

# Rapport de Série Temporelle

Analyse et Prédiction des Séries Temporelles  
à l'aide de modèles statistiques et deep learning

Projet réalisé par :

**IChaoui Chaima**

**Isaa Ayda**

Année universitaire : 2025 – 2026

# Table des matières

<b>Introduction</b>	<b>2</b>
0.1 Exploration et Prétraitement des Données . . . . .	2
0.1.1 Description des données . . . . .	2
0.1.2 Analyse exploratoire . . . . .	2
0.1.3 Prétraitement des données . . . . .	3
0.2 Modélisation avec SARIMA . . . . .	4
0.2.1 Résultats et évaluation . . . . .	4
0.3 Modélisation avec LSTM . . . . .	5
0.3.1 Justification du LSTM : . . . . .	5
0.3.2 Configuration : . . . . .	5
0.3.3 L'architecture du modèle LSTM . . . . .	5
0.3.4 Résultats . . . . .	5
0.4 Modélisation avec Prophet . . . . .	6
0.4.1 préparation des données pour prophet . . . . .	6
0.4.2 Résultats . . . . .	6
0.5 Détection d'Anomalies . . . . .	6
0.5.1 Justification des Approches . . . . .	6
0.5.2 RÉSULTATS . . . . .	6
0.6 Comparaison des Modèles . . . . .	7
0.7 Dashboard et Visualisation . . . . .	8
<b>Conclusion</b>	<b>9</b>

# Introduction

L'objectif de ce rapport est d'analyser une série temporelle, d'appliquer différentes techniques de prévision et de comparer leurs performances.

## 0.1 Exploration et Prétraitement des Données

### 0.1.1 Description des données

Présentation de la source de données, variables, fréquence temporelle et contexte du problème.

### 0.1.2 Analyse exploratoire

#### — Visualisation de la série temporelle

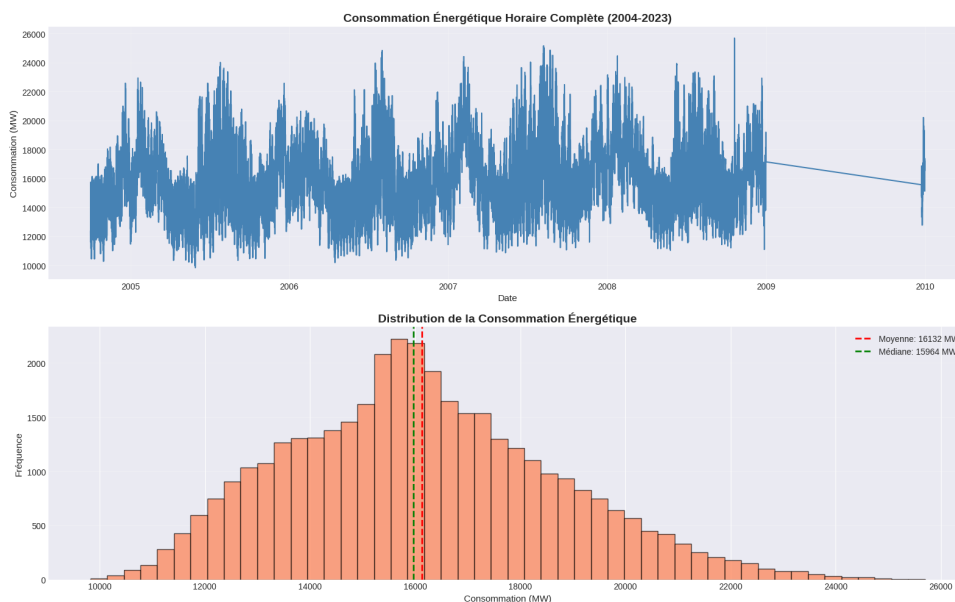


FIGURE 1 – Visualisation de la série temporelle

La série temporelle montre une consommation énergétique fortement saisonnière avec des cycles réguliers, des pics marqués et une variabilité importante au fil

du temps. La distribution est légèrement asymétrique à droite, avec une moyenne proche de la médiane, indiquant une consommation globalement stable autour de 16 000 MW .

