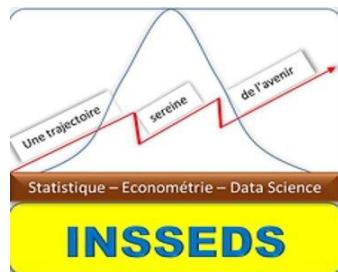


MINISTERE DE L'ENSEIGNEMENT SUPERIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE



REPUBLIQUE DE COTE D'IVOIRE



INSTITUT SUPERIEUR DES  
STATISTIQUES D'ECONOMETRIES ET  
DATASCIENCE

UNION-DISCIPLINE-TRAVAIL

MASTER 2  
STATISTIQUE-ECONOMETRIE-DATA SCIENCE

MINI-PROJET

SERIES TEMPORELLES

ANALYSE DESCRIPTIVE ET PREVISION PAR LISSAGE  
EXPONENTIEL

ANNEE ACADEMIQUE :

2024 -2025

NOM: KABA

PRENOM: MAHAMOUD TOIB

ENSEIGNANT – ENCADREUR  
AKPOSSO DIDIER MARTIAL

INSEEDS



KABA MAHAMOUD TOIB

## AVANT-PROPOS

L'analyse des séries temporelles est un domaine fondamental en statistique et en économétrie, permettant d'examiner l'évolution des phénomènes dans le temps et d'en dégager des tendances utiles pour la prise de décision. Dans un monde où la demande en électricité ne cesse de croître et où l'optimisation de la gestion énergétique devient cruciale, la prévision de la consommation électrique revêt une importance stratégique majeure.

Le présent mini-projet s'inscrit dans cette logique en proposant une analyse descriptive des séries temporelles relatives à la consommation d'électricité, suivie d'une prévision basée sur la méthode du lissage exponentiel de Holt-Winters. Cette approche, largement utilisée en prévision des séries temporelles, permet d'anticiper l'évolution de la demande énergétique en prenant en compte les tendances et les variations saisonnières des données historiques.

Ce travail repose sur l'exploitation de données historiques couvrant la période du 1er octobre 2004 au 1er Aout 2018, contenues dans le fichier "AEP\_hourly". L'objectif est de structurer et d'analyser ces données afin d'en extraire des enseignements clés sur l'évolution de la consommation électrique et d'élaborer des prévisions pour les 24 mois suivants.

Pour atteindre ces objectifs, plusieurs tâches ont été définies :

- Importation et organisation des données** : Cette première étape consistera à charger les données et à structurer la consommation totale d'électricité par mois en utilisant les packages appropriés dans R, notamment *tseries*, *dplyr* et *lubridate*.
- Analyse descriptive des données** : L'objectif ici est d'explorer les tendances générales, les fluctuations saisonnières et les éventuelles anomalies afin de mieux comprendre l'évolution de la consommation électrique.
- Prévision de la consommation pour les 24 prochains mois** : La méthode de Holt-Winters sera employée pour réaliser cette prévision, permettant ainsi d'obtenir des projections précises basées sur les tendances passées et les variations saisonnières.

Ce mini-projet sera entièrement rédigé sous *Microsoft Word* puis converti en *PDF* avant d'être soumis à l'enseignant à l'adresse *insseds@gmail.com* au plus tard le *28 février 2024 à 23h59*.

À travers cette étude, nous espérons mettre en lumière les dynamiques sous-jacentes de la consommation électrique et démontrer l'efficacité des techniques de prévision appliquées aux séries temporelles. Ce travail s'inscrit non seulement dans un cadre académique, mais aussi dans une perspective d'aide à la décision pour la gestion et l'optimisation des ressources énergétiques.

# Table des matières

1. Introduction .....	4
1.1 Contexte du projet .....	4
1.2 Objectifs .....	4
1.3 Présentation des données .....	4
2. Exploration des données .....	5
2.1 Importation et exploration initiale.....	5
2.2 Organisation et nettoyage.....	5
2.3 Détection et traitement des valeurs manquantes .....	6
2.4 Visualisation des données brutes (sans agrégation) .....	6
3.le jeu de données sous forme mensuelle .....	7
3.1 statistique descriptives des données mensuelles.....	7
3.1.1 visualisation graphique.....	7
3.1.2. résumé statistiques.....	7
3.2 Dépendance temporelle .....	9
Autocorrélation simple .....	9
Autocorrélation partielle .....	10
3.3. test de stationnarité .....	11
3.4 Test de normalité .....	12
4. Modélisation et prévision avec Holt-Winters .....	15
4.1 visualisation de la série brutes .....	15
4.2 Décomposition .....	16
4.3 ajustement du modèle de Holt Winter .....	17
4.4 Application du modèle Holt-Winters pour la prévision sur 24 mois .....	17
5. Vérification des résidus .....	19
5.2 ACF .....	20
5.3 test de Ljung-Box .....	21
5.4 test de normalité.....	21
Conclusion .....	22
Tableaux de bord.....	23
annexe .....	24

# 1. Introduction

## 1.1 Contexte du projet

La consommation d'électricité est un enjeu majeur pour les sociétés modernes, notamment en raison des fluctuations de la demande et de la nécessité d'une gestion efficace des ressources énergétiques. Une prévision précise de la consommation permet d'améliorer la planification, d'optimiser les coûts et de garantir un approvisionnement stable. L'analyse des séries temporelles fournit un cadre méthodologique rigoureux pour identifier les tendances et anticiper les variations futures.

## 1.2 Objectifs

L'objectif principal de ce projet est d'analyser et de prévoir la consommation d'électricité à partir de données historiques. Plus précisément, il s'agit de :

- Organiser et structurer les données de consommation électrique par mois.
- Réaliser une analyse descriptive afin d'identifier les tendances et variations saisonnières.
- Mettre en œuvre la méthode de Holt-Winters pour effectuer des prévisions sur les 24 mois à venir.

## 1.3 Présentation des données

Les données utilisées dans ce projet proviennent du fichier *Electric\_Production.csv*, qui contient les enregistrements mensuels de la consommation électrique de octobre 2004 à aout 2018. Ces données seront exploitées à l'aide de divers outils statistiques et logiciels tels que R et ses bibliothèques spécialisées (*tseries*, *dplyr*, *lubridate*). Une attention particulière sera portée à la qualité des données, notamment en identifiant et en traitant les valeurs manquantes ou aberrantes.

## 2. Exploration des données

### 2.1 Importation et exploration initiale

L'importation des données constitue la première étape du projet. Elle permet d'examiner la structure du fichier, de vérifier la présence d'éventuels problèmes de formatage et d'obtenir une vue d'ensemble des observations disponibles. Une exploration initiale des données sera réalisée afin de détecter les valeurs extrêmes, repérer les tendances et observer les variations saisonnières.

UN APERCU DE MES DONNEES EN HEURE-JOUR-MOIS-ANNEES

	AEP_MW
Datetime	
2004-12-31 01:00:00	13478.0
2004-12-31 02:00:00	12865.0
2004-12-31 03:00:00	12577.0
2004-12-31 04:00:00	12517.0
2004-12-31 05:00:00	12670.0
...	...
2018-01-01 20:00:00	21089.0
2018-01-01 21:00:00	20999.0
2018-01-01 22:00:00	20820.0
2018-01-01 23:00:00	20415.0
2018-01-02 00:00:00	19993.0

MES DONNEES TRANSFORMEES EN ANNEES-MOIS

month	
2004-10-01	10349073.0
2004-11-01	10677917.0
2004-12-01	12452864.0
2005-01-01	12735114.0
2005-02-01	11085745.0
2005-03-01	11847502.0
2005-04-01	10089312.0
2005-05-01	10181694.0
2005-06-01	11700240.0
2005-07-01	12546722.0
2005-08-01	12834811.0
2005-09-01	11049942.0
2005-10-01	10685367.0
2005-11-01	10940315.0
2005-12-01	13056150.0
2006-01-01	12208918.0

### INTERPRETATION

Notre jeu de données contient **121273 observations** et **deux variables** à savoir **datetime** et consommation

L'image montre la transformation des données de consommation d'électricité initialement enregistrées au format **-HEURE** en un format **année-mois**, permettant ainsi d'agréger les valeurs mensuelles.

- **À gauche** : Les données brutes avec une consommation horaire.
- **À droite** : La consommation totale pour chaque mois après agrégation.

