



Lena Verboom

Prédiction de la demande en électricité



Projet 9 : Prédisez la demande en électricité



Contexte

- ❑ Mission pour une société spécialisée dans les énergies renouvelables



- ❑ Ces énergies sont **intermittentes** : **difficile** de prévoir les capacités **de production de l'électricité** 

- ❑ La demande en électricité **varie au cours du temps** (météo, température, luminosité..)



- ❑ **Stratégie** : Créer un **modèle** permettant de prévoir la consommation sur une année



- ❑ **But** : Répondre au **besoin d'électricité** (adéquation entre l'offre et la demande)

Construction de différents modèles



Données

Mission 1

Mission 2

Mission 3

Conclusion

Données à ma disposition :



Réseau de transport d'électricité

Consommation électrique de 2010 à 2019

	Mois	qualite	territoire	consommation
13	2010-01	Données consolidées	France	56342
14	2010-02	Données consolidées	France	48698

df_energie

Degrés Jour Unifiés pour le chauffage de 2010 à 2018 pour la France

Unnamed: 0	JAN	FÉV	MAR	AVR	MAI	JUN	JUI	AOÛ	SEP	OCT	NOV	DÉC	TOTAL	
0	2010	624.8	474.7	414.9	292.7	260.6	112.6	46.5	93.8	163.6	286.4	419.3	644.1	3833.8
1	2011	507.2	408.6	368.7	202.0	165.6	117.2	122.5	84.7	101.9	232.3	338.0	424.8	3073.4

df_chauffage

Degrés Jour Unifiés pour la climatisation de 2010 à 2018 pour la France

Unnamed: 0	JAN	FÉV	MAR	AVR	MAI	JUN	JUI	AOÛ	SEP	OCT	NOV	DÉC	TOTAL	
0	2010	0	0	0.0	2.6	7.6	27.5	61.1	24.4	4.4	1.9	0	0	129.4
1	2011	0	0	0.1	7.8	10.8	27.7	11.7	28.8	26.8	6.3	0	0	120.0

df_clim



CEGIBAT
L'expertise efficacité énergétique de GRDF

Données

Mission 1

Mission 2

Mission 3

Conclusion

Nettoyage de la Dataframe df_energie :

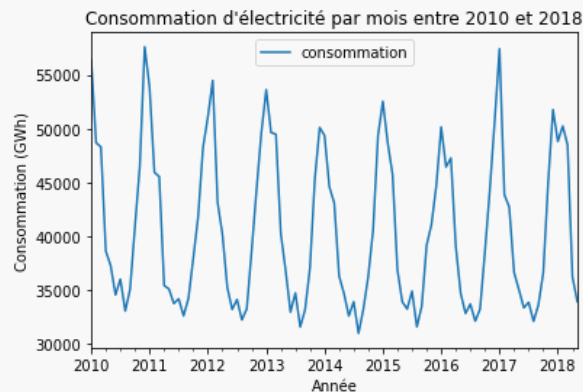
Sélection du territoire : France

	Mois	qualite	territoire	consommation
0	0000-00	Données consolidées	Grand-Est	3364
1	0000-00	Données définitives	Nouvelle-Aquitaine	3222
2	0000-00	Données définitives	Auvergne-Rhône-Alpes	5010
3	0000-00	Données définitives	Bourgogne-Franche-Comté	1533
4	0000-00	Données définitives	Bretagne	1493



	Mois	consommation
0	2010-01-01	56342
1	2010-02-01	48698
2	2010-03-01	48294
3	2010-04-01	38637
4	2010-05-01	37284
...
96	2018-01-01	48807
97	2018-02-01	50236
98	2018-03-01	48484
99	2018-04-01	36236
100	2018-05-01	33949

101 rows x 2 columns



Suppression des mois 0000

Suppression des dates non en
communs avec le DJU (< 2018-06-01)



Données

Mission 1

Mission 2

Mission 3

Conclusion

Nettoyage de la Dataframe df_chauffage et df_clim :

Addition des valeurs dju chauffage et clim

Degrés jours unifiés

- Différence entre la température extérieure et une température de référence : **estimation de la consommation d'En thermique**