- **Test de la stationnarité**

- **ADF** : p-value  $\approx 0$  la série est stationnaire.

- **KPSS** : p-value = 0.01 la série est non stationnaire.

- **Décomposition de la Série Temporelle** La série a été décomposée avec une période de 24 heures (cycle quotidien) en trois composants : trend (decomposition.trend) représentant la tendance à long terme, seasonal (decomposition.seasonal) mettant en évidence la composante cyclique quotidienne, et residuals (decomposition.resid) correspondant aux résidus ou au bruit.

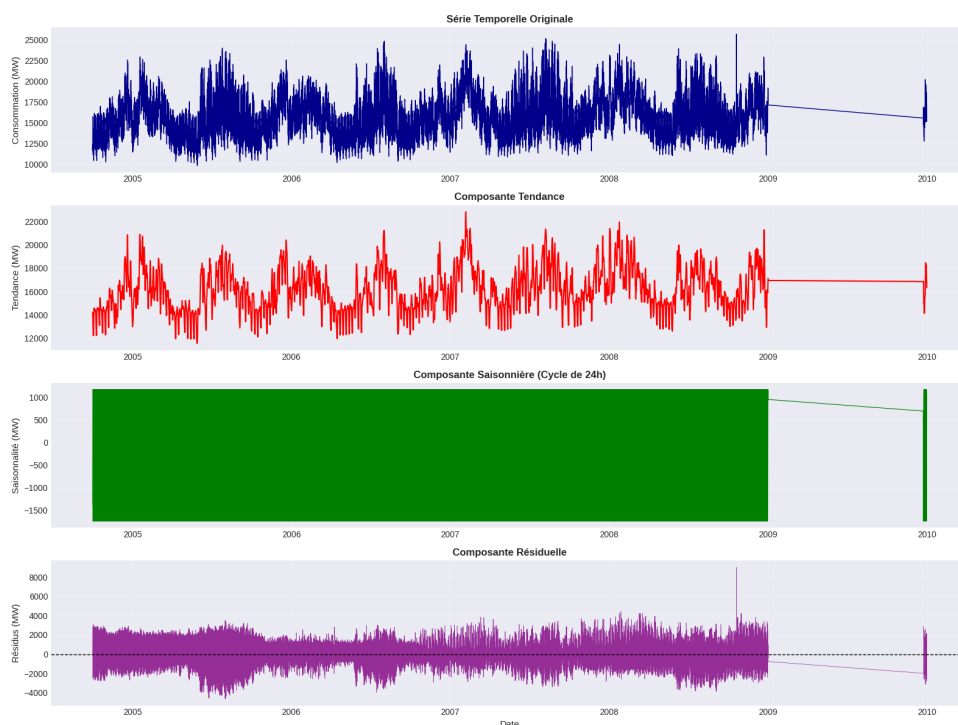


FIGURE 2 – Décomposition de la Série Temporelle

### 0.1.3 Prétraitement des données

- **Gestion des valeurs manquantes** Vérification et traitement et imputation par Interpolation linéaire
- **Détection des outliers** Détection des outliers via méthode IQR Les outliers ne sont PAS supprimés car ils peuvent représenter des événements réels
- Transformation (log, différenciation)
- **Normalisation / standardisation** Normalisation avec MinMaxScaler
- **Récupération des données météorologiques réelles (API Open-Meteo)** Ajouter température et humidité réelles pour améliorer les modèles Région : AEP

(Columbus, Ohio - 39.98°N, -82.89°W) Source : Open-Meteo API (gratuit, sans authentification).