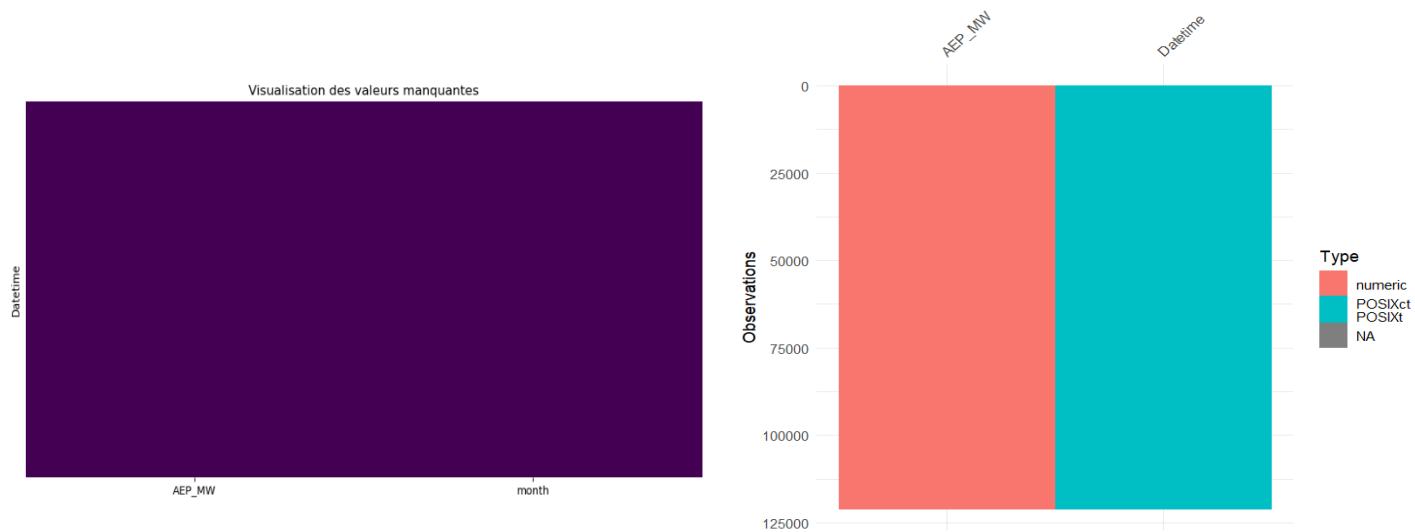
Cette transformation facilite l'analyse des tendances temporelles et la modélisation des séries chronologiques.

### 2.2 Organisation et nettoyage

Une fois les données importées, il est essentiel de les organiser et de les nettoyer pour garantir une analyse fiable. Cette étape inclut la conversion des formats de date, la gestion des doublons et la standardisation des variables. Un nettoyage rigoureux permet d'éviter les erreurs et d'assurer la cohérence des résultats obtenus.

## 2.3 Détection et traitement des valeurs manquantes

Les valeurs manquantes sont un problème fréquent dans l'analyse des séries temporelles. Leur identification et leur traitement sont essentiels pour éviter les biais et assurer la qualité des prévisions. Différentes stratégies seront explorées, comme l'imputation par interpolation ou l'utilisation de modèles spécifiques permettant de reconstruire les données manquantes en tenant compte des tendances passées. **Ici il n'y a pas de valeurs manquantes.**



## 2.4 Visualisation des données brutes (sans agrégation)

Image 1

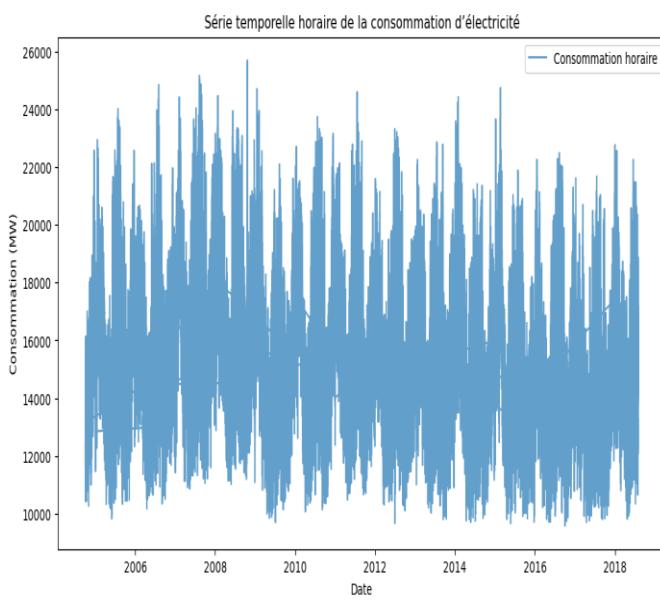
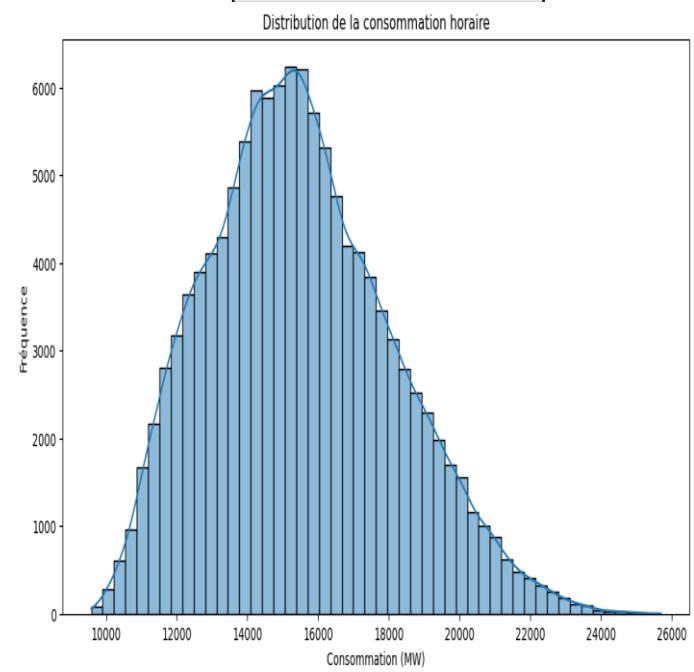


Image 2



L'image 1 représente l'évolution de la consommation horaires d'électricité . Voici une explication concise :

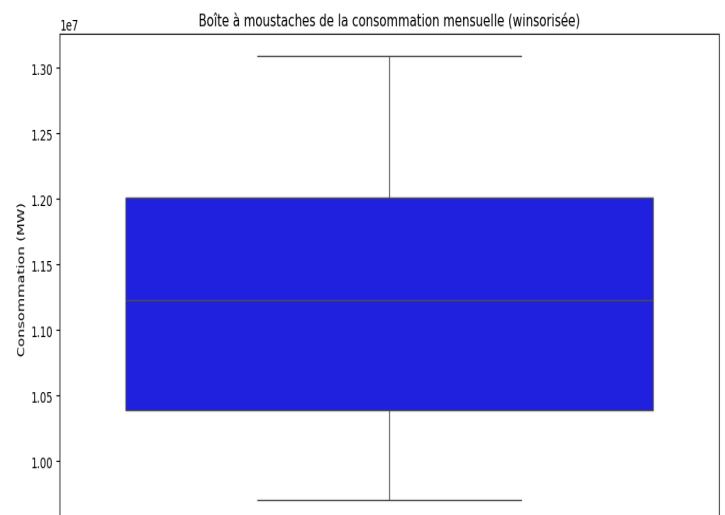
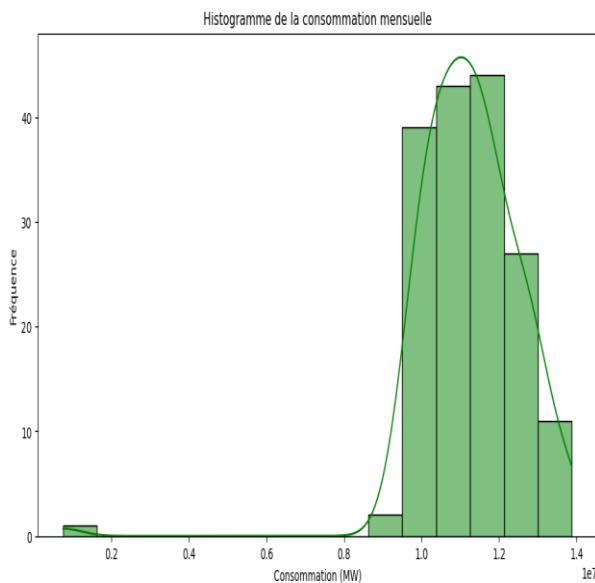
1. Tendance Générale : La consommation d'électricité montre une constance régulière depuis 2004 jusqu'à environ 2018, ce qui peut refléter un comportement neutre de la demande énergétique.

Sur l'image 2 nous voyons que la consommation honoraire semble suivre une loi normale

### 3.le jeu de données sous forme mensuelle

#### 3.1 statistique descriptives des données mensuelles

##### 3.1.1 visualisation graphique



##### 3.1.2. résumé statistiques

Les statistiques descriptives permettent de résumer les données à travers plusieurs indicateurs :

- Moyenne : valeur centrale des données.
- Médiane : valeur qui sépare les données en deux groupes égaux.
- Variance et écart-type : mesure de la dispersion des valeurs autour de la moyenne.
- Skewness (asymétrie) : indique si la distribution est symétrique ou biaisée vers la gauche/droite.
- Kurtosis (aplatissement) : montre si la distribution a des queues plus épaisses ou plus fines que la normale.

Indicateurs	Valeurs
<b>Moyenne</b>	11 289 308.78
<b>Médiane</b>	11 228 646.00
<b>Mode</b>	9 703 509.00
<b>Écart-type</b>	1 024 312.61
<b>Variance</b>	1 049 216 326 276.29
<b>IQR</b> <b>(Intervalle Interquartile)</b>	1 617 217.50
<b>Minimum</b>	9 703 509.00
<b>Maximum</b>	13 090 587.00
<b>Asymétrie (Skewness)</b>	0.21
<b>Kurtosis</b>	-1.05

## Interprétation détaillée des statistiques

### 1. Indicateurs de tendance centrale

- Moyenne (11 289 308.78 MW) : La consommation moyenne d'électricité sur la période analysée est d'environ 11,3 millions de MW. Cela donne une idée de la consommation typique, mais elle peut être influencée par des valeurs extrêmes.
  - Médiane (11 228 646 MW) : La valeur médiane, qui est le point central des données, est légèrement inférieure à la moyenne. Cela signifie que la moitié des mois ont une consommation inférieure à 11,2 millions de MW et l'autre moitié une consommation supérieure.
  - Mode (9 703 509 MW) : Le mode représente la valeur la plus fréquente dans les données. Ici, 9,7 millions de MW est la consommation la plus souvent observée au cours de la période étudiée.
- ❖ Interprétation globale : Comme la moyenne et la médiane sont proches, la distribution des consommations est assez équilibrée, sans forte asymétrie.

