



MINISTÈRE  
DE L'ENSEIGNEMENT  
SUPÉRIEUR  
ET DE LA RECHERCHE

*Liberté  
Égalité  
Fraternité*



UNIVERSITÉ  
PARIS-EST CRÉTEIL  
VAL DE MARNE

Master Méthodes Appliquées de la Statistique et de  
l'Econométrie pour la Recherche, l'Analyse et le  
Traitement de l'Information

Mémoire

# APPROCHE PROBABILISTE DE LA PREVISION DE LA CONSOMMATION D'ENERGIES A L'AIDE D'UN RESEAU BAYESIEN DYNAMIQUE

Soutenu le :  
13/05/2025

Rapporteurs :

- Sylvain **CHAREYRON**
- Zineb **ABIDI**

Auteur :

**BENGONE AKOU** Lajoie

# TABLE DES MATIERES

## 1. INTRODUCTION

- A. Problématique
- B. Enjeux

## 2. REVUE DE LITTÉRATURE

- A. Modèles conventionnels de statistique
- B. Modèles d'apprentissage automatique
- C. Modèles d'apprentissage profond
- D. Modèles hybrides

## 3. BASE DE DONNEES

- A. Présentation du programme UKDALE
- B. Préparation de la base de donnée

## 4. ANALYSES EXPLORATOIRES

- A. Tableau 1 - Statistiques descriptives
- B. Graphique 1 - Diagramme de Tukey
- C. Graphique 2 - Séries chronologiques

## 5. MODELES DE PREVISION

- A. Modèle ARIMA
- B. Modèle de Réseaux Bayésien Dynamique
- C. Analyse Comparative

## 6. CONCLUSION

- A. Synthèse
- B. Perspective

# **I – INTRODUCTION**

**1. Problématique**

**2. Enjeux**

# INTRODUCTON : La maîtrise de l'énergie et ses bienfaits

- ❑ La maitrise des différentes formes d'énergie que sont : l'eau, l'électricité, le charbon a permis aux Hommes d'améliorer leur condition de vie.

Les usines



Les transports

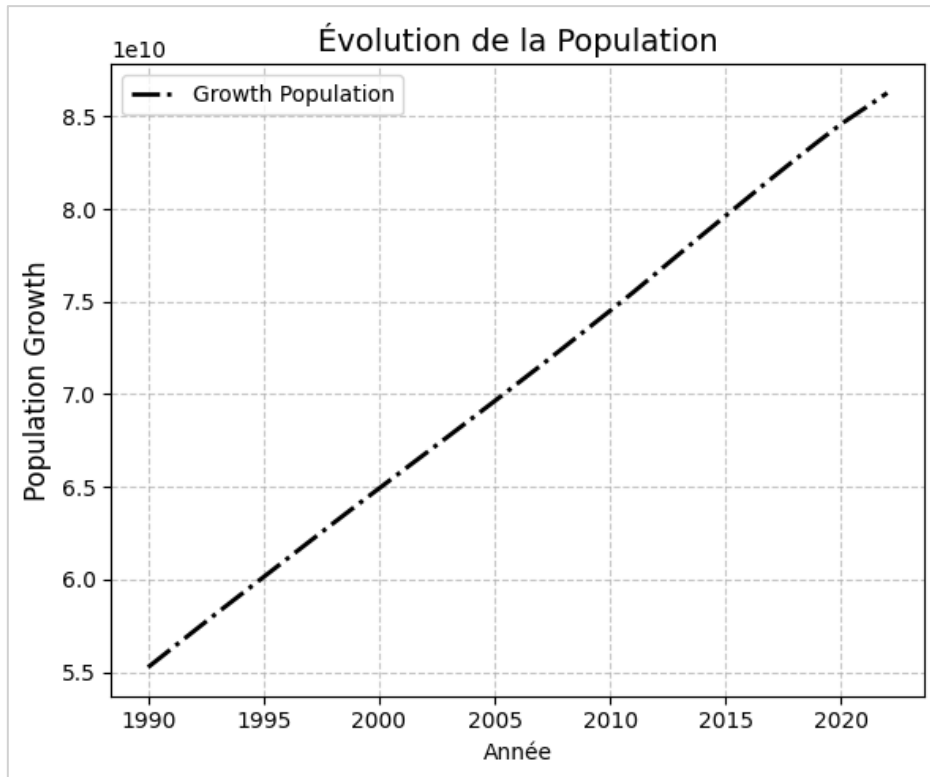


Les réseaux de télécommunications

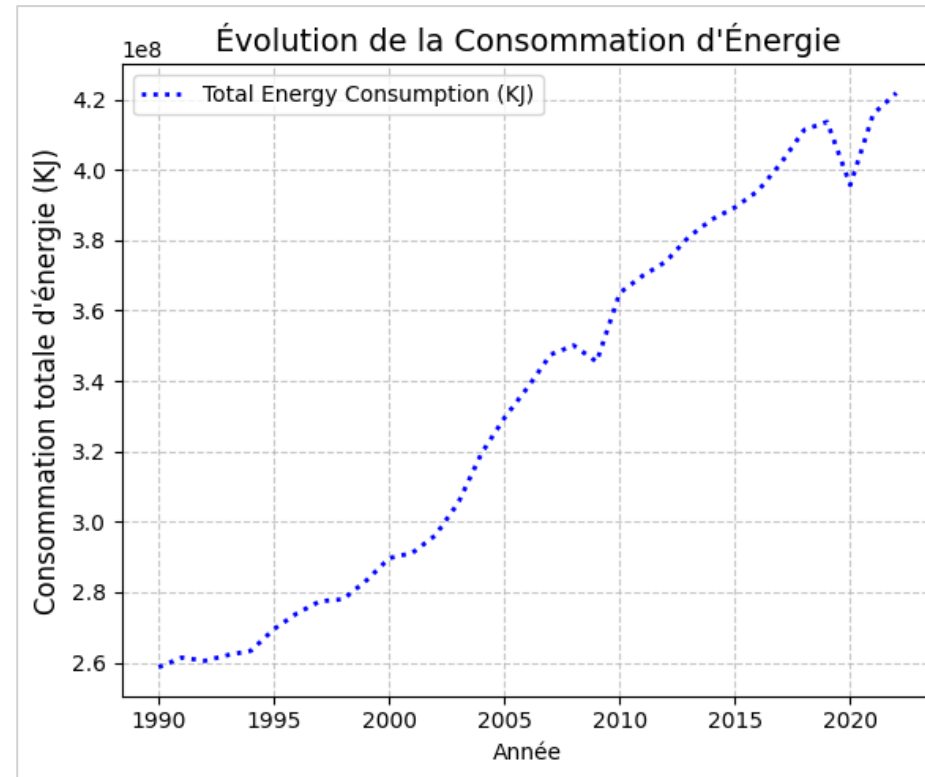


# INTRODUCTON : La maîtrise de l'énergie et ses bienfaits

- ❑ L'amélioration des conditions de vie des individus a eu comme conséquences la croissance de la population mondiale et consommation d'énergie finale.



Source : Données issues de la Banque Mondiale, 2024



Source : Données issues de l'Agence Internationale de l'Énergie, 2024

**Plus la population augmente plus les besoins en termes d'énergie augmentent également**

# INTRODUCTION : La raréfaction des ressources naturelles et l'efficacité énergétique

## Contrainte :

- ❑ Cependant, les ressources naturelles qui servent à la production de l'énergie sont limitées

## Problématique :

- ❑ Comment peut-on utiliser de façon plus efficace, l'énergie encore disponible, dans ce contexte de raréfaction des ressources naturelles et de croissance de la population.

## Solution envisagée :

- ❑ Développement d'un modèle mathématique pour ajuster de façon optimale la production d'énergie et la demande d'énergie.

## Enjeux :

- ❑ Enjeux d'ordre écologique et économique.

# II – REVUE DE LITTÉRATURE

- 1. Modèles conventionnels de statistiques**
- 2. Modèles d'apprentissage automatique**
- 3. Modèles d'apprentissage profond**
- 4. Modèles hybrides**

# REVUE DE LITTÉRATURE : Méthodes d'analyse de données

La prévision de la consommation d'énergie est un phénomène complexe pour lequel plusieurs méthodes mathématiques ont été mobilisées. Parmi ces méthodes on retrouve les méthodes d'analyse de données.

- ☐ Modèles conventionnels de statistique
- ☐ Modèles d'apprentissage automatique
- ☐ Modèles d'apprentissage profond
- ☐ Modèles hybrides

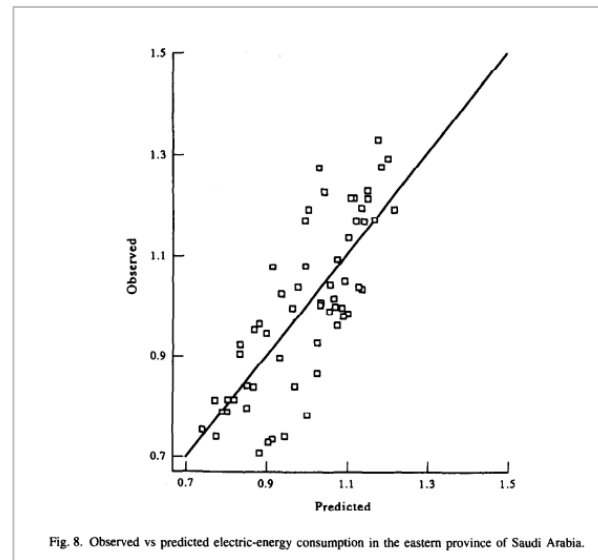
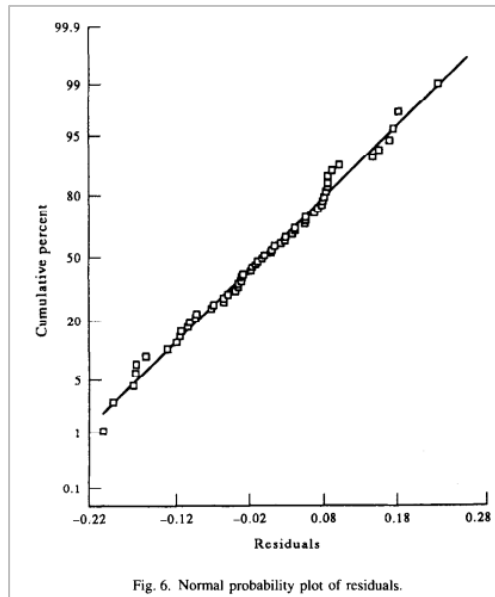


# REVUE DE LITTÉRATURE : Modèles conventionnels de statistiques

- ❑ Les modèles conventionnels regroupent les modèles de régression linéaire, non linéaire et les séries temporelles
- ❑ Ces approches font partie des premières méthodes d'analyse de données utilisées dans le cadre de cette thématique.
- ❑ Elles ont notamment été utilisées pour prédire la consommation d'énergie à l'échelle nationale, sectorielle et résidentielle.