Unnamed: 0	JAN	FÉV	MAR	AVR	MAI	JUN	JUI	AOÛ	SEP	OCT	NOV	DÉC	TOTAL
2010	624.8	474.7	414.9	292.7	260.6	112.6	46.5	93.8	163.6	286.4	419.3	644.1	3833.8
2011	507.2	408.6	368.7	202.0	165.6	117.2	122.5	84.7	101.9	232.3	338.0	424.8	3073.4



Unnamed: 0	JAN	FÉV	MAR	AVR	MAI	JUN	JUI	AOÛ	SEP	OCT	NOV	DÉC	TOTAL
2010	0	0	0.0	2.6	7.6	27.5	61.1	24.4	4.4	1.9	0	0	129.4
2011	0	0	0.1	7.8	10.8	27.7	11.7	28.8	26.8	6.3	0	0	120.0



Unnamed: 0	JAN	FÉV	MAR	AVR	MAI	JUN	JUI	AOÛ	SEP	OCT	NOV	DÉC	TOTAL
4020	624.8	474.7	414.9	295.3	268.2	140.1	107.6	118.2	168.0	288.3	419.3	644.1	3963.2
4022	507.2	408.6	368.8	209.8	176.4	144.9	134.2	113.5	128.7	238.6	338.0	424.8	3193.4

	Annee	01	02	03	04	05	06	07	08	09	10	11	12
0	2010	624.8	474.7	414.9	295.3	268.2	140.1	107.6	118.2	168.0	288.3	419.3	644.1
1	2011	507.2	408.6	368.8	209.8	176.4	144.9	134.2	113.5	128.7	238.6	338.0	424.8
2	2012	454.5	555.6	335.4	344.5	205.0	143.0	119.6	112.8	175.6	263.2	403.3	455.7



Colonne renommé et l'année
est divisé par 2



Données

Mission 1

Mission 2


Mission 3

Conclusion


Nettoyage de la Dataframe df_chauffage et df_clim :

Transformation en datetime

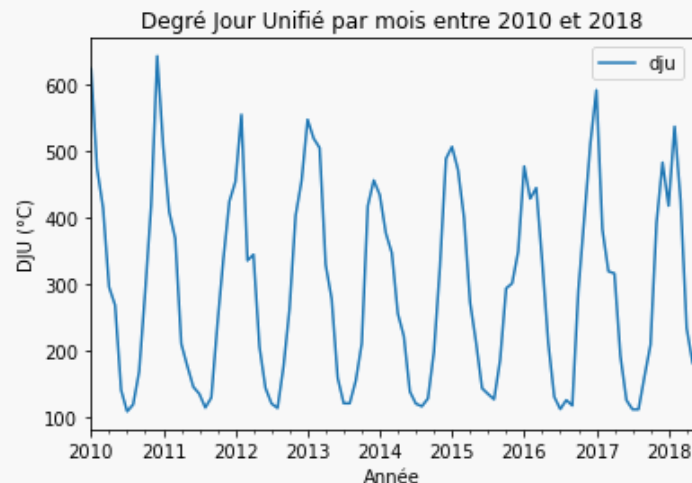
```
for Annee in dju_total.index.values:  
    for Mois in dju_total.columns:  
        dju_final['Mois'].append(f"{Annee}-{Mois}-01")  
        dju_final['dju'].append(dju_total.loc[Annee,Mois])
```



	Mois	dju
0	2010-01-01	624.8
1	2010-02-01	474.7
2	2010-03-01	414.9
3	2010-04-01	295.3
4	2010-05-01	268.2
...
103	2018-08-01	0.0
104	2018-09-01	0.0
105	2018-10-01	0.0
106	2018-11-01	0.0
107	2018-12-01	0.0



Suppression des dates
< 2018-06-01



Données

Mission 1

Mission 2

Mission 3

Conclusion

Merge des Dataframes :