### 2. Indicateurs de dispersion

- Écart-type (1 024 312.61 MW) : Cette valeur indique que la consommation varie en moyenne d'environ 1 million de MW autour de la moyenne. Une valeur élevée montre une forte variabilité dans les données.
- Variance (1 049 216 326 276.29 MW<sup>2</sup>) : Comme l'écart-type, la variance reflète la dispersion des données, mais avec une unité différente (MW<sup>2</sup>). Une variance aussi grande confirme une importante variabilité des consommations.
- IQR (1 617 217.50 MW) : Cet intervalle interquartile indique que 50 % des valeurs sont comprises entre Q1 et Q3, c'est-à-dire une variation de 1,6 million de MW. Cela montre une dispersion modérée des valeurs centrales.
- Minimum (9 703 509 MW) & Maximum (13 090 587 MW) : La consommation minimale et maximale montrent que la consommation a varié d'environ 3,4 millions de MW entre les mois les plus bas et les plus hauts.

❖ Interprétation globale : Il y a une grande variabilité dans la consommation d'électricité, ce qui pourrait être dû à des fluctuations saisonnières ou économiques.

### 3. Indicateurs de forme de distribution

- Asymétrie (Skewness = 0.21) : Une asymétrie proche de 0 signifie que la distribution est quasiment symétrique. Il n'y a donc pas de biais significatif vers des valeurs extrêmes à droite (forte consommation) ou à gauche (faible consommation).
- Kurtosis (-1.05) : Une kurtosis négative signifie que la distribution est plus aplatie que la normale, ce qui signifie qu'il y a moins de valeurs extrêmes que dans une distribution normale standard.

► Interprétation globale : La distribution des consommations est équilibrée et peu influencée par des valeurs extrêmes. Cela suggère que les variations sont relativement stables au fil du temps.

## Conclusion générale

■ L'analyse montre que la consommation mensuelle d'électricité est globalement stable, avec une variabilité modérée et une distribution équilibrée. Cependant, les fortes variations (évidentes dans l'écart-type et la variance) pourraient être liées à des facteurs saisonniers ou économiques, ce qui justifie une analyse plus approfondie avec des modèles de prévision.

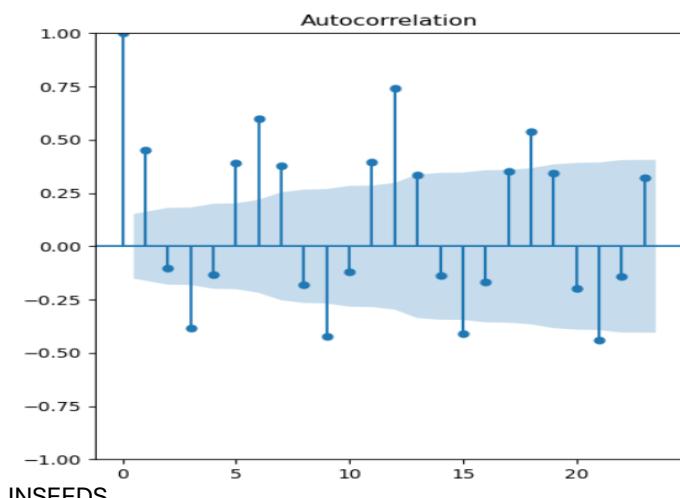
## 3.2 Dépendance temporelle

La dépendance temporelle dans une série chronologique signifie que les observations successives ne sont pas indépendantes, mais qu'elles sont corrélées entre elles. Cela implique que les valeurs passées influencent les valeurs futures. L'analyse de cette dépendance se fait souvent à l'aide des fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF).

### Autocorrélation simple

L'autocorrélation simple, ou fonction d'autocorrélation (ACF), mesure la corrélation entre une série temporelle et elle-même à différents décalages (lags). Elle permet d'identifier la structure de dépendance dans le temps.

Si une série présente une forte autocorrélation pour des lags proches de zéro, cela signifie que ses valeurs passées ont une influence significative sur ses valeurs futures.



### Interprétation du graphique d'autocorrélation (ACF)

Ce graphique représente la **fondation d'autocorrélation (ACF)** d'une série temporelle, permettant d'analyser la dépendance entre une observation et ses valeurs passées (retards ou lags).

#### 1. Lecture générale du graphique

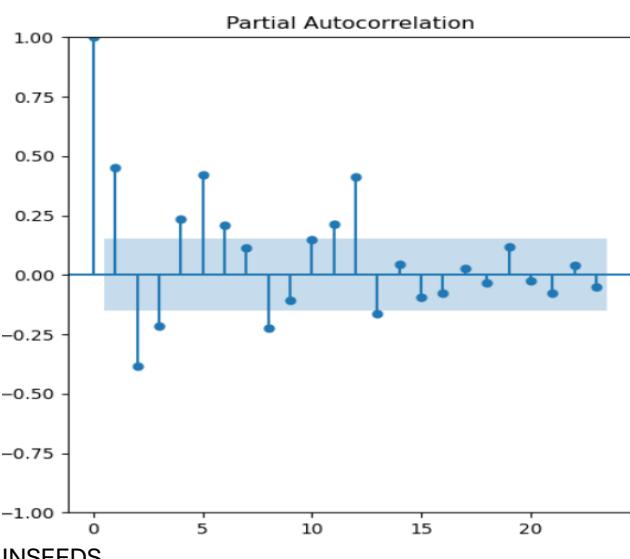
KABA MAHAMOUD TOIB

- **L'axe horizontal (x) :** Représente les décalages (lags), c'est-à-dire les périodes précédentes par rapport à la valeur actuelle.
- **L'axe vertical (y) :** Représente le coefficient d'autocorrélation, variant entre -1 et 1.
- **Les barres verticales :** Indiquent la force de la corrélation pour chaque décalage.
- **La zone bleue :** Correspond à l'intervalle de confiance. Les valeurs en dehors de cette zone sont statistiquement significatives.

## 2. Interprétation des résultats

1. **Valeur à lag 0 :** Toujours égale à 1, car une série est parfaitement corrélée avec elle-même.
2. **Présence de pics significatifs à certains lags :**
  - ✓ Des lags comme 5, 10 et 15 montrent des pics au-dessus de l'intervalle de confiance.
  - ✓ Cela signifie que la série présente une **corrélation périodique**, suggérant une certaine répétition dans les données (potentiel effet saisonnier).
3. **Diminution progressive de l'autocorrélation :**
  - ✓ L'autocorrélation diminue avec l'augmentation du lag, ce qui est typique d'une série stationnaire.
  - ✓ La tendance de certaines valeurs à rester significatives pourrait suggérer une mémoire persistante.
4. **Valeurs négatives à certains lags :**
  - ✓ Indique que la série a des cycles avec des périodes de hausse suivies de baisses après certains intervalles de temps.

## Autocorrélation partielle



L'autocorrélation partielle (PACF) mesure la corrélation entre une série temporelle et elle-même à différents décalages (lags), en éliminant l'effet des lags intermédiaires. Elle permet d'identifier l'ordre d'un modèle AR (AutoRégressif).

Un pic significatif au premier lag suivi d'une chute rapide suggère un processus AR(1), tandis que plusieurs lags significatifs indiquent un processus AR( $p$ ).

## Interprétation du graphique de l'Autocorrélation Partielle (PACF)

Le **PACF (Partial Autocorrelation Function)** permet de mesurer la corrélation d'une série temporelle avec ses

retards (lags), après avoir contrôlé l'influence des autres lags intermédiaires. Ce graphique est crucial pour identifier l'ordre du modèle AR (AutoRegressif) dans les modèles ARIMA.

## 1. Lecture du graphique

- **Axe horizontal (x)** : Indique les lags (retards) de la série.
- **Axe vertical (y)** : Montre les coefficients d'autocorrélation partielle.
- **Les barres verticales** : Représentent la corrélation partielle de chaque lag avec la valeur actuelle, une fois l'effet des lags intermédiaires retiré.
- **La zone bleue** : Correspond à l'intervalle de confiance. Les valeurs significatives en dehors de cette zone indiquent des dépendances importantes.

## 2. Interprétation

1. **Valeur à lag 0** : Toujours égale à 1, car la série est parfaitement corrélée avec elle-même.
2. **Premier lag (1) significatif** :
  - ✓ L'autocorrélation partielle à **lag 1** est élevée et significative.
  - ✓ Cela suggère une dépendance immédiate entre une valeur et celle qui la précède.
  - ✓ Une telle structure indique qu'un **modèle AR(1) pourrait être approprié**.
3. **Quelques autres lags significatifs** :
  - ✓ Des pics sporadiques apparaissent à certains lags (ex : lag 5, 10, 15).
  - ✓ Cela peut indiquer une **saisonnalité ou une dépendance à long terme**.
4. **Diminution rapide des valeurs après quelques lags** :
  - ✓ Cela montre que l'effet des observations précédentes disparaît assez vite.
  - ✓ Si seule la première valeur est significative, un modèle **AR(1) ou AR(2)** est souvent suffisant.
5. **Valeurs proches de zéro après un certain lag** :
  - ✓ À partir d'un certain lag (~10-12), les valeurs deviennent non significatives (dans la zone bleue).
  - ✓ Cela signifie que les observations plus anciennes n'ont plus d'effet direct sur la valeur actuelle après correction des autres lags.