— **FEATURE ENGINEERING ET DONNÉES**

- Variables retardées : capture les valeurs passées de la série à différents décalages (heures, jours, semaine)
- Moyennes mobiles : moyennes sur différentes fenêtres temporelles .
- Écart-type mobile : mesure la volatilité sur une fenêtre donnée
- Transformations cycliques : encode les cycles horaires, journaliers et mensuels pour capturer la circularité
- Interactions croisées : combine plusieurs variables pour modéliser les effets conjoints
- Ratios et dérivées : différences et pourcentages de variation .
- Indicateurs calendaires : distingue week-end, jours fériés .
- Pattern quotidien : capture le comportement moyen par heure et l'écart par rapport à ce pattern

## 0.2 Modélisation avec SARIMA

### Prévision univariée avec SARIMA

— **Justification du SARIMA :**

- SARIMA : Modèle statistique robuste pour les séries temporelles avec saisonnalité.
- Auto ARIMA : Trouve automatiquement les meilleurs paramètres  $(p, d, q)(P, D, Q)_s$ .
- Avantages : Excellent pour capturer les patterns saisonniers, interprétabilité élevée.
- Limitation : Supposé que la future saisonnalité sera similaire à celle passée.

— **Configuration de l'ensemble de test :**

- Test Set : Derniers 30 jours (720 heures) pour évaluation robuste.
- Train Set : Reste des données ( 19 ans).

### 0.2.1 Résultats et évaluation

Le modèle ARIMA(2,1,2)(1,0,1)[24] avec intercept est le meilleur selon l'AIC.

- **MAE** : 1857.83 MW en moyenne, les prévisions s'écartent de  $\pm 1858$  MW par rapport aux valeurs réelles.
- **RMSE** : 2455.52 MW l'erreur quadratique moyenne est un peu plus élevée que la MAE, ce qui indique que quelques erreurs importantes influencent le modèle.
- **MAPE** : 0.11% en pourcentage, les prévisions sont très précises, avec un écart moyen de seulement 0.11% par rapport à la consommation réelle.

## 0.3 Modélisation avec LSTM

### 0.3.1 Justification du LSTM :

- Mémoire à long terme : Capable de capturer des dépendances sur longues périodes
- Flexibilité : Peut apprendre des patterns non-linéaires complexes
- Séquences variables : Gère bien les dépendances temporelles
- Avantage compétitif : Souvent surpasse les modèles statistiques purs

### 0.3.2 Configuration :

- Entrée : 24 heures de données (séquence d'entrée)
- Sortie : 6 prochaines heures (horizon de prévision)
- Normalisation : MinMaxScaler pour données entre 0 et 1
- Régularisation : Dropout et Early Stopping pour éviter le surapprentissage

### 0.3.3 L'architecture du modèle LSTM

- **Entrée** : séquence de 24 heures avec 6 features par pas de temps (24, 6)
- **Première couche LSTM** : 128 unités avec **Dropout 0.2** pour régularisation
- **Deuxième couche LSTM** : 64 unités avec **Dropout 0.2**
- **Couche Dense** : 32 neurones avec activation **ReLU** pour capturer des combinaisons non linéaires
- **Couche de sortie Dense** : 6 neurones correspondant aux **prévisions sur 6 heures**

### 0.3.4 Résultats

- **MAE** : 322.41 MW — en moyenne, les prévisions s'écartent de  $\pm 322$  MW par rapport aux valeurs réelles.
- **RMSE** : 456.87 MW — l'erreur quadratique moyenne est plus élevée que la MAE, indiquant la présence de quelques écarts plus importants.
- **MAPE** : 2.01% — l'erreur relative est faible, montrant que le modèle prédit avec une bonne précision par rapport à l'échelle de la consommation.

## 0.4 Modélisation avec Prophet

### 0.4.1 préparation des données pour prophet

Mettre la colonne de datetime comme ds, la colonne cible d'énergie comme y, et ajouter les autres features comme régresseurs externes avec add\_regressor.

Régresseurs : Température, Humidité, Features cycliques, Lags, Moyennes mobiles

### 0.4.2 Résultats

L'erreur moyenne absolue est d'environ 1108 MW (**MAE**) , l'erreur quadratique moyenne **RMSE** est légèrement plus élevée (1334 MW), ce qui indique que quelques écarts restent plus importants, et l'erreur relative est extrêmement faible **MAPE** ( $\text{MAPE} = 0.07\%$ ), montrant que le modèle capture presque parfaitement la consommation réelle.