	Mois	consommation
0	2010-01-01	56342
1	2010-02-01	48698
2	2010-03-01	48294

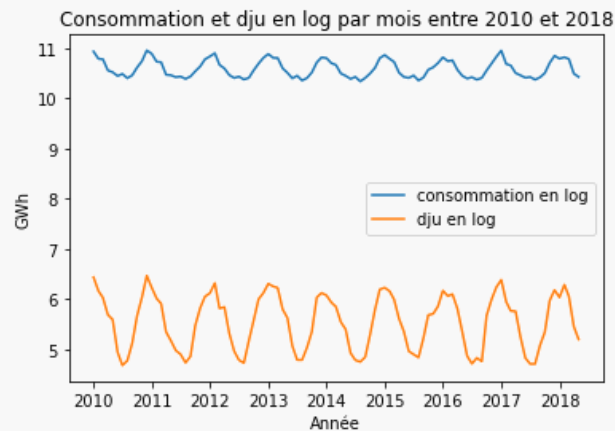
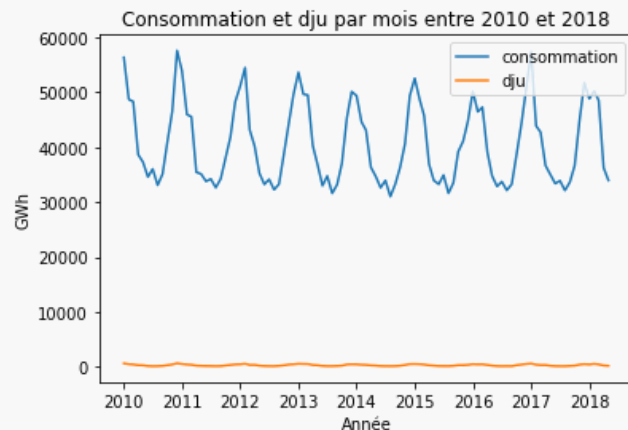


	Mois	dju
0	2010-01-01	624.8
1	2010-02-01	474.7
2	2010-03-01	414.9

Merge sur les Mois



	consommation	dju
Mois		
2010-01-01	56342	624.8
2010-02-01	48698	474.7
2010-03-01	48294	414.9



Données

Mission 1

Mission 2

Mission 3

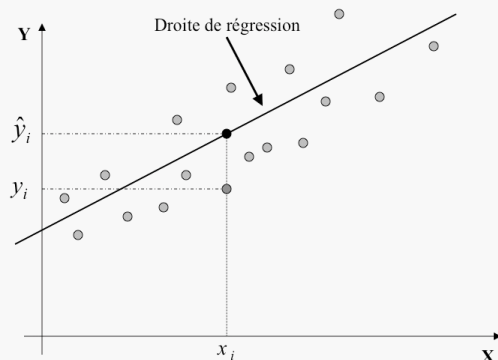
Conclusion

Correction de l'effet température avec une régression linéaire :

➔ Régression linéaire

Régression linéaire

- **Méthode statistique de prédiction** : une variable (X) expliquée est modélisée par une fonction affine d'une autre variable (y)



OLS : Ordinary Least Squares

- **Technique d'analyse** : comparaison de la différence entre les points du data set et les prédictions afin de mesurer l'erreur

OLS Regression Results					
Dep. Variable:	consommation		R-squared:	0.952	
Model:	OLS		Adj. R-squared:	0.952	
Method:	Least Squares		F-statistic:	1983.	
Date:	Thu, 19 Aug 2021		Prob (F-statistic):	2.69e-67	
Time:	10:01:11		Log-Likelihood:	-887.64	
No. Observations:	101		AIC:	1779.	
Df Residuals:	99		BIC:	1785.	
Df Model:	1				
Covariance Type:	nonrobust				
	coef	std err	t	P> t	[0.025 0.975]
const	2.638e+04	356.879	73.919	0.000	2.57e+04 2.71e+04
dju	48.0958	1.080	44.533	0.000	45.953 50.239
Omnibus:	0.588		Durbin-Watson:		1.691
Prob(Omnibus):	0.745		Jarque-Bera (JB):		0.266
Skew:	-0.103		Prob(JB):		0.875
Kurtosis:	3.144		Cond. No.		739.



