'influence des paramètres
 - Modèle
 - Régression linéaire
 - Arbres de décision
 - Réseaux de neurones
- 6 Optimisation
- 7 Conclusion et Perspective
- 8 Annexe

Hiérarchisation de l'influence des paramètres

Modèle

Quantifier l'importance relative de chaque paramètre

- Les variables d'entrée : la température, l'humidité, le nombre de cycle par jour, la puissance de l'ESS et le nombre de B CAB.

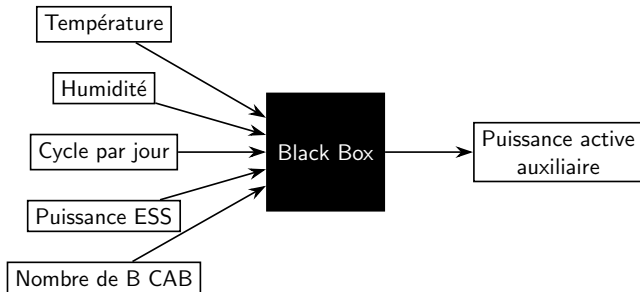


Figure: Schéma de la modélisation de la puissance auxiliaire

- Variable de sortie : puissance active des auxiliaires.

Hiérarchisation de l'influence des paramètres

Régression linéaire

Régression linéaire

- 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 :

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

- Elastic Net Cross-Validation :

$$\min_{\beta} \frac{1}{2n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \alpha \left(\frac{1 - l_1 \text{ ratio}}{2} \sum_{j=1}^p \beta_j^2 + l_1 \text{ ratio} \sum_{j=1}^p |\beta_j| \right)$$

Hierarchisation de l'influence des paramètres

Régression linéaire

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
RMSE	3464.1	0.0110	0.974	3464.1	0.0110	0.974	3464.1	0.0110	0.974
R²	0.052	0.049	0.052	0.052	0.048	0.052	$3,16.10^{-4}$	0.049	0.051
Classement	1. C	1. H	1. R	1. R	1. C	1. R	1. P 2. THCR	1. H	1. R
	2. R	2. R	2. H	2. H	2. R	2. H		2. R	2. H
	3. T	3. C	3. T	3. T	3. T	3. T		3. T	3. C
	4. H	4. T	4. C	4. C	4. H	4. C		4. C	4. P
	5. P	5. P	5. P	5. P	5. P	5. P		5. P	5. T

Figure: Tableau résumé des résultats des régressions cross validées. MM = normalisation min-max, S = normalisation centrer-réduire. C = Cycles par jour, R = nombre de B-CAB, T = Température, H = Humidité et P = Puissance ESS.

Indice de Shapley : Mesurent l'importance relative de chaque variable en évaluant sa contribution moyenne sur toutes les combinaisons possibles.

- Arbres de décision : Random Forest et Gradient Boosting
- Réseau de neurones : LSTM et GRU

Hiérarchisation de l'influence des paramètres

Arbres de décision

- **Gradient Boosting** : Modèle séquentiel corrigeant les erreurs des arbres précédents, très précis mais sensible aux hyperparamètres. Nombre optimal d'estimateurs : 1500. *RMSE* de 666 et un R^2 de 0,71 :
 - ① Humidité (548)
 - ② Nombre de B-CAB (278)
 - ③ Nombre de Cycle par jour (180)
 - ④ Puissance ESS (145)
 - ⑤ Température (135)
- **Random Forest** : Ensemble d'arbres indépendants sur échantillons aléatoires, robuste et rapide à calculer. Nombre d'estimateurs : 100, échantillon limité à 10 000 lignes.

Hiérarchisation de l'influence des paramètres

Arbres de décision

Nombre de B-CAB	Gradient Boosting				Random Forest			
	RMSE	R ²	Classement	N lignes	RMSE	R ²	Classement	N lignes
1	289	0,84	H, T, P, C	26847	265	0,86	H, T, P, C	10000
2	411	0,76	H, T, P, C	39773	377	0,80	H, T, C, P	10000
3	819	0,60	H, C, T, P	40553	799	0,62	H, C, P, T	10000
4	962	0,32	H, C, P, T	1463	835	0,48	P, T, C, H	1463
5	681	0,56	P, T, H, C	2943	573	0,70	H, T, C, P	2943
6	841	0,73	H, T, C, P	14720	790	0,76	H, C, T, P	10000

Figure: Tableau résumé des résultats des indices de Shapley avec les arbres de décisions. C = Cycle par jour, R = nombre de Racks, T = Température, H = Humidité et P = Puissance ESS.