## 3.3. test de stationnarité

Le test ADF (Augmented Dickey-Fuller) est utilisé pour vérifier si une série temporelle est **stationnaire** ou non. Il teste l'hypothèse nulle  $H_0$  selon laquelle la série a une racine unitaire, ce qui signifie qu'elle **n'est pas stationnaire**.

### 1. Rappel des hypothèses du test

- **$H_0$  (Hypothèse nulle)** : La série **n'est pas stationnaire** (elle a une racine unitaire).
- **$H_1$  (Hypothèse alternative)** : La série **est stationnaire** (elle n'a pas de racine unitaire).

### 2. Analyse de la p-value obtenue

- **p-value = 0.6113**

- Cette valeur est **bien supérieure** au seuil de 5 % (0.05).
  - ◆ **Règle de décision :**
    - Si **p-value < 0.05**, on **rejette H<sub>0</sub>** → La série est **stationnaire**.
    - Si **p-value > 0.05**, on **ne rejette pas H<sub>0</sub>** → La série **n'est pas stationnaire**.
- 

### 3. Conclusion

Puisque **0.6113 > 0.05**, nous **ne pouvons pas rejeter H<sub>0</sub>**.

→ La série **n'est pas stationnaire**.

## 3.4 Test de normalité

Les tests de **Shapiro-Wilk** et **Kolmogorov-Smirnov** sont utilisés pour vérifier si une variable suit une distribution normale.

---

### 1. Test de Shapiro-Wilk

#### Interprétation des résultats

- **p-value = 1.5163e-05** (soit environ **0.000015**)
  - **Hypothèses du test :**
    - $H_0$  : Les données suivent une distribution normale.
    - $H_1$  : Les données ne suivent pas une distribution normale.
  - **Décision :**
    - Si **p-value < 0.05**, on rejette  $H_0$  → les données **ne suivent pas** une distribution normale.
    - Si **p-value > 0.05**, on ne rejette pas  $H_0$  → les données **sont normales**.
  - ◆ **Conclusion :** Ici, **p-value < 0.05**, donc on rejette l'hypothèse nulle. Cela signifie que les données **ne suivent pas** une distribution normale.
- 

### 2. Test de Kolmogorov-Smirnov

#### Interprétation des résultats

- **p-value = 0.4368**
- **Hypothèses du test :**
  - H<sub>0H\_0H0</sub> : Les données suivent une distribution normale.
  - H<sub>1H\_1H1</sub> : Les données ne suivent pas une distribution normale.
- **Décision :**
  - Si **p-value < 0.05**, on rejette H<sub>0H\_0H0</sub> → les données **ne suivent pas** une distribution normale.
  - Si **p-value > 0.05**, on ne rejette pas H<sub>0H\_0H0</sub> → les données **sont normales**.
- ◆ **Conclusion :** Ici, **p-value > 0.05**, donc on **ne rejette pas H<sub>0H\_0H0</sub>**. Cela signifie que le test de Kolmogorov-Smirnov **ne détecte pas de déviation significative** par rapport à la normalité.

Autres test

## 1 .Test d'Anderson-Darling

- **Hypothèses du test :**
  - ✓ **H<sub>0H\_0H0</sub>** (Hypothèse nulle) : Les données suivent une distribution normale.
  - ✓ **H<sub>1H\_1H1</sub>** (Hypothèse alternative) : Les données ne suivent pas une distribution normale.
- **Interprétation des résultats :**
  - ✓ La **statistique de test** est **1.7219**.
  - ✓ Cette statistique est **supérieure** aux valeurs critiques à **tous les niveaux de signification** (15%, 10%, 5%, 2.5% et 1%).
  - ✓ **Conclusion :** On rejette H<sub>0H\_0H0</sub> à tous ces niveaux, ce qui signifie que **les données ne suivent pas une distribution normale**.

## 2. test de Jarque-Bera

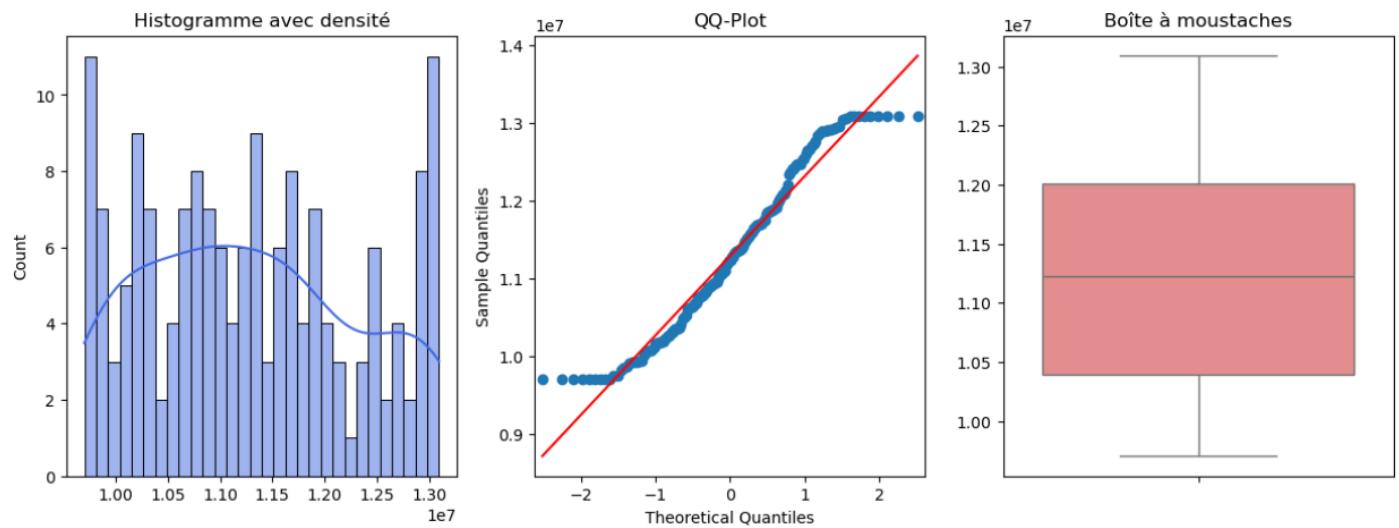
- **Hypothèses du test :**
  - ✓ **H<sub>0H\_0H0</sub>** : Les données suivent une distribution normale.
  - ✓ **H<sub>1H\_1H1</sub>** : Les données ne suivent pas une distribution normale.
- **Interprétation des résultats :**
  - ✓ La **statistique de test** est **8.8725** avec une **p-value = 0.0118**.

- ✓ La p-value est inférieure à 0.05, ce qui signifie que l'on rejette  $H_0$  au seuil de 5%.
- ✓ Conclusion : On conclut que les données ne suivent pas une distribution normale.

## Conclusion générale

- ✓ Les deux tests (Anderson-Darling et Jarque-Bera) montrent que les données ne suivent pas une loi normale.

Voyons



Q-Q plot montre si tes données suivent une distribution normale. Voici comment l'interpréter :

1. **Alignement sur la ligne rouge :**
  - ✓ Si les points sont bien alignés sur la ligne, les données suivent une distribution normale.
  - ✓ Ici, les points s'écartent de la ligne aux extrémités, suggérant des queues plus lourdes ou une asymétrie.
2. **Déviation aux extrémités :**
  - ✓ Une déviation aux deux bouts indique une distribution avec des valeurs extrêmes (kurtosis élevée).
  - ✓ Ici, la partie inférieure est sous la ligne et la supérieure au-dessus, ce qui peut indiquer une asymétrie positive.
3. **Complément aux tests de normalité :**
  - ✓ Le test de Shapiro-Wilk donne une p-value très faible ( $< 0.05$ ), donc on rejette l'hypothèse de normalité.
  - ✓ Le test KS a une p-value plus élevée ( $> 0.05$ ), ce qui ne permet pas de rejeter directement la normalité.
  - ✓ Le test de Jarque-Bera donne une p-value très faible ( $< 0.05$ ), donc on rejette l'hypothèse de normalité.