Données

Mission 1

Mission 2

Mission 3

Conclusion

Correction de l'effet température avec une régression linéaire :

Résultats

OLS Regression Results

```

=====
Dep. Variable:      consommation      R-squared:      0.952
Model:              OLS               Adj. R-squared: 0.952
Method:             Least Squares     F-statistic:    1983.
Date:               Thu, 19 Aug 2021  Prob (F-statistic): 2.69e-67
Time:               10:01:11          Log-likelihood: -887.64
No. Observations:   101              AIC:             1779.
Df Residuals:       99               BIC:             1785.
Df Model:            1
Covariance Type:    nonrobust
=====

```

```

=====
coef      std err      t      P>|t|      [0.025      0.975]
-----
const     2.638e+04    356.879    73.919    0.000     2.57e+04    2.71e+04
dju        48.0958      1.080     44.533    0.000     45.953     50.239
=====

```

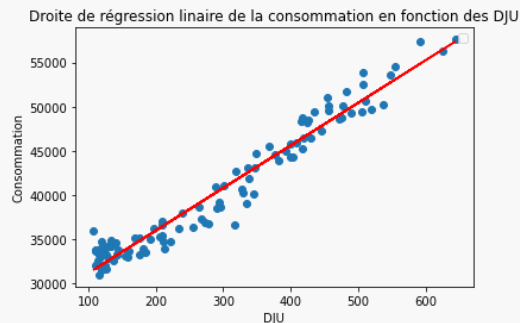
```

=====
Omnibus:      0.588    Durbin-Watson:      1.691
Prob(Omnibus): 0.745    Jarque-Bera (JB):    0.266
Skew:         -0.103   Prob(JB):            0.875
Kurtosis:     3.144    Cond. No.            739.
=====

```

Indique la multicollinéarité : plusieurs variables donnent la même information

Pas de multicollinéarité



Explique à 95 % la variabilité de la consommation

Coefficients de la fonction affine

Significativité : P-value < 0.05

Toutes les variables sont significatives



Données

Mission 1

Mission 2

Mission 3

Conclusion

Correction de l'effet température avec une régression linéaire :

➔ Test sur les résidus

Résidus

- Différence entre la valeur observée et la valeur prédite du modèle

Test shapiro

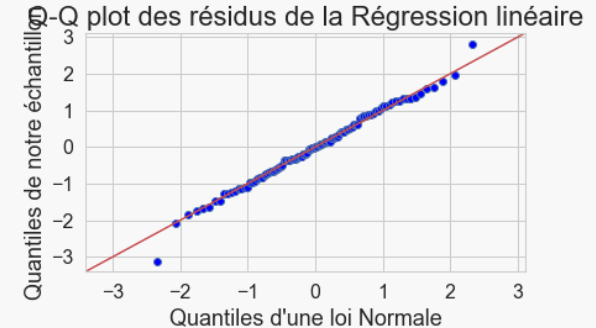
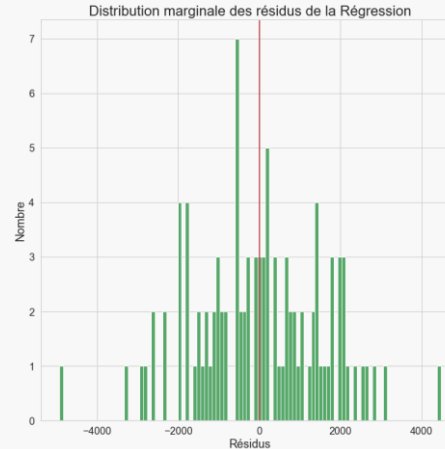
p-value = 0.96

H₀ : L'échantillon est normalement distribué

H_a : L'échantillon n'est pas normalement distribué

p value > 0.05 : L'échantillon suit une loi normale

Normalité : distribution normale



Les résidus suivent une loi normale



Données

Mission 1

Mission 2

Mission 3

Conclusion

Correction de l'effet température avec une régression linéaire :