## 0.5 Détection d'Anomalies

### 0.5.1 Justification des Approches

#### Méthodes

- **Méthode Statistique (3-sigma)** : simple, interprétable, basée sur la théorie statistique
- **Isolation Forest** : non-paramétrique, adaptée aux distributions non-gaussiennes

#### Configuration

- **Seuil 3-sigma** :  $|résidu| > 3 \times \text{écart-type} = \text{anomalie}$
- **Isolation Forest** : anomalies basées sur l'isolabilité du point

### 0.5.2 RÉSULTATS

#### Méthode 1 :Détection d'anomalie par méthode de seuil

- **Nombre d'anomalies détectées** : 1 (0.14%)
- **Détail de l'anomalie** :
  - Date/heure : 2009-12-30 10 :00 :00
  - Valeur : 19328.00 MW
  - Écart par rapport à la moyenne : 7498.08 MW

#### Méthode 2 : Isolation Forest

- **Nombre d'anomalies détectées** : 36 (5.00%)

- **Top 5 anomalies (score d'anomalie) :**
  - 2008-12-28 07 :00 :00 : 11774.00 MW (score : -0.664)
  - 2008-12-28 06 :00 :00 : 11327.00 MW (score : -0.657)
  - 2008-12-28 08 :00 :00 : 12351.00 MW (score : -0.649)
  - 2008-12-22 09 :00 :00 : 22930.00 MW (score : -0.648)
  - 2008-12-22 08 :00 :00 : 22727.00 MW (score : -0.646)

## 0.6 Comparaison des Modèles

### Métriques Utilisées

- **MAE** : Erreur absolue moyenne (même unité que les données)
- **RMSE** : Racine de l'erreur quadratique moyenne (pénalise davantage les grands écarts)
- **MAPE** : Erreur en pourcentage (indépendant de l'échelle)

### Comparaison des Modèles de Prévision

Modèle	MAE (MW)	RMSE (MW)	MAPE (%)	MAE_Score	RMSE_Score	MAPE_Score	Score Global
SARIMA	1857.83	2455.52	0.1062	0.0000	0.0000	0.9472	0.3157
Prophet	1108.30	1334.45	0.0690	0.4034	0.4566	0.9657	0.6086
LSTM	322.41	456.87	2.0121	0.8265	0.8139	0.0000	0.5468

TABLE 1 – Comparaison des performances des modèles SARIMA, Prophet et LSTM sur la série temporelle d'énergie.

### Interprétation

- SARIMA : erreurs élevées (MAE et RMSE) malgré un MAPE correct, Score Global faible → modèle moins performant.
- Prophet : erreurs modérées, MAPE très faible (0.07%), Score Global le plus élevé (0.6086) → modèle le plus équilibré et précis.
- LSTM : MAE et RMSE très faibles mais MAPE élevé (2.01%), Score Global moyen → modèle sensible aux variations relatives.

### Justification du choix

Le modèle **Prophet** est choisi comme meilleur modèle car il offre un bon compromis entre précision absolue et relative, avec un Score Global le plus élevé. Il capture efficacement la tendance et la saisonnalité tout en maintenant une erreur relative très faible, ce qui est crucial pour la prévision énergétique.