**ETUDE DE CAS :** (Al-Garni et al, 1994) : A regression model for electric-energy consumption forecasting in eastern Saudi

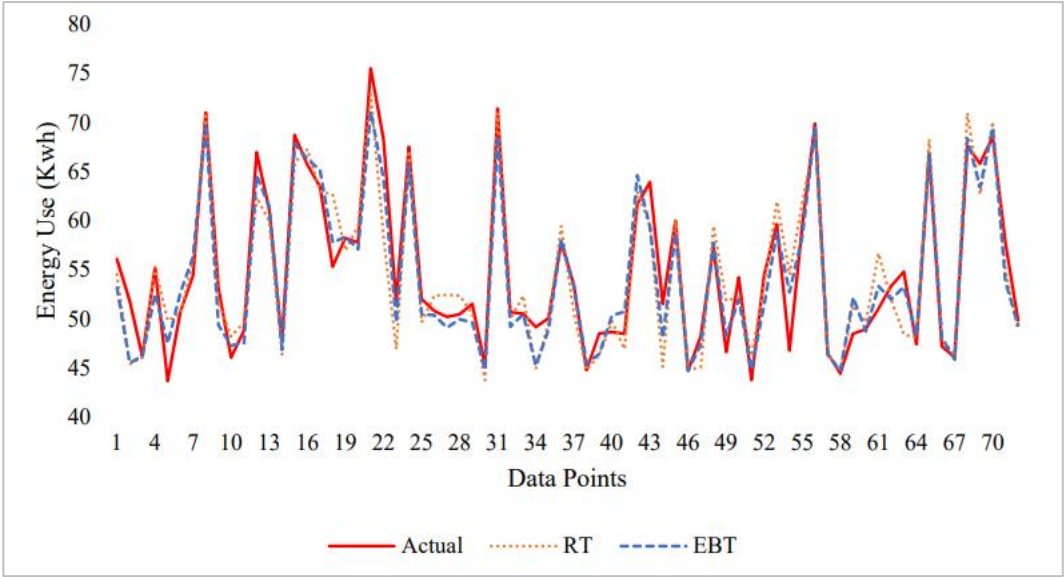
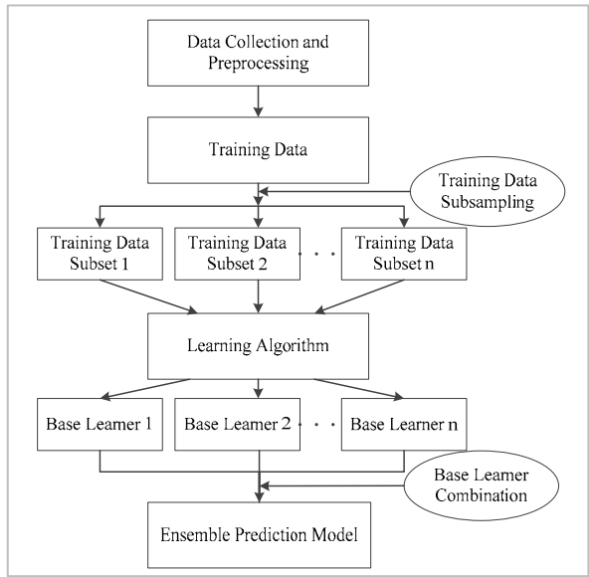
**REGRESSION MULTIPLE :**  $E_p = \alpha_1 Pop + \alpha_2 Temp + \alpha_3 Hum + \alpha_4 Rad$



# REVUE DE LITTÉRATURE : Modèles d'apprentissage automatique

- ❑ Ces modèles permettent d'éviter les hypothèses des modèles conventionnels de statistique
- ❑ Leur popularité a été favorisée avec l'émergence des données massives
- ❑ Ils regroupent : les modèles de random forest, Gradient boosting, Machine à vecteur de support.

**ETUDE DE CAS :** (Wang et al, 2016) : Homogeneous Ensemble Model for Building Energy Prediction: A Case Study Using Ensemble Regression Tree

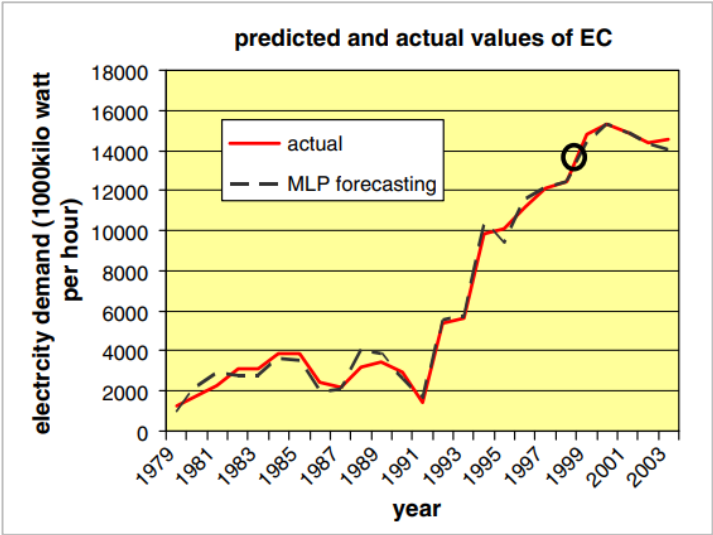
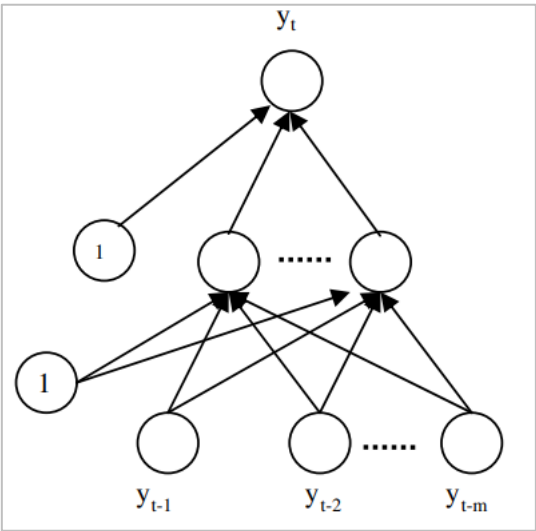


	RT	EBT
R2	0.93	0.88
RMSE	2.13	2.83
MAPE	3.17	3.89

# REVUE DE LITTÉRATURE : Modèles d'apprentissage profond

- ❑ Ils permettent d'obtenir un temps de calcul plus court et revêtement une meilleure gestion des données massives.
- ❑ Ils regroupent les perceptrons multicouches, les réseaux de neurones convolutifs et récurrents.

**ETUDE DE CAS :** (Azadeh et al, 2022) : Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors.

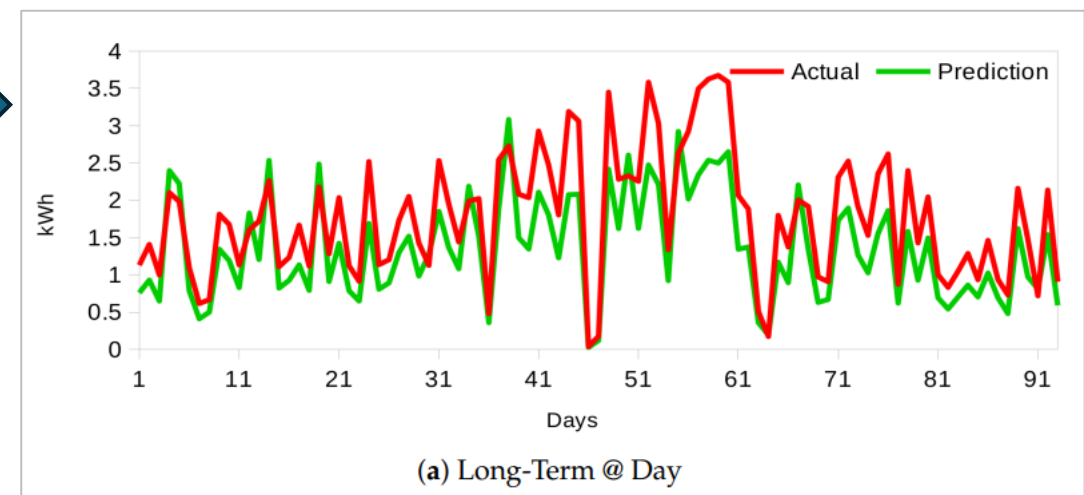
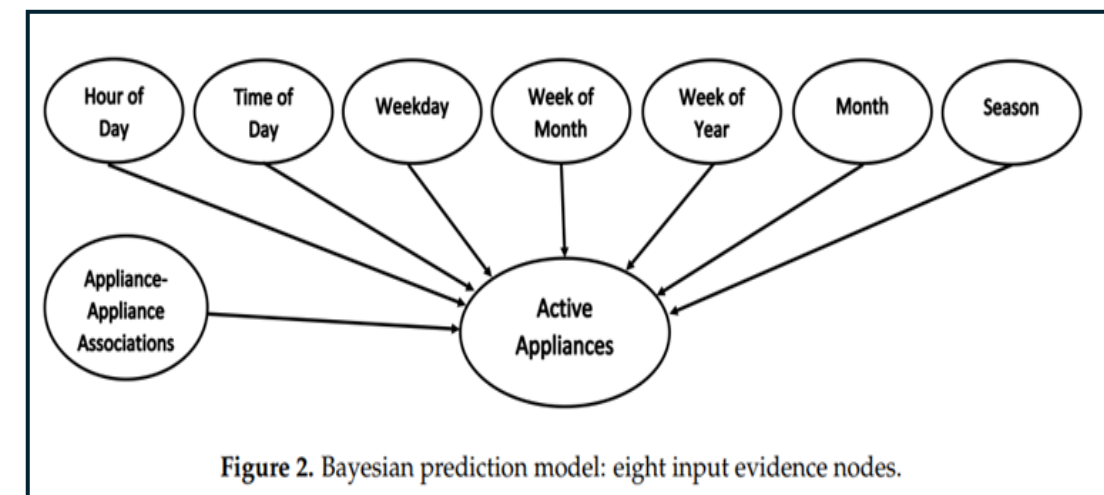


	MLP
MSE	0.22
MAPE	0.03

# REVUE DE LITTÉRATURE : Modèles hybrides

- ❑ Ils possèdent de très bonnes capacités pour traiter des données massives
- ❑ Ils permettent une analyse très détaillée de l'historicité des données et du caractère non linéaire des données
- ❑ Ces modèles permettent de résoudre le problème d'explicabilité des modèles d'apprentissage automatique
- ❑ Ces méthodes regroupent les réseaux bayésiens, les LSSVM, GRU