➔ Test sur les résidus

Test d'homoscédasticité

Test Goldfeld et Quandt

p-value = 0.40

H0 : Les échantillons possèdent une variance égale

Ha : Les échantillons possèdent des variances différentes

p value > 0.05 : Il y a homoscédasticité des résidus

Test de corrélation

Test Durbin-Watson

r = 1,69

r -> 2 : Il n'y a pas d'auto-corrélation

Il y a homoscédasticité et il n'y a pas d'auto-corrélation

Données

Mission 1

Mission 2

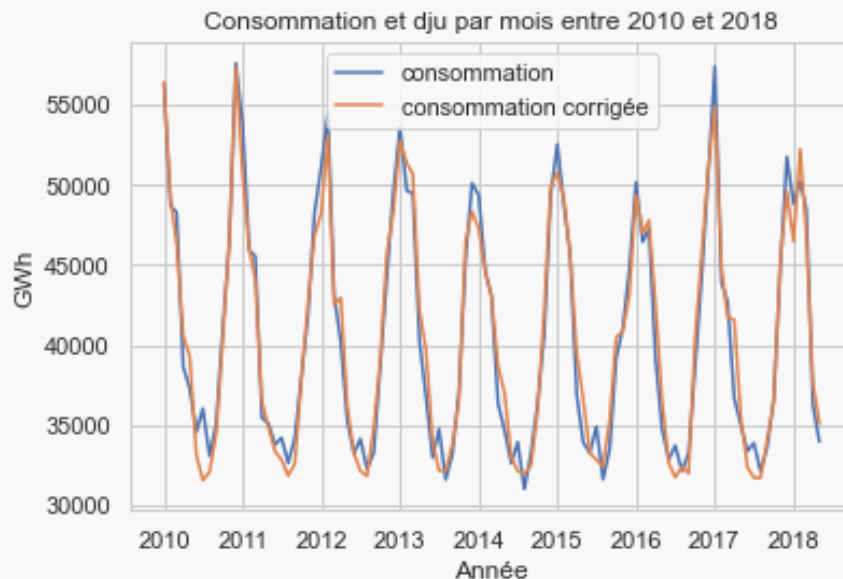
Mission 3

Conclusion

Correction de l'effet température avec une régression linéaire :

➔ Récupération des consommations prédites par le modèle

	consommation	dju	conso_corrige
Mois			
2010-01-01	56342	624.8	56430.430531
2010-02-01	48698	474.7	49211.254213
2010-03-01	48294	414.9	46335.126673
2010-04-01	38637	295.3	40582.871593
2010-05-01	37284	268.2	39279.476002
...
2018-01-01	48807	417.9	46479.414008
2018-02-01	50236	537.6	52236.478666
2018-03-01	48484	426.6	46897.847279
2018-04-01	36236	233.6	37615.362074
2018-05-01	33949	180.9	35080.714559



Données

Mission 1

Mission 2

Mission 3

Conclusion

Correction de l'effet température avec une régression linéaire :

➔ Test de stationnarité

Série temporelle stationnaire

- Les propriétés statistiques (espérance, variance, auto-corrélation) ne varient pas dans le temps

Test d'adfuller

p-value = 0.10

H0 : La série comporte une racine unitaire

Ha : Les séries ne comportent pas de racine unitaire (stationnaire)

p value > 0.05 : La série n'est pas stationnaire

Il n'y a pas de stationnarité

Racine unitaire

- Caractéristique qui cause des problèmes d'inférence statistique (non stationnaire)