Conclusion :

- Les performances dépendent du nombre B-CAB.
- Humidité variable la plus influente.

Hiérarchisation de l'influence des paramètres

Réseaux de neurones

Modèle	Portes	Normalisation	R^2	RMSE (W)
LSTM	1. Oubli 2. Entrée 3. Sortie	Centrée-réduite	0,88	427
		Min-Max	0,70	670
GRU	1. Réinitialisation 2. Mise à jour	Centrée-réduite	0,88	428
		Min-Max	0,71	657

Figure: Comparaison des performances des modèles RNN (LSTM et GRU) selon le type de normalisation et les portes utilisées.

- Sensibles à l'échelle des données = la normalisation influence les performances.
- Indices de Shapley : l'humidité variable la plus influente, suivie du nombre de B-CAB et de la température.

Sommaire

- 1 Présentation de l'entreprise
- 2 Contexte et mission du stage
- 3 Présentation des données
 - Visualisation
 - Puissance active, réactive et apparente
- 4 Analyse des données
 - Contrôle de qualité des données
 - Résultats de qualité des données
- 5 Hiérarchisation de l'influence des paramètres
 - Modèle
 - Régression linéaire
 - Arbres de décision
 - Réseaux de neurones
- 6 Optimisation
- 7 Conclusion et Perspective
- 8 Annexe

5 méthodes : Nelder-Mead, CMA-ES, Descente de Gradient avec contraintes, Simulated Annealing et Optimisation Bayésienne

Problèmes rencontrés :

- Instabilité élevée
- Convergence aux bornes : sensibilité aux limites
- Forte dépendance aux conditions initiales : résultats très variables






Solutions explorées :





- Amélioration entraînement sur systèmes L : non concluante
- Segmentation par nombre de B-CAB : trop long, non finalisé

Sommaire

- 1 Présentation de l'entreprise
- 2 Contexte et mission du stage
- 3 Présentation des données
 - Visualisation
 - Puissance active, réactive et apparente
- 4 Analyse des données
 - Contrôle de qualité des données
 - Résultats de qualité des données
- 5 Hiérarchisation de l'influence des paramètres
 - Modèle
 - Régression linéaire
 - Arbres de décision
 - Réseaux de neurones
- 6 Optimisation
- 7 Conclusion et Perspective
- 8 Annexe

Conclusion

-  Stage enrichissant, développement de compétences en analyse de données, modélisation statistique et apprentissage automatique.
-  Modèles linéaires insuffisants ; arbres de décision et réseaux de neurones performants.
-  Humidité identifiée comme la variable la plus influente.
-  Performances et classement des variables varient selon la configuration des systèmes : modèle unique peu adapté.
-  Pas de sur-apprentissage significatif, robustesse des modèles validée.

-  Modèles spécifiques par configuration
-  Optimisation énergétique et modèles robustes
-  Conditions opérationnelles optimales
-  Analyse des leviers de performance

Sommaire

- 1 Présentation de l'entreprise
- 2 Contexte et mission du stage
- 3 Présentation des données
 - Visualisation
 - Puissance active, réactive et apparente
- 4 Analyse des données
 - Contrôle de qualité des données
 - Résultats de qualité des données
- 5 Hiérarchisation de l'influence des paramètres
 - Modèle
 - Régression linéaire
 - Arbres de décision
 - Réseaux de neurones
- 6 Optimisation
- 7 Conclusion et Perspective
- 8 Annexe

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: Informations système de l'ordinateur utilisé

- Python : 3.9.6
 - IPython : 8.18.1, ipykerne : 6.29.5, jupyter_client : 8.6.3, jupyter_core : 5.7.2, traitlets : 5.14.3
 - Library : Pandas : 2.2.3, Matplotlib : 3.9.4, Numpy : 2.0.2, Scikit-learn : 1.6.1, Shap : 0.47.2

- [1] *Ivan STEYERT*. Consulté le 17 juin 2025. URL: URL.
- [2] *Joseph SIAT*. Consulté le 17 juin 2025. URL: URL.