**ETUDE DE CAS :** (Singh et Yassine, 2018) : Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting



# REVUE DE LITTÉRATURE : Résumé

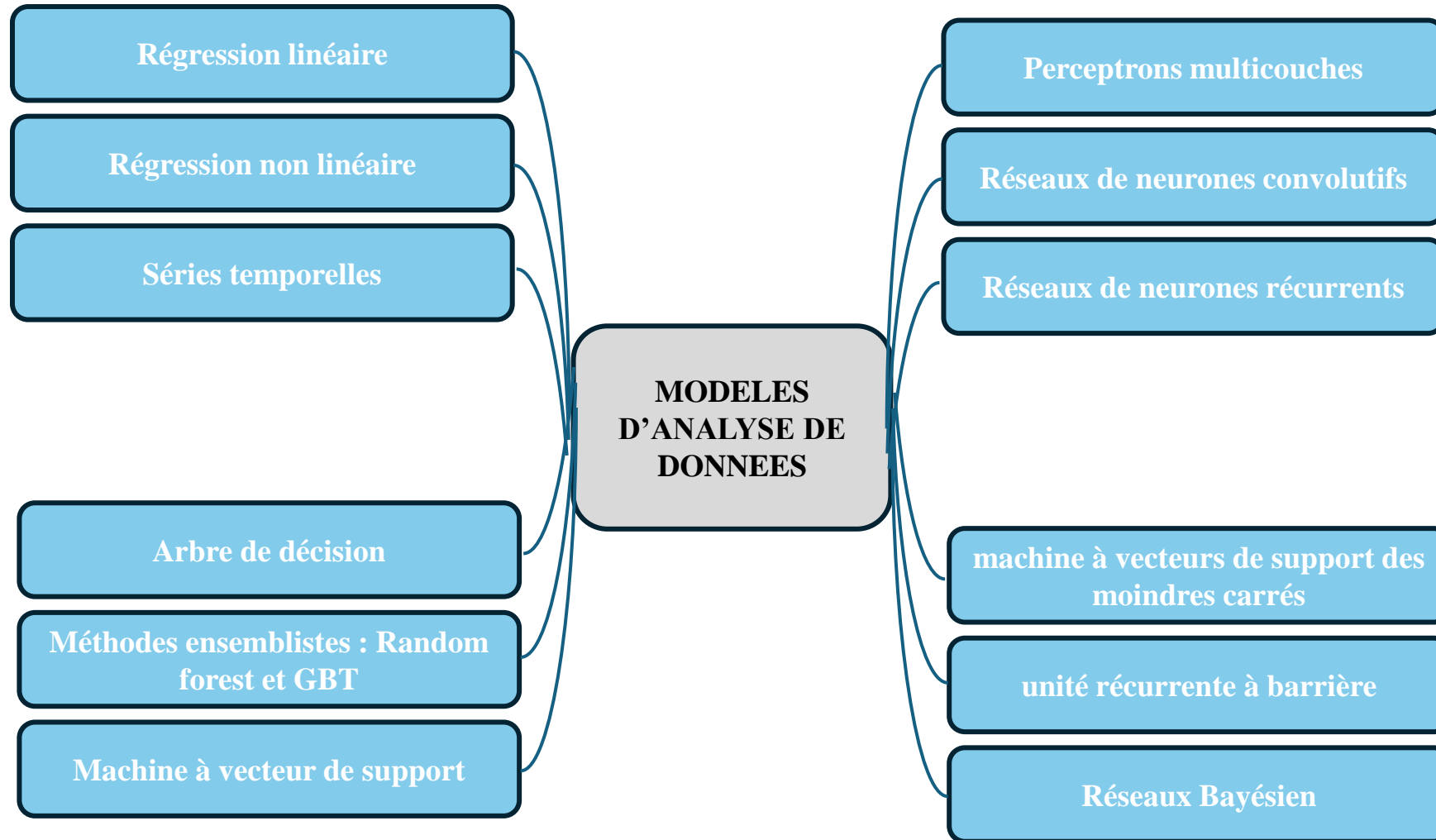


Figure : Résumé des méthodes d'analyse de données pour la prévision de la consommation d'énergie

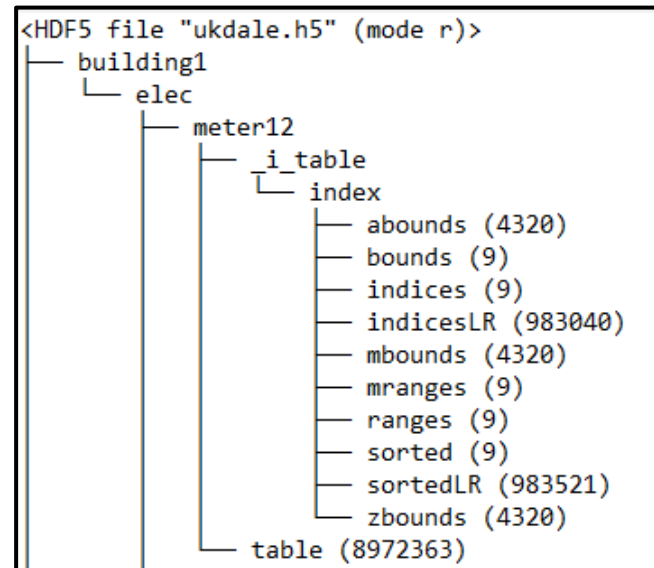
# III – BASE DE DONNEES

**1. Présentation du programme de recherche UKDALE**

**2. Préparation de la base de données**

# BASE DE DONNEES : Présentation du programme de recherche UKDALE

- ❑ La base de données UKDALE a été développée en 2013 au sein du département informatique de Imperial College dans le cadre d'un programme de recherche en lien avec les « energy disaggregation ».
- ❑ Elle a été citée 1194 fois sur Google Scholar (à ce jour).
- ❑ Elle contient les données de panel de 5 foyers en matière de consommation d'électricité. Ces données ont été collectées à partir des systèmes embarqués entre 2013 et 2015 avec une résolution de 1 Hz ou 6 Hz.

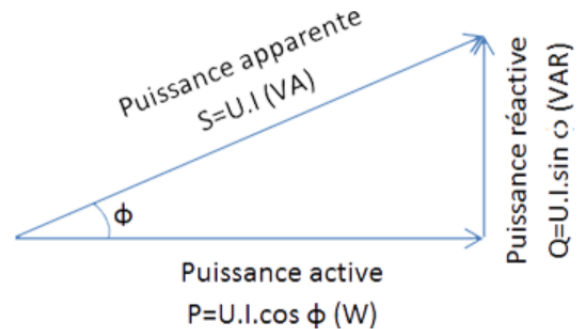


# BASE DE DONNEES : Présentation du programme de recherche UKDALE

Base de données brute en format DataFrame Pandas

	unix	active_power	apparent_power	main_rms_voltage
0	1.363548e+09	337.88	431.04	240.15
1	1.363548e+09	339.43	427.94	240.56
2	1.363548e+09	340.63	429.66	241.07
3	1.363548e+09	338.80	426.99	240.44
4	1.363548e+09	340.88	429.13	241.01

- ☐ Unix : Variable temporelle. Date à laquelle les données ont été enregistrées
- ☐ Active power : puissance active
- ☐ Apparent power : puissance apparente
- ☐ Main rms voltage : tension





# BASE DE DONNEES : Présentation du programme de recherche UKDALE

Base de données traitées au format DataFrame Pandas

	unix	unix_year	unix_month	unix_day	unix_hour	unix_minute	unix_second	unix_microsecond	active_power	apparent_power	main_rms_voltage
unix											
2013-03-17 19:12:43	1.363548e+09	2013.0	3.0	17.0	19.0	12.0	43.0	0.0	337.88	431.04	240.15
2013-03-17 19:12:44	1.363548e+09	2013.0	3.0	17.0	19.0	12.0	44.0	0.0	339.43	427.94	240.56
2013-03-17 19:12:45	1.363548e+09	2013.0	3.0	17.0	19.0	12.0	45.0	0.0	340.63	429.66	241.07
2013-03-17 19:12:46	1.363548e+09	2013.0	3.0	17.0	19.0	12.0	46.0	0.0	338.80	426.99	240.44
2013-03-17 19:12:47	1.363548e+09	2013.0	3.0	17.0	19.0	12.0	47.0	0.0	340.88	429.13	241.01

- ☐ Création de nouvelles variables temporelles
- ☐ Changement de type de variables
- ☐ Vérification des valeurs manquantes
- ☐ Vérification des valeurs nulles
- ☐ Utilisation de Python, SQL, Microsoft Databricks et Google Colab
- ☐ Utilisation de la variable : active power comme variable dépendante

# IV – ANALYSES EXPLORATOIRES

- 1. Tableau 1 - Statistique descriptive**
- 2. Graphique 1 - Diagramme de Tukey**
- 3. Graphique 2 - Série chronologique**
- 4. Graphique 3 - Histogramme**

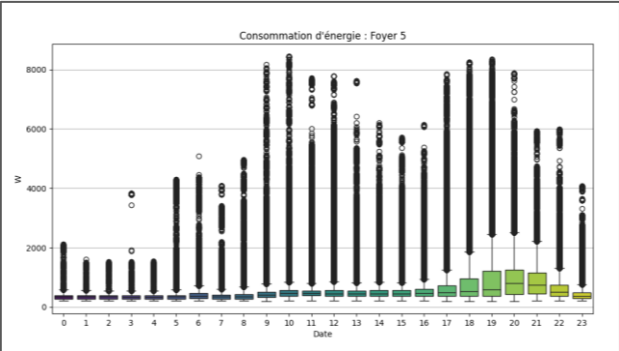
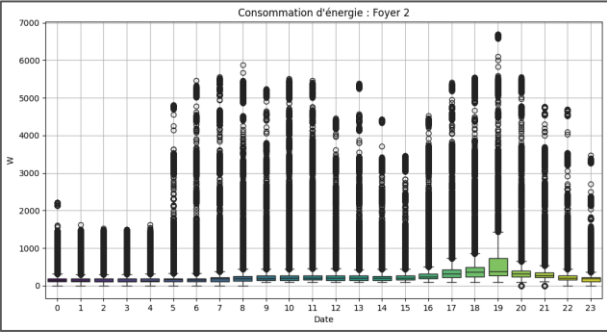
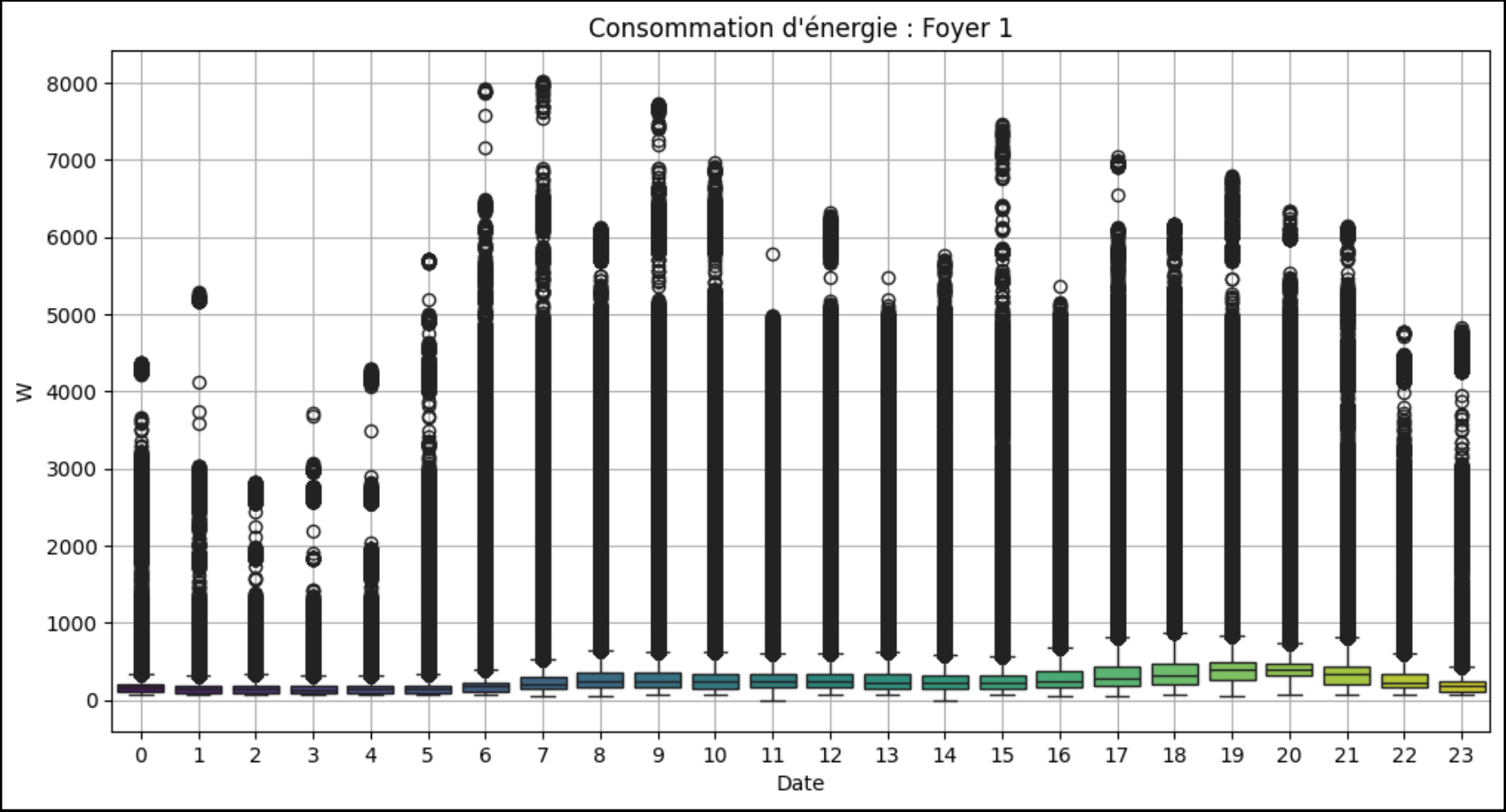
# ANALYSES EXPLORATOIRES : Tableau 1 – Statistique descriptives