Données

Mission 1

Mission 2

Mission 3

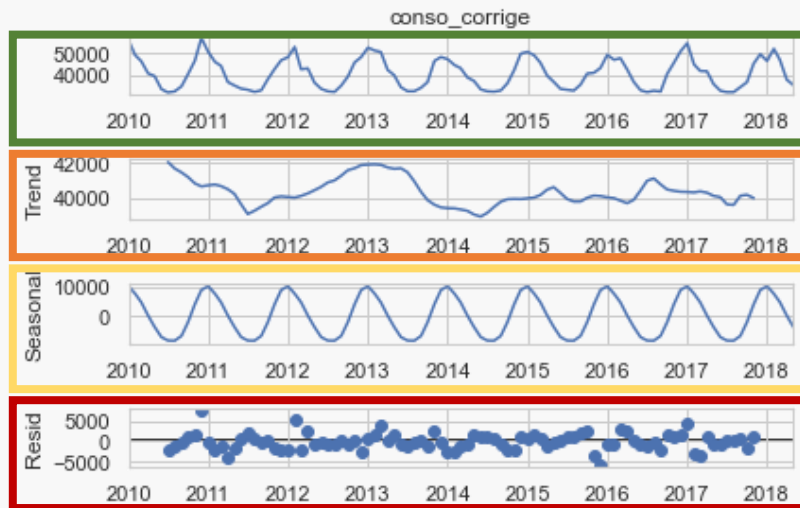
Conclusion

Désaisonnalisation de la consommation corrigée avec les MM :

➔ **Seasonal_decompose**

Seasonal_decompose

- Séparation des séries en composantes : - **La tendance potentielle** (hausse ou baisse de la moyenne) - **La saisonnalité** (un cycle récurrent) – **Les résidus aléatoires**



Consommation corrigée

Tendance globale :
Moyenne mobile

Saisonnalité de 12 mois

Résidus : observé –
(tendance + saisonnalité)

Données

Mission 1

Mission 2

Mission 3

Conclusion

Désaisonnalisation de la consommation corrigée avec les MM :

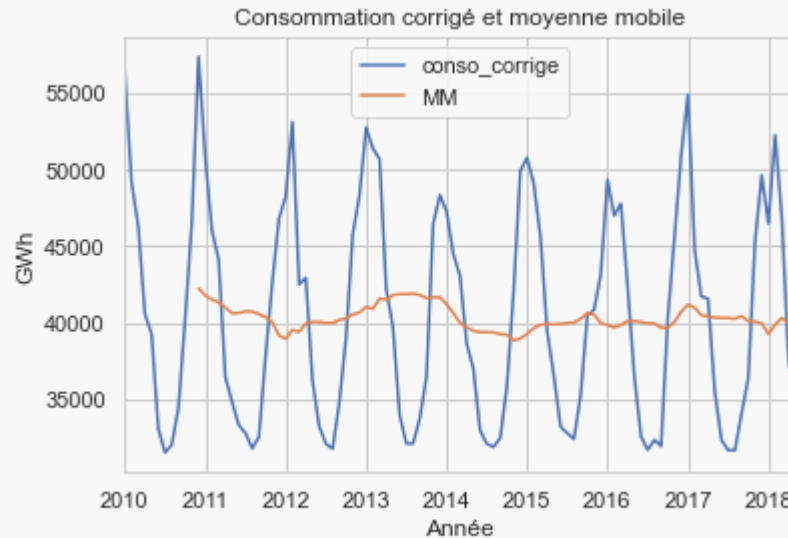
➔ Détermination des moyennes mobiles (MM)

Moyenne mobile

- Moyenne utilisée pour supprimer les fluctuations transitoires (recalculée de façon continue)



	conso_corrige	MM
Mois		
2010-01-01	56430.430531	NaN
2010-02-01	49211.254213	NaN
2010-03-01	46335.126673	NaN
2010-04-01	40582.871593	NaN
2010-05-01	39279.476002	NaN
...
2018-01-01	46479.414008	39279.876800
2018-02-01	52236.478666	39901.113936