- ✓ Le test de Anderson donne une p-value très faible ( $< 0.05$ ), donc on rejette l'hypothèse de normalité

**Conclusion :** mes données ne semblent pas suivre une distribution normale strictement.

## 4. Modélisation et prévision avec Holt-Winters

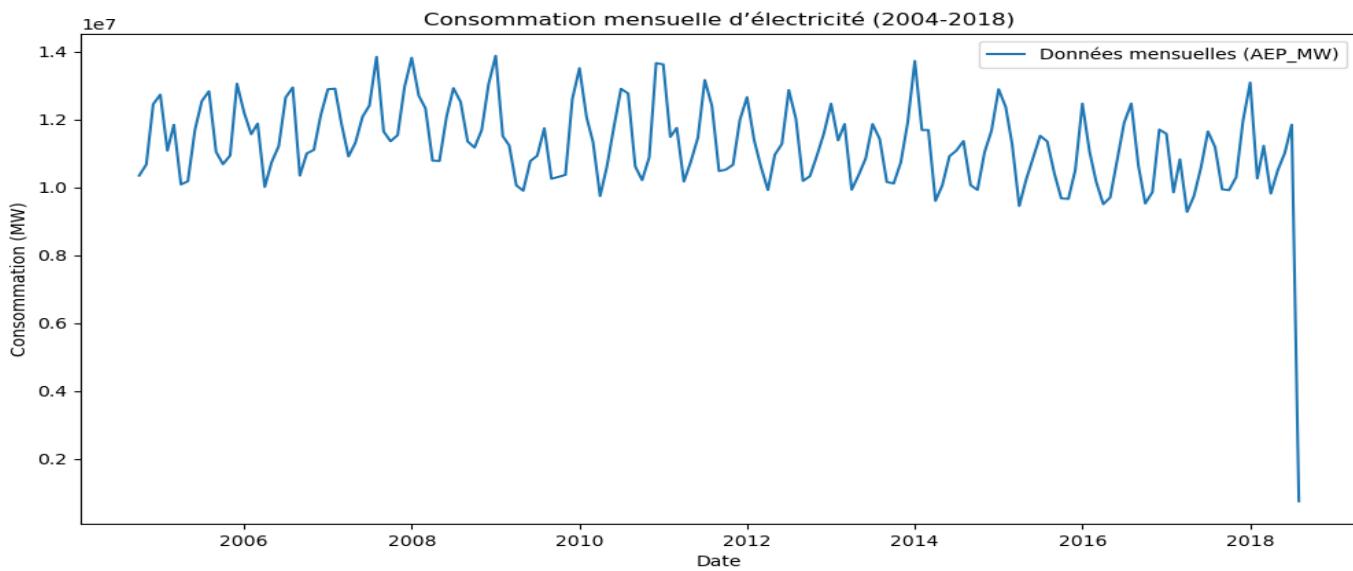
Le modèle de Holt-Winters est une méthode de lissage exponentiel qui permet de modéliser et de prévoir des séries temporelles avec tendance et saisonnalité. Il existe trois variantes du modèle :

1. **Holt-Winters simple** (sans saisonnalité) : utilisé lorsque les données présentent une tendance mais pas de saisonnalité.
2. **Holt-Winters additif** : adapté aux séries avec saisonnalité additive (l'amplitude des variations saisonnières reste constante).
3. **Holt-Winters multiplicatif** : adapté aux séries avec saisonnalité multiplicative (l'amplitude des variations saisonnières augmente avec le niveau de la série).

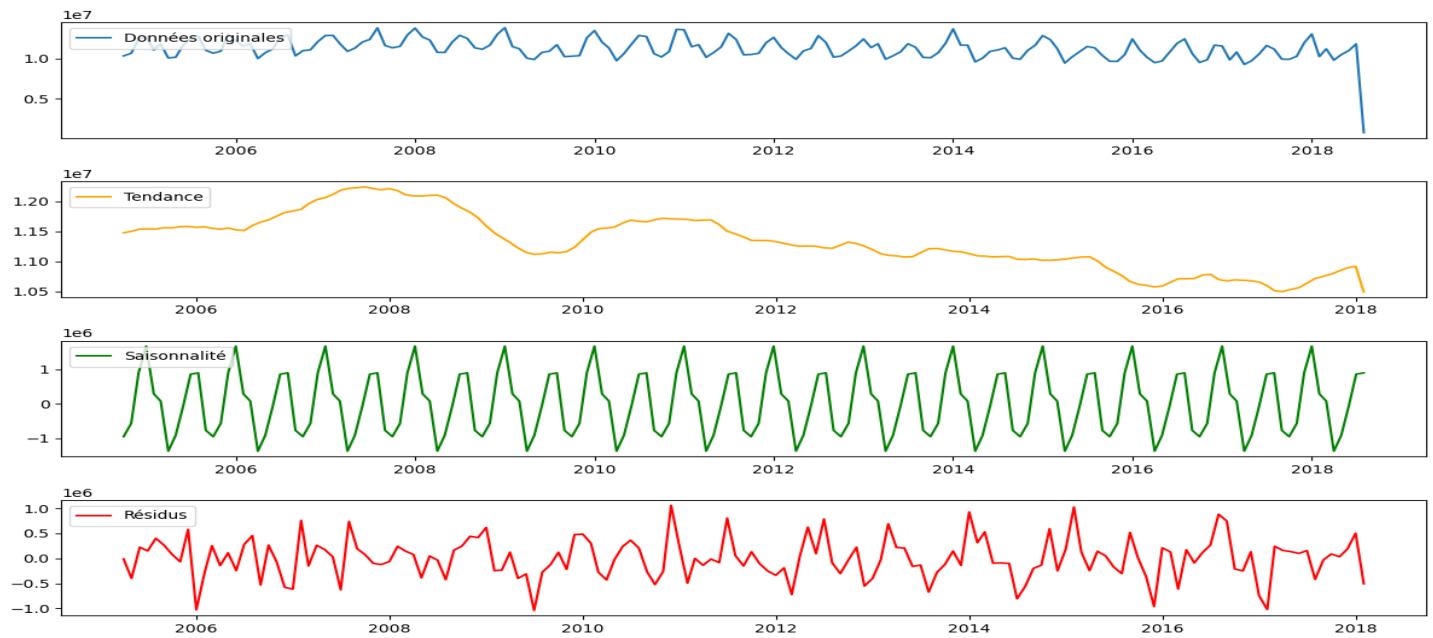
Dans notre cas, étant donné que les données sont mensuelles et présentent une tendance ainsi qu'une saisonnalité,. Ce modèle ajuste trois composantes :

- **Le niveau ( $L_t$ )** : la valeur de base de la série.
- **La tendance ( $T_t$ )** : la direction générale de la série.
- **La saisonnalité ( $S_t$ )** : les fluctuations périodiques.

### 4.1 visualisation de la série brutes



## 4.2 Décomposition



### 1. Données originales (Première partie du graphique) :

- C'est la série temporelle brute de la consommation électrique (2004-2018), avant toute analyse.
- Tu peux y observer des motifs saisonniers et des variations annuelles, ainsi qu'une forte baisse vers la fin.
- Interprétation : Cette série brute contient à la fois des tendances, des cycles saisonniers et des irrégularités.

### 2. Tendance (Deuxième partie du graphique) :

- La courbe de tendance représente la variation à long terme des données.
- On observe une augmentation entre 2005 et 2009, suivie d'une diminution progressive vers 2018.
- Interprétation : La tendance montre une période de croissance soutenue suivie d'un déclin. Cela peut être lié à des facteurs externes comme une baisse de la demande d'électricité ou des événements économiques.

### 3. Saisonnalité (Troisième partie du graphique) :

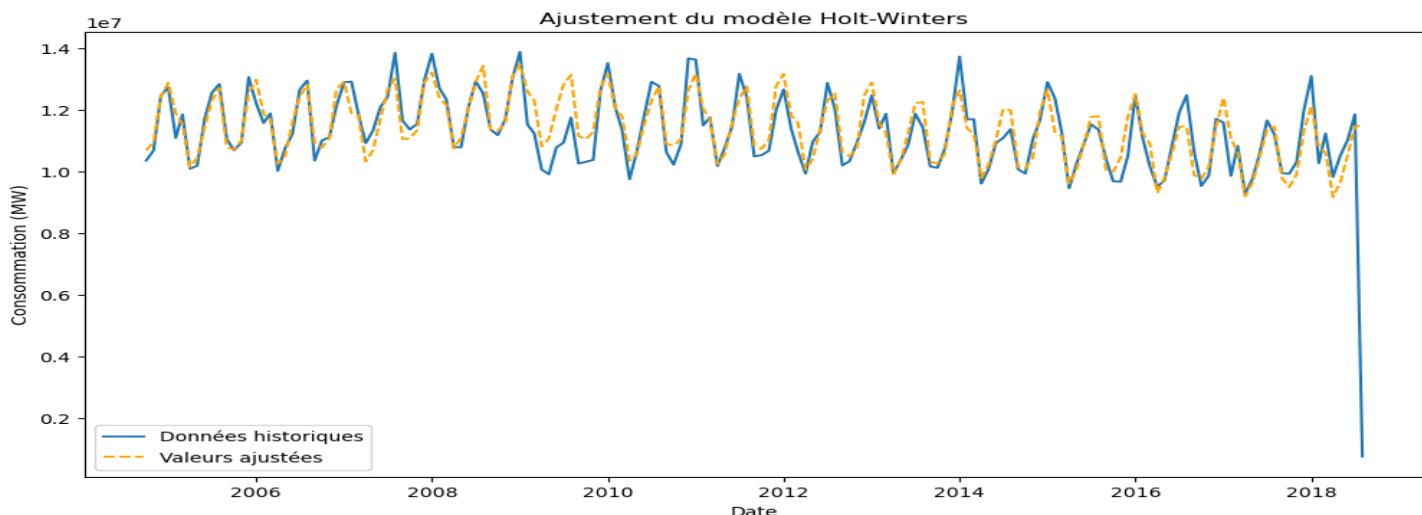
- La composante saisonnière capture les variations périodiques régulières qui se répètent chaque année.
- On voit un motif cyclique stable, indiquant que la consommation suit un comportement récurrent d'une année à l'autre.
- Interprétation : Cela pourrait refléter un schéma saisonnier, comme une hausse en hiver (chauffage) et une baisse en été.