Critères	Variables	Foyer 1	Foyer 2	Foyer 5
Max	Puissance active	2829.45	1863.51	3530.69
	Puissance apparente	2902.86	1877.75	3626.27
	Tension	246.64	244.63	250.59
Min	Puissance active	87.51	0.00	226.11
	Puissance apparente	125.26	145.60	399.48
	Tension	237.77	236.62	238.76
Moyenne	Puissance active	318.55	298.63	571.98
	Puissance apparente	371.27	333.24	730.42
	Tension	243.18	241.15	246.12

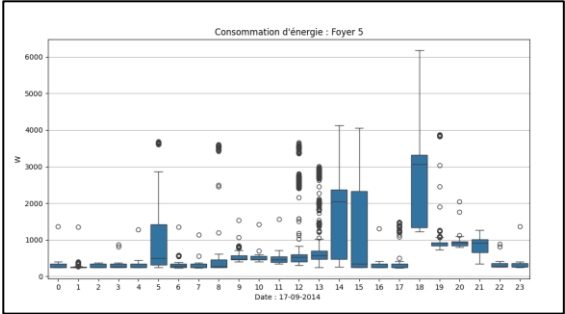
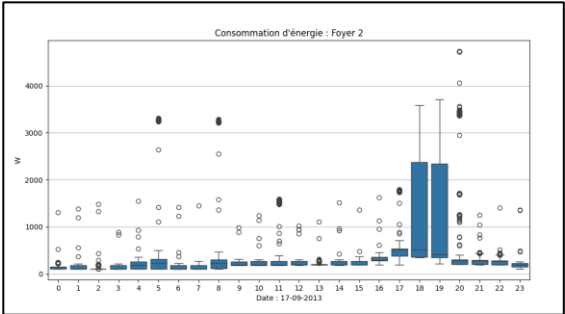
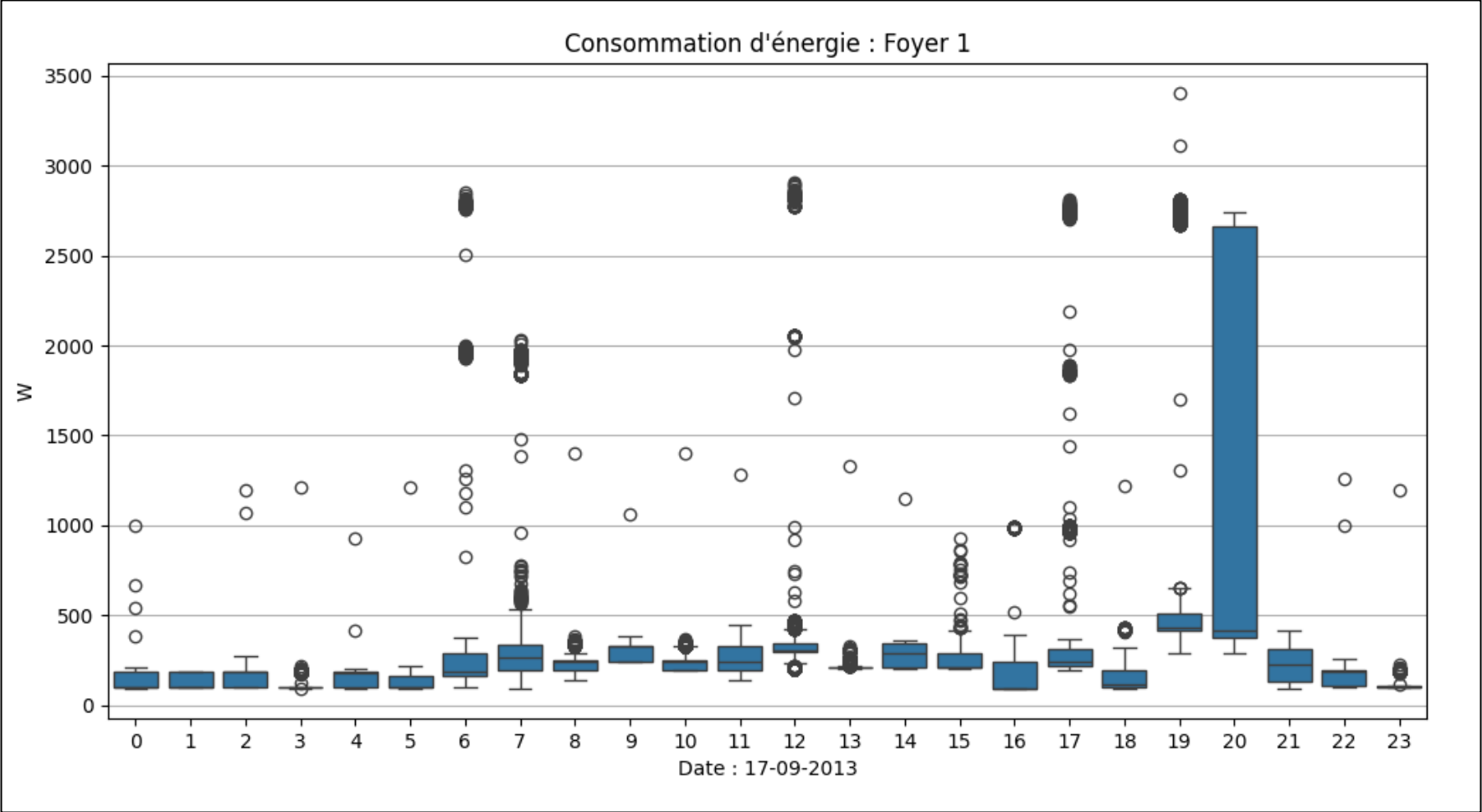
# ANALYSES EXPLORATOIRES : Tableau 1 – Statistique descriptives (Suite)

Critères	Variables	Foyer 1	Foyer 2	Foyer 5
Médiane	Puissance active	222.43	207.76	429.95
	Puissance apparente	264.29	240.36	601.73
	Tension	243.23	241.19	246.10
Ecart-type	Puissance active	242.06	260.33	396.12
	Puissance apparente	261.11	261.29	377.18
	Tension	1.17	1.13	1.68
Valeurs nulles	Puissance active	0.00	0.00	0.00
	Puissance apparente	0.00	0.00	0.00
	Tension	0.00	0.00	0.00

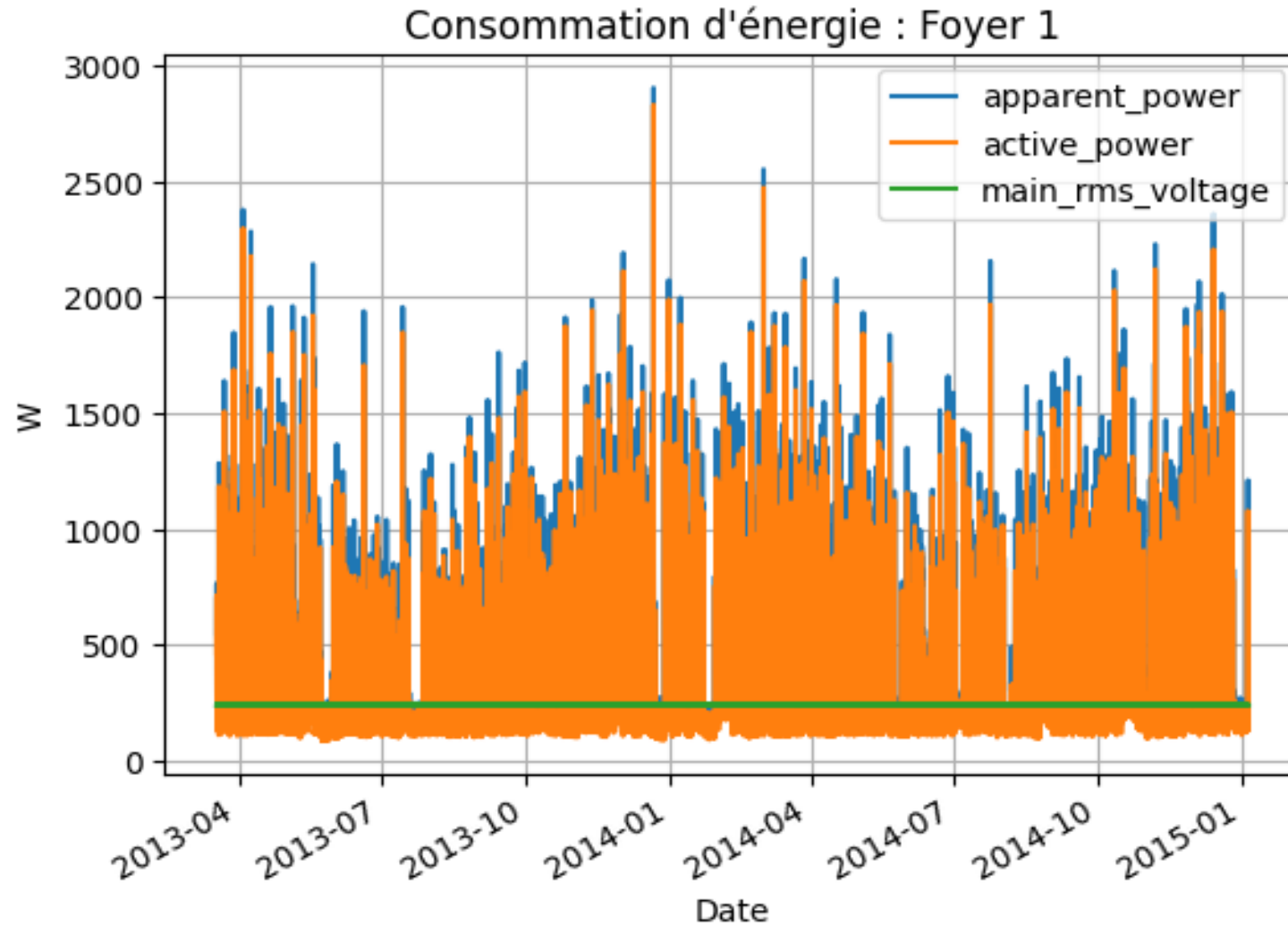
# ANALYSES EXPLORATOIRES : Graphique 1 – Diagramme de Tukey