## 0.7 Dashboard et Visualisation

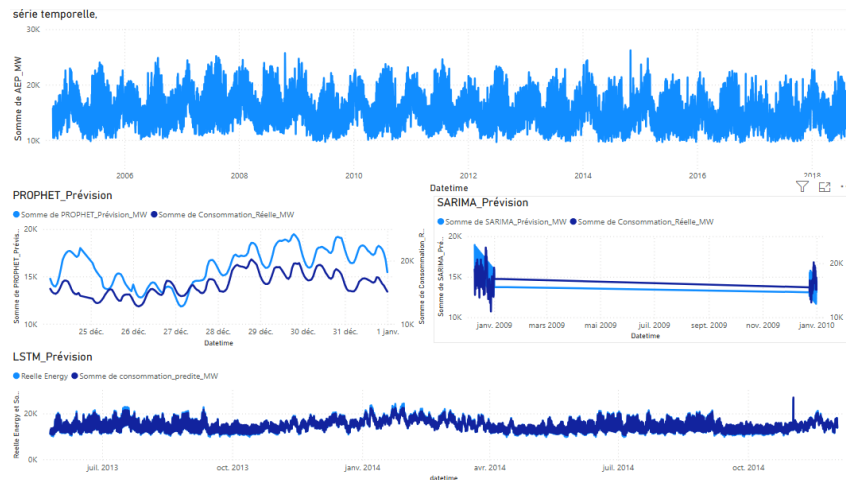


FIGURE 3 – Dashboard

ce tableau de bord présente la série temporelle ainsi que les différentes prévisions de chaque modèle, en comparant les valeurs réelles aux valeurs prédites.

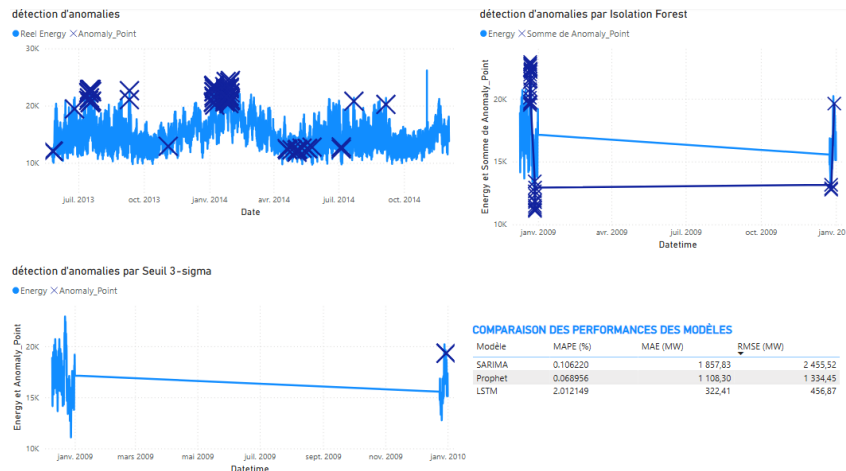


FIGURE 4 – Dashboard

ce tableau de bord illustre la détection d'anomalies à l'aide de deux méthodes, à savoir le SUIL et Isolation Forest.

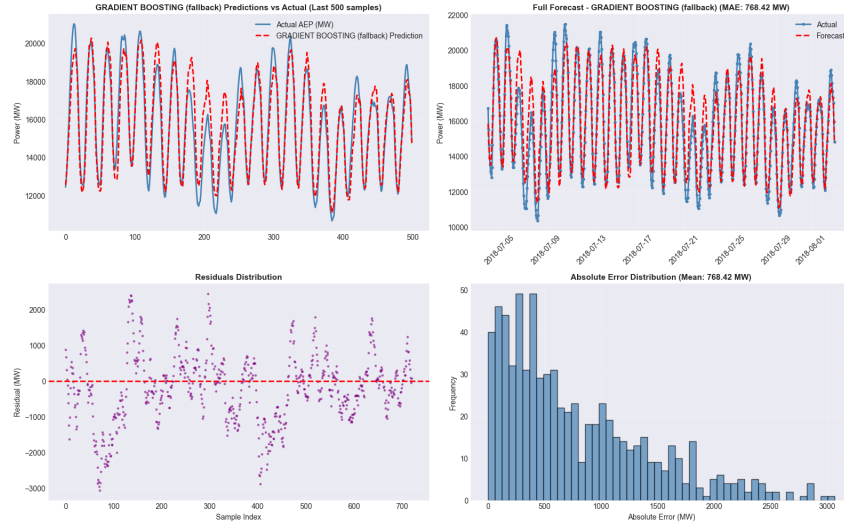


FIGURE 5 – XGboost

ce tableau de bord présente la série temporelle ainsi que les différentes prévisions de modèle XGboost(MAPE=0.05), en comparant les valeurs réelles aux valeurs prédites.

## Conclusion

Ce rapport a présenté une approche comparative entre des méthodes statistiques classiques et des modèles de deep learning pour la prévision des séries temporelles.