### 4. Résidus (Quatrième partie du graphique) :

- Les résidus sont ce qui reste après avoir retiré la tendance et la saisonnalité des données.

- Ils devraient être imprévisibles et sans motif si le modèle est bon.
- Interprétation : Si les résidus montrent des motifs ou une corrélation dans le temps, cela signifie que des composantes importantes n'ont pas été capturées par la décomposition ou le modèle.

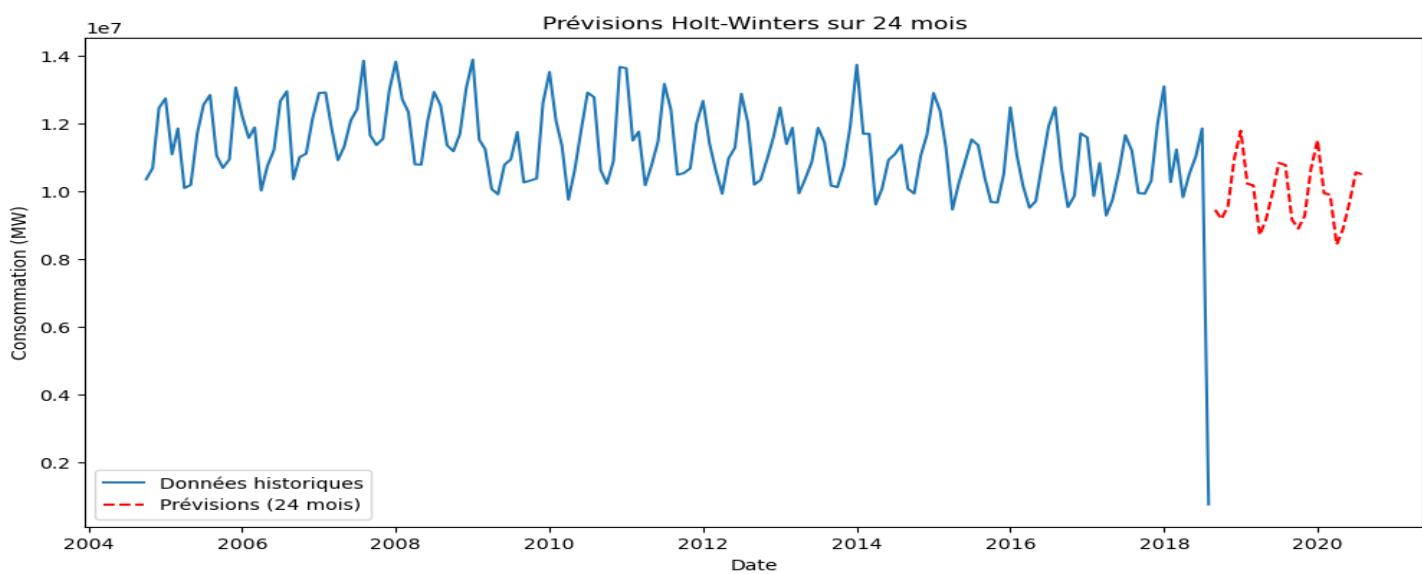
#### 4.3 ajustement du modèle de Holt Winter



Les valeurs ajustées (ligne orange pointillée) suivent de très près les données historiques (ligne bleue), ce qui indique que le modèle capture bien les composantes de tendance et de saisonnalité de la série temporelle. Cela suggère que le modèle explique correctement les variations globales.

Cependant, pour confirmer sa qualité, il faut analyser les résidus (graphiquement, tests de bruit blanc, ACF, etc.). Cela permettra de vérifier si le modèle est vraiment fiable pour des prévisions futures.

#### 4.4 Application du modèle Holt-Winters pour la prévision sur 24 mois



## 1. Données historiques (ligne bleue) :

- Les données historiques montrent un **motif saisonnier stable** avec des fluctuations régulières d'année en année. Cela indique une variation périodique bien définie dans la consommation d'électricité.
- Une chute marquée en 2018 est visible, ce qui pourrait refléter un événement spécifique ou une anomalie dans les données.

## 2. Prévisions (ligne rouge pointillée) :

- Les prévisions suivent la **tendance saisonnière observée** dans les données historiques, mais avec une légère réduction de l'amplitude des fluctuations. Cela est typique des modèles Holt-Winters qui tendent à lisser les variations extrêmes.
- La continuité entre les données historiques et les prévisions suggère que le modèle capture bien les motifs présents dans la série temporelle.

## 3. Observations importantes :

- **Validité des prévisions** : Le modèle semble pertinent pour prolonger les données historiques sur 24 mois, car il maintient les motifs saisonniers identifiés dans les données précédentes.
- **Changement d'amplitude** : L'amplitude légèrement réduite des prévisions peut indiquer que le modèle Holt-Winters applique un effet de lissage, ce qui peut être utile pour éviter de sur-réagir à des anomalies.

Date	Prévisions (MW)
V 01/09/2018	9.445036e+06
01/10/2018	9.174746e+06
01/11/2018	9.555486e+06
01/12/2018	1.095898e+07
01/01/2019	1.178023e+07
01/02/2019	1.021866e+07
01/03/2019	1.017121e+07
01/04/2019	8.703317e+06
01/05/2019	9.168909e+06
01/06/2019	9.952534e+06
01/07/2019	1.083338e+07
01/08/2019	1.077157e+07
01/09/2019	9.163091e+06
01/10/2019	8.892801e+06
01/11/2019	9.273541e+06
01/12/2019	1.067704e+07
01/01/2020	1.149828e+07
01/02/2020	9.936712e+06
01/03/2020	9.889267e+06
01/04/2020	8.421372e+06

01/05/2020	8.886964e+06
01/06/2020	9.670589e+06
01/07/2020	1.055144e+07
01/08/2020	1.048962e+07

Les prévisions générées sur 24 mois montrent des projections de consommation mensuelle d'électricité qui suivent les motifs saisonniers et la tendance détectés dans les données historiques. Voici une interprétation concise des valeurs et leur pertinence :

### 1. Stabilité saisonnière dans les prévisions

- Les prévisions montrent un comportement cyclique cohérent avec les années précédentes, reflétant une saisonnalité régulière :
  - **Hauts** en hiver (par exemple, janvier et décembre), probablement dus à une demande accrue pour le chauffage.
  - **Bas** au printemps et en été (par exemple, avril ou août), lorsque la consommation diminue.

### 2. Amplitude légèrement modérée

- Les valeurs prévues montrent une amplitude un peu plus réduite que les données historiques. Cela peut être lié à l'effet de lissage de Holt-Winters, qui tend à limiter les fluctuations extrêmes.
- Cela peut aussi refléter un modèle conservateur qui minimise l'impact des anomalies.

### 3. Confiance dans les projections

- L'ajustement entre les données historiques et le modèle est bon, ce qui inspire confiance dans ces prévisions. Cependant, des événements inattendus (comme des facteurs économiques ou climatiques) pourraient influencer la fiabilité.

En résumé, ces prévisions permettent de planifier avec une certaine assurance, car elles maintiennent les tendances saisonnières et les variations annuelles des données passées.

### Prochaines étapes : Vérification des résidus

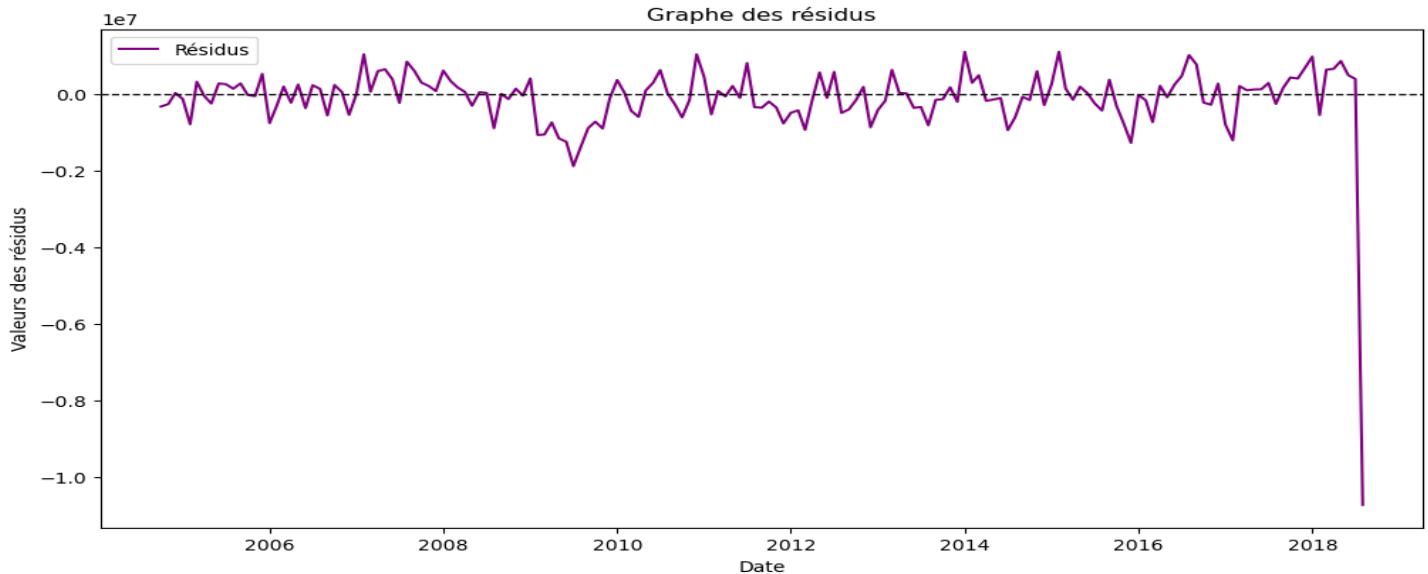
Pour valider la qualité du modèle, il est nécessaire de vérifier si les résidus (différence entre données et ajustements) forment un **bruit blanc gaussien**. Si c'est le cas :

1. **Le modèle est bon** pour capturer les dynamiques sous-jacentes.
2. Les prévisions peuvent être jugées fiables.