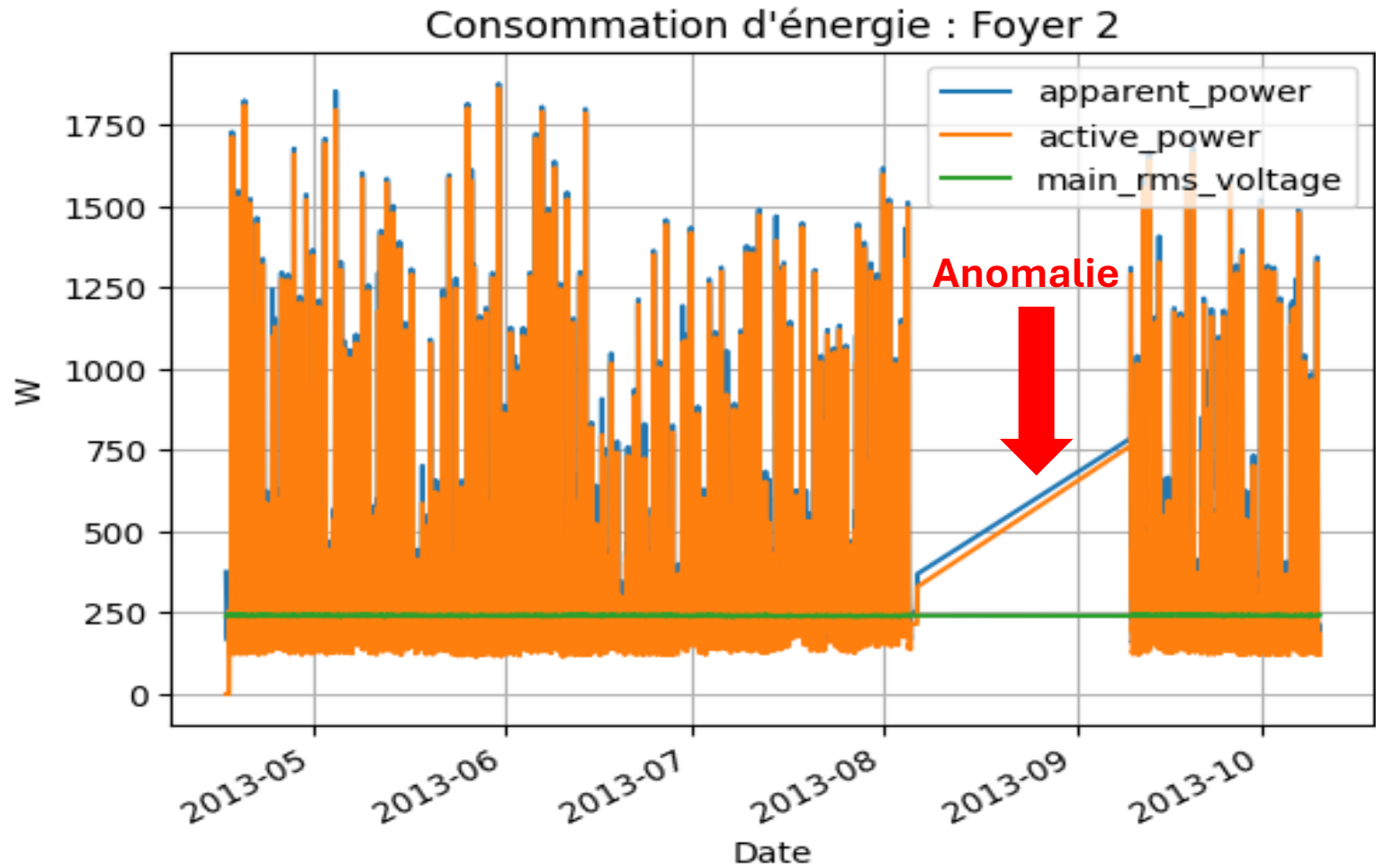
# ANALYSES EXPLORATOIRES : Graphique 1 – Diagramme de Tukey



# ANALYSES EXPLORATOIRES : Graphique 1 – Série chronologique

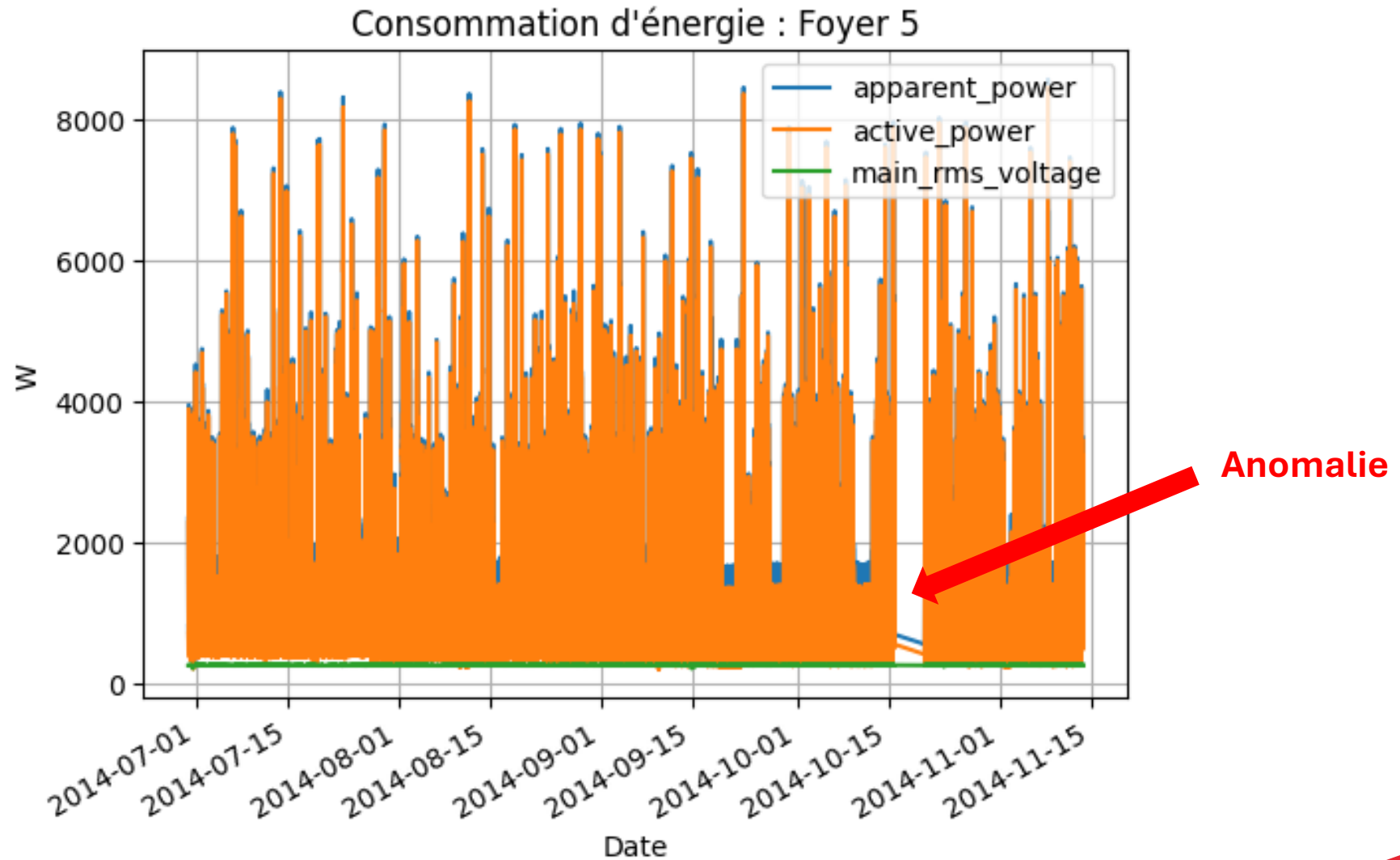


# ANALYSES EXPLORATOIRES : Graphique 2 – Série chronologique





# ANALYSES EXPLORATOIRES : Graphique 3 – Série chronologique



# V – MODELES DE PREVISION

**1. ARIMA**

**2. Réseaux Bayésiens Dynamiques**

**3. Comparaison**

# MODELES DE PREVISION : ARIMA

- ❑ Le modèle ARIMA est composée trois parties : La partie Autorégressive (AR), Moyenne Mobile (MA) et Intégrée ( I ).

- ❑ La partie Autorégressive vise à modéliser la série chronologique en fonction de ces valeurs passée.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

- ❑ La partie Moyenne Mobile vise à modéliser la série chronologique en fonction de ces termes d'erreur passés.

$$y_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_p \epsilon_{t-p} + \epsilon_t$$

- ❑ La partie Intégration ( différenciation ) vise à rendre la série chronologique stationnaire

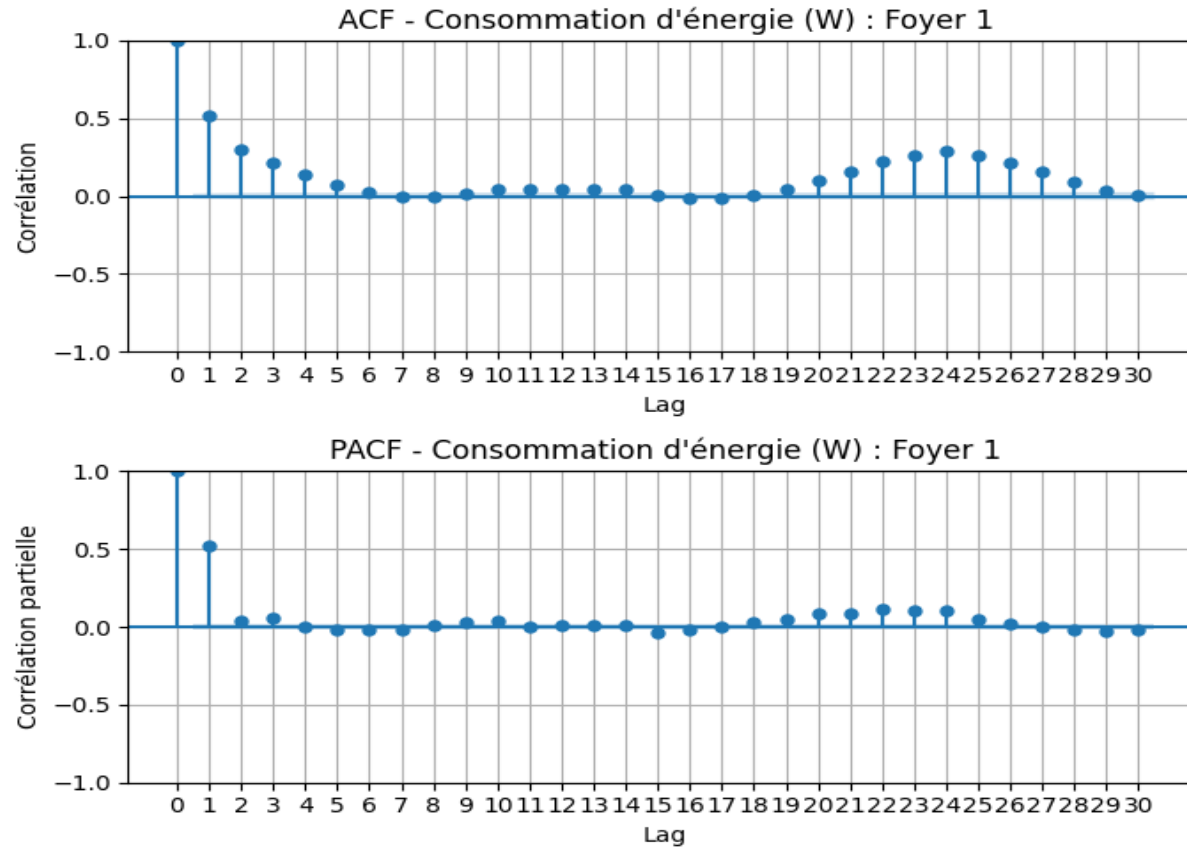
$$\Delta^k y_t = y_t - y_{t-k}$$

- ❑ Pour développer le modèle prévision, on va utiliser la Méthodologie de Box-Jenkins.

- 1) Vérification de la stationnarité
- 2) Estimation des paramètres du modèle  $ARIMA(p, d, q)$
- 3) Evaluation du modèle

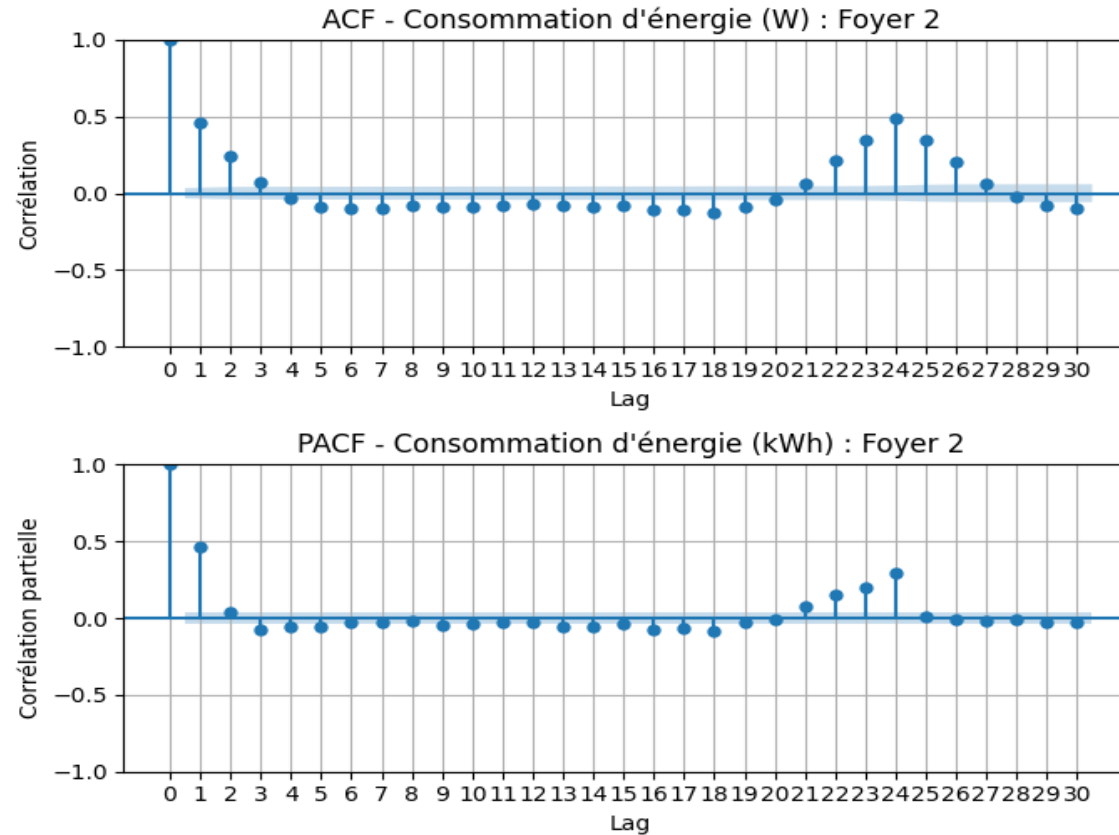
# MODELES DE PREVISION : ARIMA – Etude de la stationnarité

- ❑ Analyse des autocorrélations et des autocorrélations partielles ( Foyer 1 )



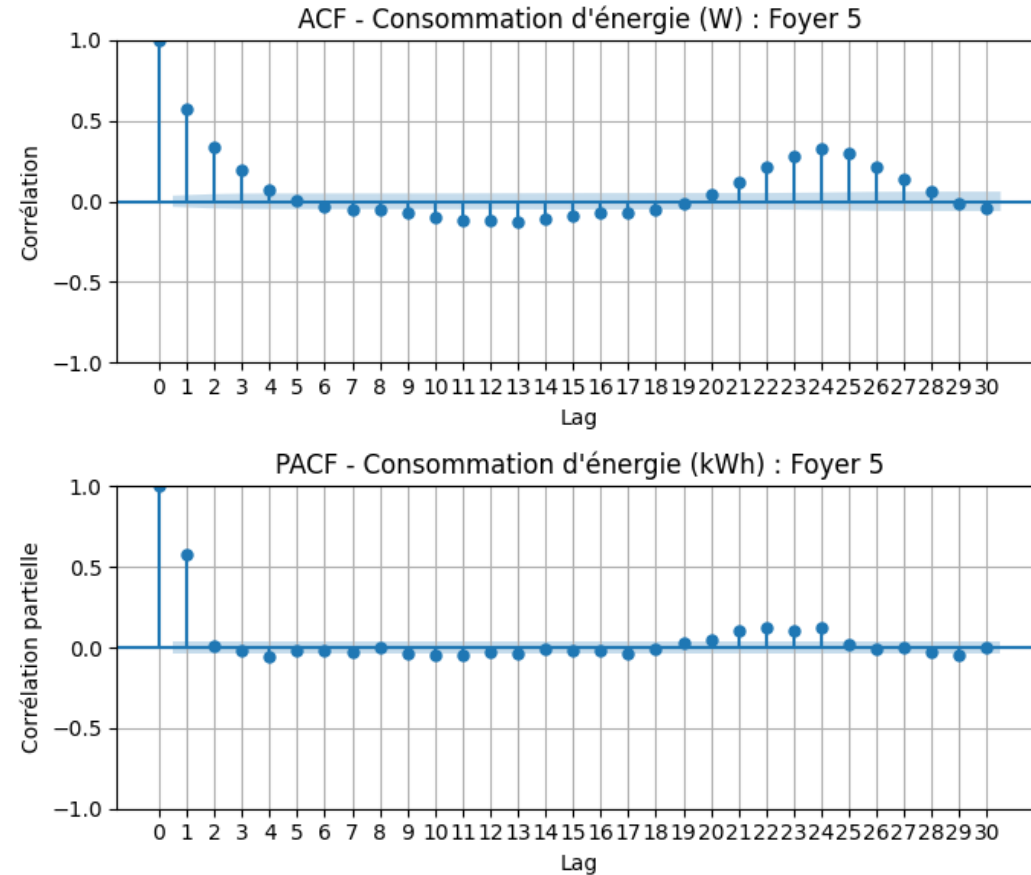
# MODELES DE PREVISION : ARIMA – Etude de la stationnarité

- ❑ Analyse des autocorrélations et des autocorrélations partielles ( Foyer 2)



# MODELES DE PREVISION : ARIMA – Etude de la stationnarité

- ❑ Analyse des autocorrélations et des autocorrélations partielles ( Foyer 5)



# MODELES DE PREVISION : ARIMA – Etude de la stationnarité

- ❑ Test de Dickey-Fuller : Le test de Dickey-Fuller est un test statistique qui permet de déterminer si une série chronologique est stationnaire ou pas.
- ❑ Hypothèse nulle  $H_0: \phi = 1 \Rightarrow$  Le processus contient une racine unitaire ( Processus non stationnaire )  
Hypothèse alternative  $H_1: \phi < 1 \Rightarrow$  Le processus ne contient aucun racine unitaire ( processus stationnaire )
- ❑ Résultats du test de Dickey-Fuller sur les données des foyers 1, 2, 5.

	Foyer 1	Foyer 2	Foyer 5
Valeur de test	-12.85	-7.88	-7.92
Pvaleur	5.49e-24	4.57-12	3.57-12
Conclusion	Stationnaire	Stationnaire	Stationnaire

# MODELES DE PREVISION : ARIMA – Estimations des paramètres du modèles

- ❑ Les modèles ARIMA ont été sélectionnés à partir d'un processus itératif de la méthodologie de Box-Jenkins. Les informations clés ont été résumées dans le tableau suivant :

	Foyer 1	Foyer 2	Foyer 5
<b>Model :</b>	(1, 1, 1)	(5, 0, 3)	(0, 1, 1)
<b>No Observations</b>	12596	2706	2539
<b>Ljun-Box (Q)</b>	0.45	0.01	9.26
<b>Prob (Q)</b>	0.50	0.91	0.00
<b>Heteroskedasticity (H)</b>	1.18	0.75	0.98
<b>Prob (H)</b>	0.00	0.00	0.76
<b>Jarque-Bera (JB)</b>	103769.31	15696.62	11679.90
<b>(JB)</b>	0.00	0.00	0.00



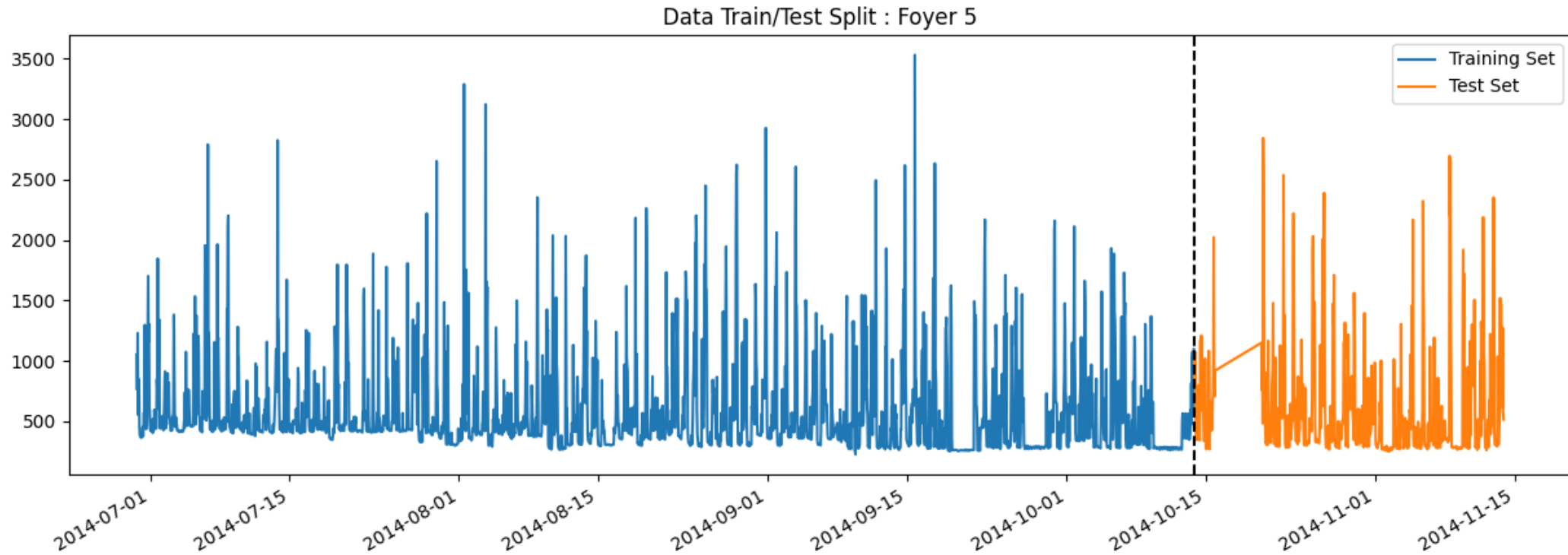
# MODELES DE PREVISION : ARIMA – Résultats et Evaluations du modèle