## 5. Vérification des résidus

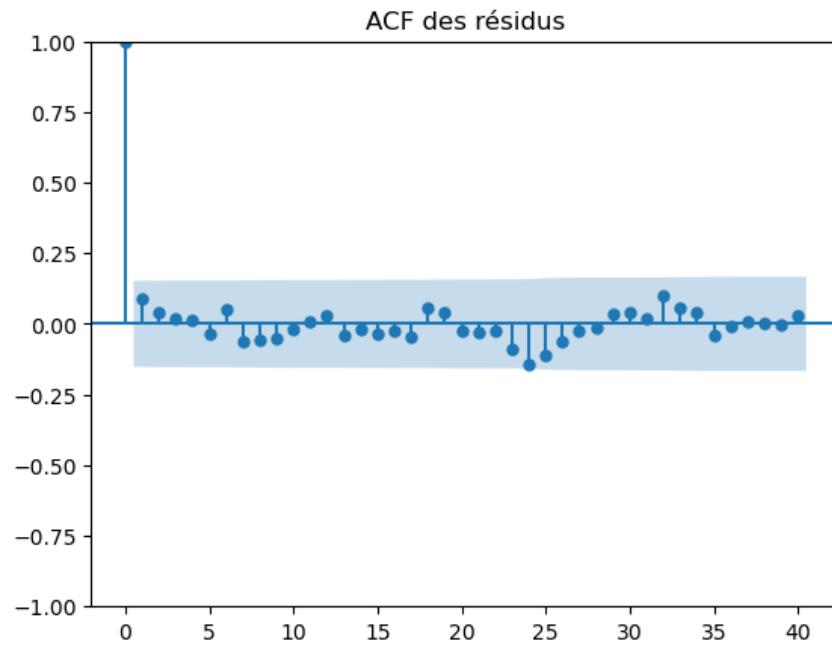
Pour valider la qualité du modèle, il est nécessaire de vérifier si les résidus

### 5.1 graphe des résidus



Le graphe des résidus montre une fluctuation autour de zéro pendant la majorité de la période (2004-2018), ce qui est un bon signe d'un modèle qui capture bien les tendances et la saisonnalité. Cependant, on observe une anomalie marquée avec une chute significative autour de 2018. Cela pourrait indiquer soit un événement exceptionnel non expliqué par le modèle, soit une faiblesse dans la capacité du modèle à gérer cette variation spécifique. Pour confirmer, il est nécessaire d'examiner l'autocorrélation des résidus et leur normalité.

## 5.2 ACF



Le graphe ACF des résidus montre l'autocorrélation à différents décalages temporels (lags). Voici une interprétation concise :

- Majorité des lags dans l'intervalle de confiance (zone bleue) :**
  - Les barres d'autocorrélation restent principalement dans l'intervalle de confiance, ce qui suggère que les résidus sont **faiblement corrélés**. Cela est un bon indicateur que le modèle capture bien les motifs dans les données.
- Lag 0 avec autocorrélation élevée (1.0) :**

➤ Cela est normal pour un ACF, car à lag 0, les résidus sont parfaitement corrélés avec eux-mêmes.

### 3. Léger dépassement à certains lags :

➤ Si des pics significatifs dépassent l'intervalle de confiance à certains lags, cela pourrait indiquer que des dépendances subsistent. Cela demande une vérification supplémentaire.

### Conclusion :

- Les résidus semblent proches d'un **bruit blanc**, ce qui signifie qu'ils sont en grande partie aléatoires.
- Cependant, pour une validation complète du modèle, il faudra effectuer le **test de Ljung-Box** pour s'assurer que les résidus ne montrent pas de dépendances significatives, et des tests de normalité pour vérifier s'ils sont gaussiens.

## 5.3 test de Ljung-Box

Le résultat du test de Ljung-Box montre les statistiques suivantes :

- **lb\_stat (statistique du test)** : 4.213256
- **lb\_pvalue (valeur p)** : 0.937214

### Interprétation :

#### 1. Valeur p élevée (**0.937214 > 0.05**) :

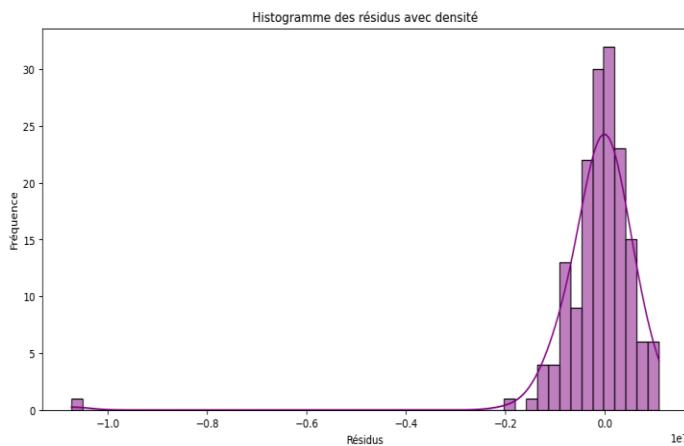
- Une valeur p supérieure à 0.05 indique qu'il n'y a pas de corrélation significative entre les résidus à différents lags. En d'autres termes, les résidus ne présentent pas de dépendances temporelles mesurables.
- Cela suggère que les résidus ressemblent bien à un **bruit blanc**, ce qui est un bon indicateur de la qualité du modèle.

#### 2. Conclusion pour le modèle Holt-Winters :

- Ces résultats signifient que le modèle capture efficacement les motifs des données (tendance et saisonnalité), laissant des résidus aléatoires. Cela renforce la confiance dans les prévisions générées.

Cependant, pour confirmer que les résidus sont également un **bruit blanc gaussien**, il est nécessaire de vérifier leur normalité.

## 5.4 test de normalité



Voici l'interprétation des résultats des tests de normalité :

#### 1. Test de Shapiro-Wilk : p-value = **1.8628918304848945e-21**

- La **p-value** est extrêmement faible (très proche de 0).
- Cela signifie que l'hypothèse nulle du test, selon laquelle les résidus suivent une distribution normale, est rejetée.

- Les résidus ne suivent pas une distribution normale.

## 2. Test de D'Agostino : p-value = 4.377666412042315e-61

- Là encore, la **p-value** est extrêmement faible.
- Le test confirme que les résidus **ne sont pas gaussiens**, c'est-à-dire qu'ils ne suivent pas une distribution en cloche typique de la normalité.

### Conclusion

- Les résidus ne satisfont pas le critère d'être un **bruit blanc gaussien**, car ils ne suivent pas une distribution normale.
- Cela n'implique pas nécessairement que le modèle est invalide, mais cela peut indiquer qu'il reste des motifs inexpliqués ou des anomalies dans les données que le modèle Holt-Winters n'a pas capturés.

## Conclusion

L'analyse de la consommation totale d'électricité a permis d'explorer les tendances et la saisonnalité des données historiques. Après avoir organisé les données en agrégats mensuels et réalisé une décomposition temporelle, nous avons identifié une tendance générale et des motifs saisonniers récurrents. Ces résultats ont servi de base à l'application de la méthode Holt-Winters pour effectuer des prévisions sur les 24 prochains mois.

Les prévisions obtenues montrent une continuité logique avec les données historiques, en conservant les motifs saisonniers identifiés. L'évaluation des résidus indique que le modèle capture efficacement les principales dynamiques des données, avec des résidus s'apparentant à un bruit blanc. Cependant, les tests de normalité ont révélé que les résidus ne sont pas parfaitement gaussiens, ce qui pourrait suggérer l'existence de fluctuations non capturées par le modèle.

Ainsi, bien que le modèle Holt-Winters soit performant pour des prévisions à court terme, des ajustements ou des approches complémentaires (comme ARIMA ou SARIMA) pourraient être envisagés pour affiner les analyses. Ces résultats offrent une base solide pour anticiper les variations futures de la consommation d'électricité et pour soutenir la prise de décision stratégique.