❑ Les modèles ont été entraînés sur 80% de la base de données et testés les 20% restants

Métriques	Foyer 1	Foyer 2	Foyer 5
Log Likelihood	-84486.127	-18555.961	-18415.439
AIC	168978.253	37131.923	36834.879
BIC	169000.576	37190.955	36846.557
MAE	193.888374	175.349217	306.666331
MAPE	0.594796	0.784164	0.596263
RMSE	276.896095	254.146390	435.986433

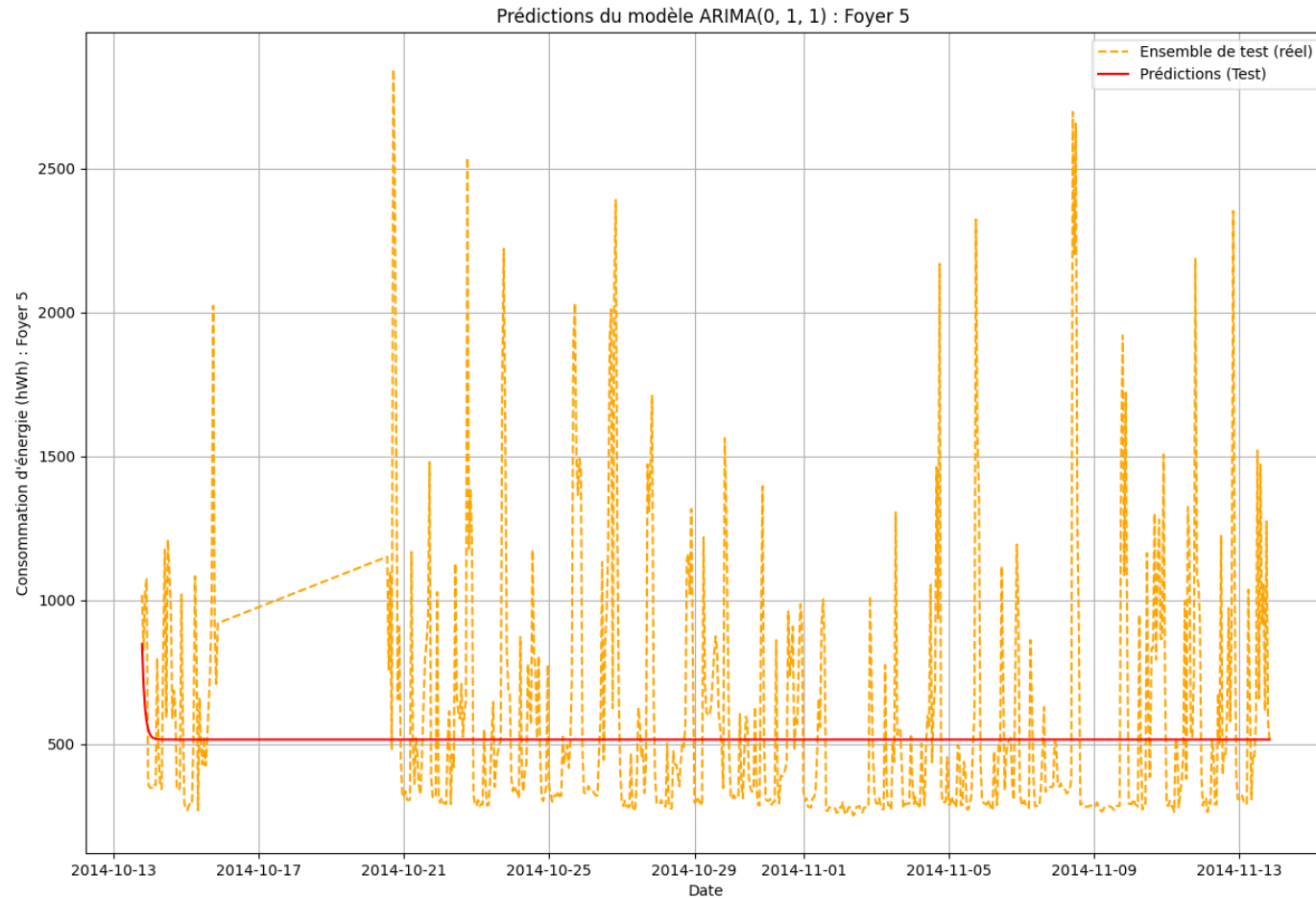
# MODELES DE PREVISION : ARIMA – Résultats et Evaluations du modèle

❑ Exemple de séparation de dataset en base d'entrainement et base test - Foyer 5



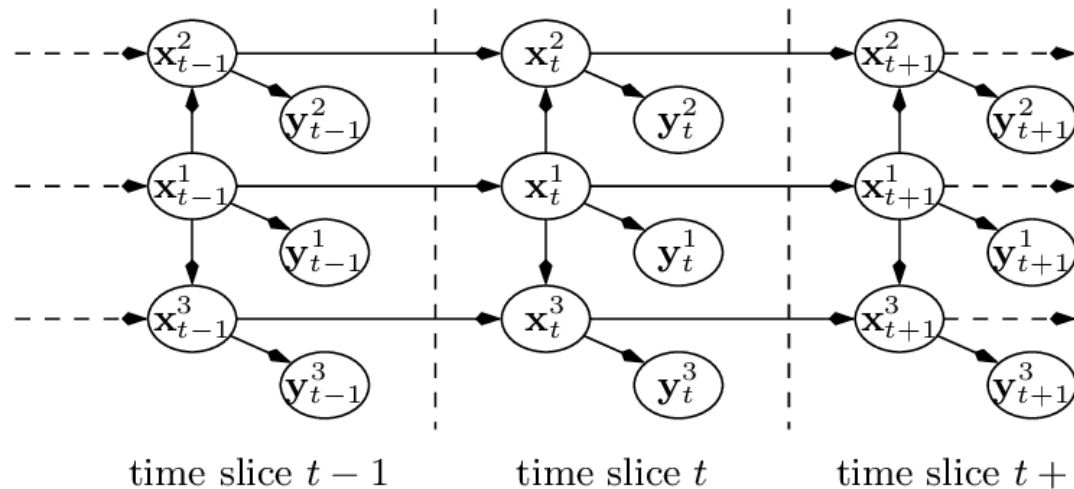
# MODELES DE PREVISION : ARIMA – Résultats et Evaluations du modèle

❑ Exemple : Prédiction à l'aide du modèle ARIMA – Foyer 5



# MODELES DE PREVISION : Réseau Bayésien Dynamique

- ❑ Un Réseau Bayésien Dynamique est un modèle statistique représenté sous la forme d'un graphe acyclique orienté qui permet de représenter l'évolution des variables aléatoires.
- ❑ Ils sont composés principalement de deux éléments : les nœuds et les arcs (arrêts).

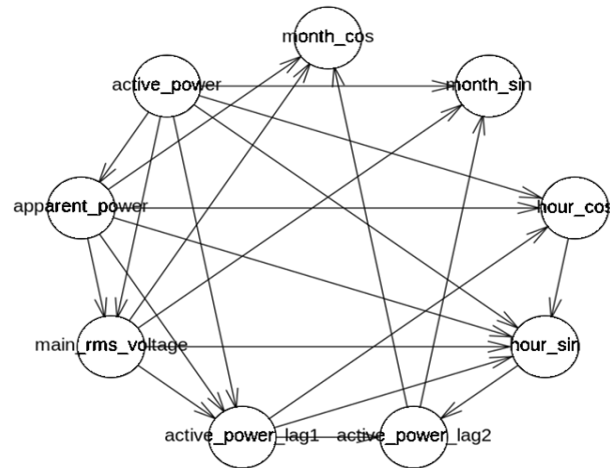


- ❑ Dans le cadre de ce mémoire, chaque nœud est modélisé par une variable aléatoire continue qui suit une loi normale.  $P(X, Y) = \prod_{t=1}^{T-1} P(x_t | x_{t-1}) \prod_{t=0}^{T-1} P(y_t | x_t) P(x_0)$

# MODELES DE PREVISION : Réseau Bayésien Dynamique

□ Le développement du modèle probabiliste s'est effectué en 3 grandes étapes.

1. La première étape concernait le traitement des données qui se présente comme suit
  - Suppression des variables discrètes
  - Création des variables périodiques
  - Normalisation des données
2. Apprentissage du Réseau Bayésien Dynamique
  - Le réseau bayésien Dynamique a été élaboré à partir du Critère d'Information Bayésien (BIC).
  - Il possède l'architecture suivante :



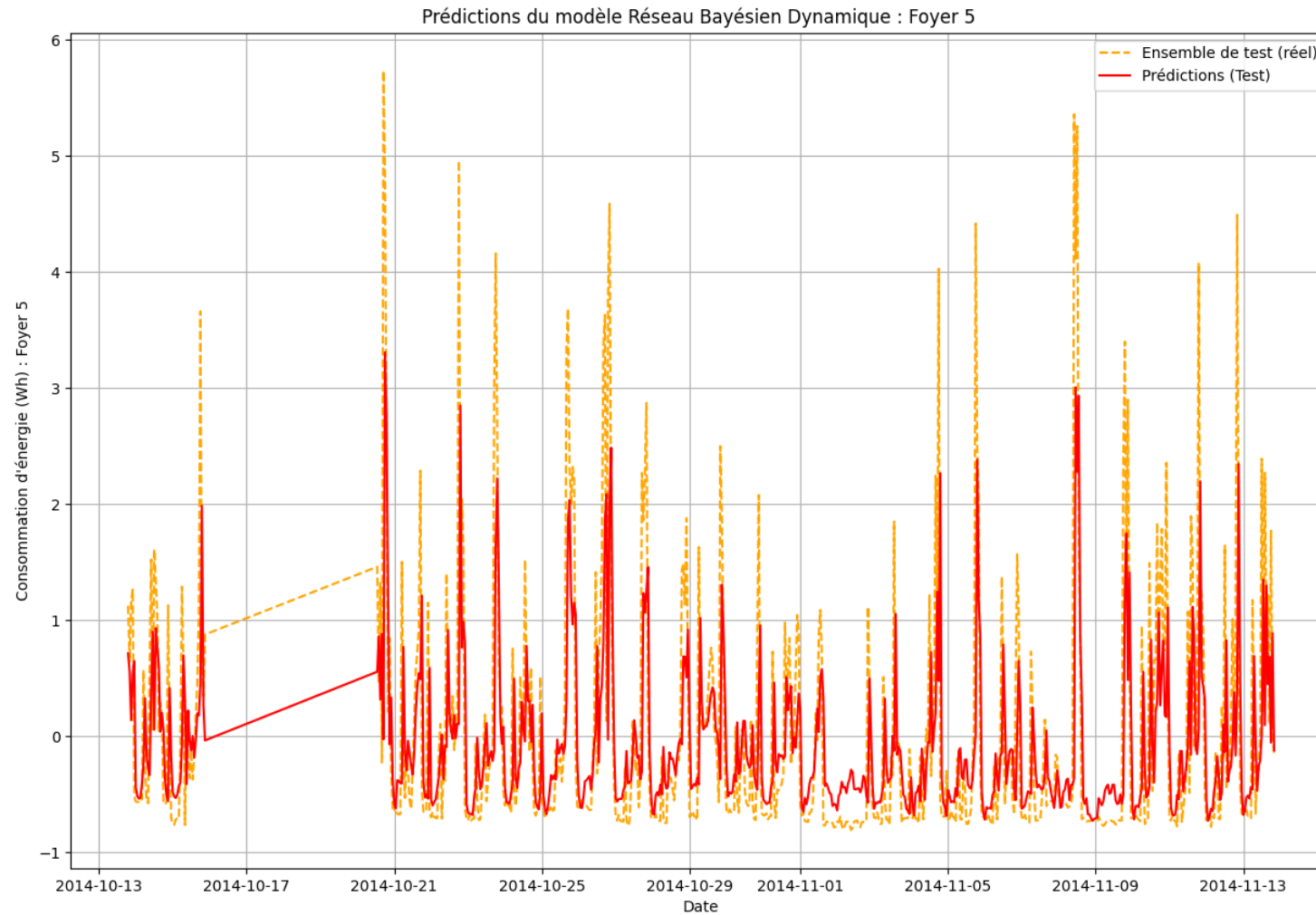
# MODELES DE PREVISION : Réseau Bayésien Dynamique