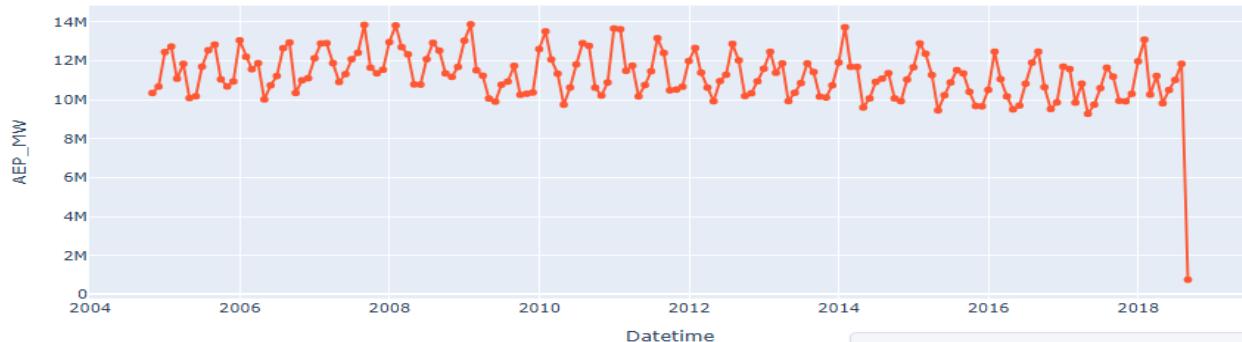
# Tableaux de bord

## Tableau de Bord des Données

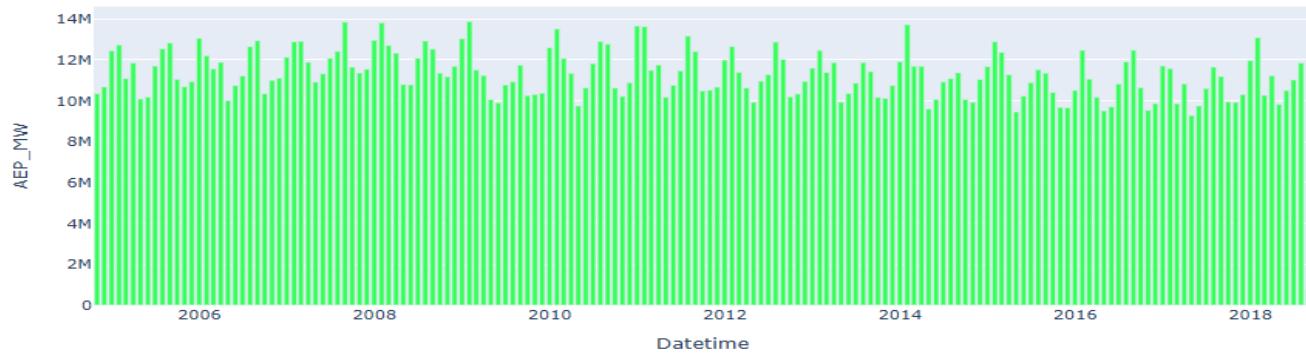
2004-10-31 → 2018-08-31

Valeur Moyenne: 11255524.11 | Variance: 1848427560640.20 | Skewness: -2.54 | Kurtosis: 20.21 | Somme Totale: 1879672527.00 | Min: 754941.00 | Max: 13877850.00 | Mois max consommation: 2009-01

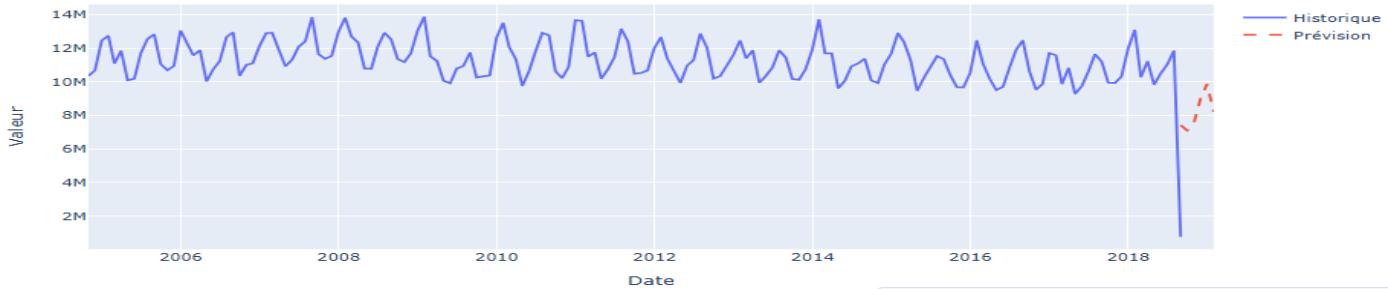
Évolution des valeurs



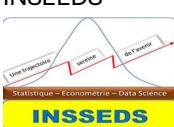
Distribution des valeurs



Prévision Holt-Winters



INSEEDS



KABA MAHAMOUD TOIB

## annexe

```
# Importation des bibliothèques nécessaires
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.holtwinters import ExponentialSmoothing
from scipy.stats import shapiro, kstest, norm

# Chargement des données
data = pd.read_csv(file_path, parse_dates=['Datetime'], index_col='Datetime')
print(data)

# Vérification des valeurs manquantes
missing_values_total = data.isnull().sum().sum()
print(f"Nombre total de valeurs manquantes : {missing_values_total}")
missing_values_per_column = data.isnull().sum()
print("Nombre de valeurs manquantes par colonne :")
print(missing_values_per_column)

# Visualisation des données brutes
plt.figure(figsize=(12, 6))
plt.plot(data.index, data['AEP_MW'], label='Consommation horaire', alpha=0.7)
plt.xlabel('Date')
plt.ylabel('Consommation (MW)')
plt.title('Série temporelle horaire de la consommation d'électricité')
plt.legend()
plt.show()

# Histogramme de la distribution des consommations
plt.figure(figsize=(12, 6))
sns.histplot(data['AEP_MW'], bins=50, kde=True)
```

INSEEDS



KABA MAHAMOUD TOIB

```
plt.xlabel('Consommation (MW)')
plt.ylabel('Fréquence')
plt.title('Distribution de la consommation horaire')
plt.show()

# Agrégation mensuelle
data['month'] = data.index.to_period('M')
monthly_data = data.groupby('month')['AEP_MW'].sum().to_timestamp()
monthly_data = monthly_data.sort_index()
print(monthly_data)

# Visualisation des données mensuelles
plt.figure(figsize=(12, 6))
plt.plot(monthly_data, marker='o', linestyle='-', label='Consommation mensuelle')
plt.xlabel('Date')
plt.ylabel('Consommation (MW)')
plt.title('Série temporelle mensuelle de la consommation d'électricité')
plt.legend()
plt.show()

# Diagramme en bâtons
plt.figure(figsize=(12, 6))
plt.bar(monthly_data.index, monthly_data.values, color='blue', alpha=0.7)
plt.xlabel('Date')
plt.ylabel('Consommation (MW)')
plt.title('Diagramme en bâtons de la consommation mensuelle')
plt.xticks(rotation=45)
plt.show()

# Boîte à moustaches pour la consommation mensuelle
plt.figure(figsize=(12, 6))
sns.boxplot(y=monthly_data.values, color='red')
plt.ylabel('Consommation (MW)')
plt.title('Boîte à moustaches de la consommation mensuelle')
```

```
plt.show()

# Décomposition de la série temporelle
from statsmodels.tsa.seasonal import seasonal_decompose
decomposition = seasonal_decompose(monthly_data, model='additive', period=12)
decomposition.plot()
plt.show()

# Ajustement du modèle Holt-Winters
model = ExponentialSmoothing(monthly_data, trend="add", seasonal="add", seasonal_periods=12)
fit = model.fit()

# Visualisation de l'ajustement Holt-Winters
plt.figure(figsize=(12, 6))
plt.plot(monthly_data.index, monthly_data.values, label='Données historiques')
plt.plot(monthly_data.index, fit.fittedvalues, label='Valeurs ajustées', linestyle='--', color='orange')
plt.xlabel('Date')
plt.ylabel('Consommation (MW)')
plt.title('Ajustement du modèle Holt-Winters')
plt.legend()
plt.show()

# Prévisions sur 24 mois
forecast_24_months = fit.forecast(steps=24)
plt.figure(figsize=(12, 6))
plt.plot(monthly_data.index, monthly_data.values, label='Données historiques')
plt.plot(pd.date_range(start='2018-09-01', periods=24, freq='MS'), forecast_24_months,
        label='Prévisions (24 mois)', linestyle='--', color='red')
plt.xlabel('Date')
plt.ylabel('Consommation (MW)')
plt.title('Prévisions Holt-Winters sur 24 mois')
plt.legend()
plt.show()
```

```
# Analyse des résidus
residuals = monthly_data - fit.fittedvalues
plt.figure(figsize=(12, 6))
plt.plot(residuals, label='Résidus', color='purple')
plt.axhline(y=0, color='black', linestyle='--', linewidth=1)
plt.xlabel('Date')
plt.ylabel('Valeurs des résidus')
plt.title('Graphe des résidus')
plt.legend()
plt.show()

# Analyse de l'auto-corrélation (ACF) des résidus
plot_acf(residuals.dropna(), lags=40)
plt.title('ACF des résidus')
plt.show()

# Test de Ljung-Box pour les résidus
from statsmodels.stats.diagnostic import acorr_ljungbox
ljung_box_test = acorr_ljungbox(residuals.dropna(), lags=[10], return_df=True)
print("Résultats du test de Ljung-Box :")
print(ljung_box_test)

# Test de normalité des résidus
shapiro_test = shapiro(residuals.dropna())
print(f"Test de Shapiro-Wilk : p-value = {shapiro_test.pvalue}")
kstest_test = kstest((residuals - residuals.mean()) / residuals.std(), 'norm')
print(f"Test de Kolmogorov-Smirnov : p-value = {kstest_test.pvalue}")
```