- ❑ Séparation de la base de données en Base de données d’entraînement (80%) et en base de données test (20%)
- ❑ Evaluation à l’aide des critère information et des métriques de prédictions

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100$$

Métrique	Foyer 1	Foyer 2	Foyer 5
Log-Likelihood	-245947	-48271.58	-44230.11
AIC	491974	96619.16	88538.22
BIC	492271.6	492271.6	88765.92
MAE	0.8011749	0.8011749	0.5263825
MAPE	211.6339	211.6339	128.8375
RMSE	1.144049	1.144049	0.891147

# MODELES DE PREVISION : Réseau Bayésien Dynamique



# MODELES DE PREVISION : Comparaison du modèle ARIMA et RBD

## ARIMA

Métriques	Foyer 1	Foyer 2	Foyer 5
Log Likelihood	-84486.127	-18555.961	-18415.439
AIC	168978.253	37131.923	36834.879
BIC	169000.576	37190.955	36846.557
MAE	193.888374	175.349217	306.666331
MAPE	0.594796	0.784164	0.596263
RMSE	276.896095	254.146390	435.986433

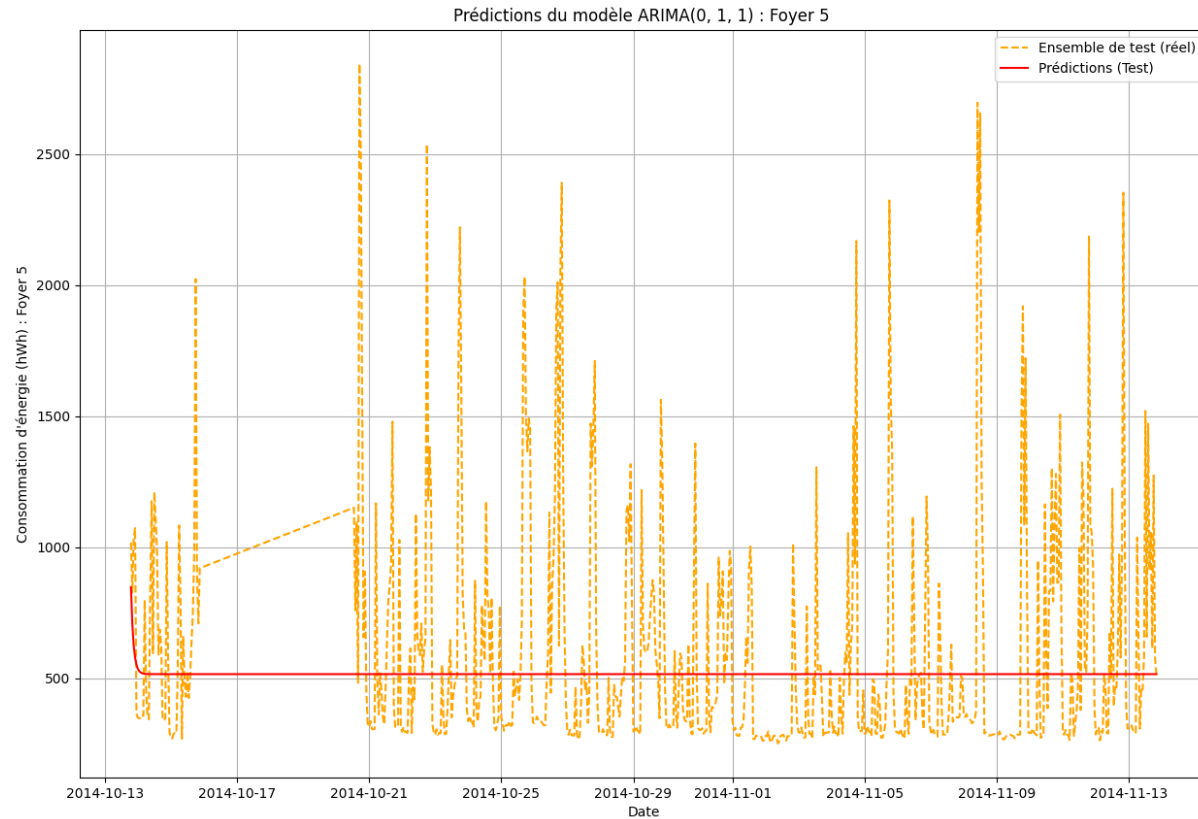
## RESEAU BAYESIEN DYNAMIQUE

Métrique	Foyer 1	Foyer 2	Foyer 5
Log-Likelihood	-245947	-48271.58	-44230.11
AIC	491974	96619.16	88538.22
BIC	492271.6	492271.6	88765.92
MAE	0.8011749	0.8011749	0.5263825
MAPE	211.6339	211.6339	128.8375
RMSE	1.144049	1.144049	0.891147

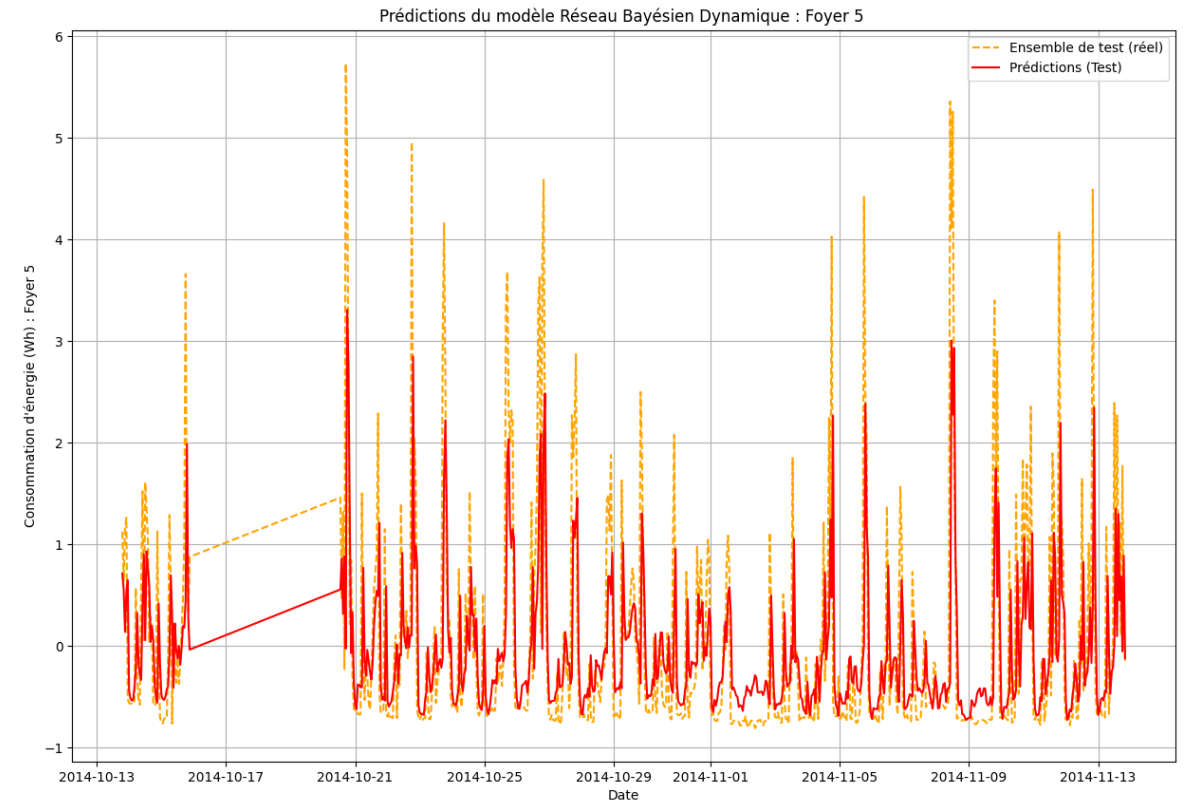


# MODELES DE PREVISION : Comparaison du modèle ARIMA et RBD

## ARIMA



## RESEAU BAYESIEN DYNAMIQUE



# VI – CONCLUSION

**1. Synthèse**

**2. Perspective**

# MODELES DE PREVISION : Synthèse

- ☐ Le Réseau Bayésien Dynamique est très promoteur en matière de prévision et d'explicabilité.
- ☐ Des analyses bayésiennes plus approfondies doivent être effectuées : Correction du MAPE et rapport de vraisemblance.
- ☐ Utilisation d'autres critères de Score pour l'apprentissage des réseaux probabilistes
- ☐ Réalisation d'une étude bibliographique sur les Réseaux Bayésiens Dynamiques ( Non Gaussiens)
- ☐ Implémentation en temps réel ( Data Streaming)

# MODELES DE PREVISION : Perspectives

- ❑ La prévision de la consommation d'énergie est un phénomène complexe en raison de son caractère non linéaire et multifactoriel.
- ❑ 4 grandes familles de méthodes d'analyse de données sont de plus en plus privilégiées : modèles conventionnels, modèles d'apprentissage automatique, profond et hybrides.
- ❑ Les réseaux Bayésiens Dynamiques représentent à ce jour, l'une des méthodes avec le plus grand potentiel.
- ❑ Utilisation de la base de données UKDALE dans le cadre de cet ouvrage.
- ❑ Conclusion par rapport au modèle ARIMA. Globalement, le modèle ARIMA a obtenu des bons résultats en ce qui concerne l'ajustement des modèles. Cependant, les modèles restent très mauvais en termes de prédictions.
- ❑ Conclusion par rapport au Réseau Bayésien Dynamique. L'ajustement des Réseaux Bayésiens Dynamiques est légèrement inférieur à celui des modèles ARIMA. Cependant, les Réseaux Bayésiens Dynamiques ont obtenu des résultats nettement supérieurs à ceux du modèle ARIMA en termes de prévision de la consommation d'électricité

# FIN