Données

Mission 1

Mission 2

Mission 3

Conclusion

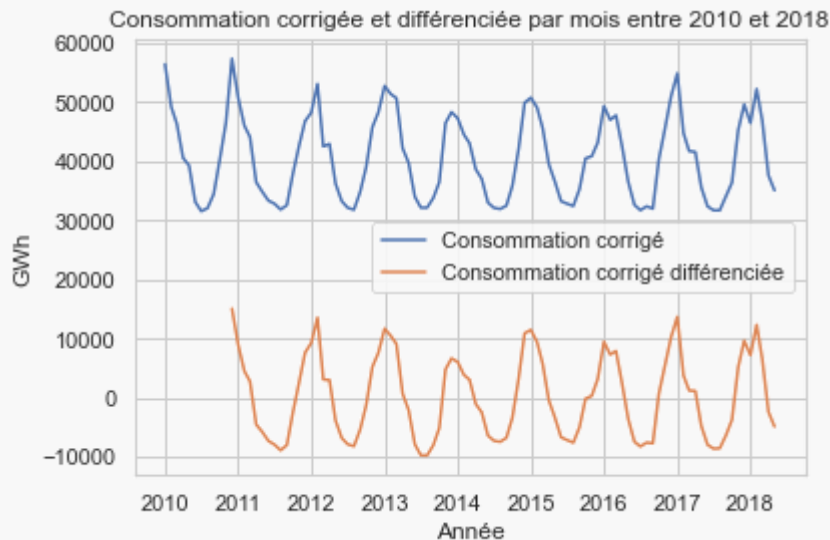
Désaisonnalisation de la consommation corrigée avec les MM :

➔ Différenciation de la moyenne mobile

$$Conso_{corrig_{diff}} = conso_{corrigé} - moyenne_{mobile} (12 \text{ périodes})$$



Mois	conso_corrigé	MM	conso_corrig_diff
2010-01-01	56430.430531	NaN	NaN
2010-02-01	49211.254213	NaN	NaN
2010-03-01	46335.126673	NaN	NaN
2010-04-01	40582.871593	NaN	NaN
2010-05-01	39279.476002	NaN	NaN
...
2018-01-01	46479.414008	39279.876800	7199.537208
2018-02-01	52236.478666	39901.113936	12335.364730
2018-03-01	46897.847279	40331.571151	6566.276127
2018-04-01	37615.362074	40001.313474	-2385.951400
2018-05-01	35080.714559	39960.832861	-4880.118301



Données

Mission 1

Mission 2

Mission 3

Conclusion

Désaisonnalisation de la consommation corrigée avec les MM :

➔ Test de stationnarité de la consommation corrigée différenciée

Test d'adfuller

p-value = $1.19e-9$

H_0 : La série comporte une racine unitaire

H_a : Les séries ne comportent pas de racine unitaire (stationnaire)

$p\text{ value} < 0.05$: La série est stationnaire

La consommation corrigée
différenciée est stationnaire

Données

Mission 1

Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec le lissage exponentiel :

➔ Lissage exponentiel

Lissage exponentiel

- Méthode de lissage et de prévision de données chronologiques : **3 types**

Additif

La série est la somme de ses composantes

Multiplicative

La série est le produit de ses composantes

Lissage exponentiel simple (SES)

Ce lissage suppose que la série n'a pas de changement (tendance et saisonnalité)

Lissage exponentiel double (HES)

Ce lissage permet à la série d'avoir une composante de tendance

Lissage exponentiel triple (WES)

Ce lissage permet à la série d'avoir une composante de tendance et d'inclure la saisonnalité

Données

Mission 1

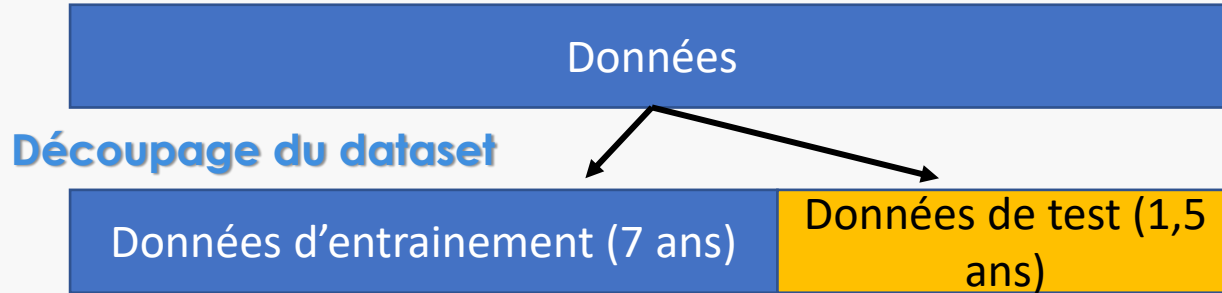
Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec le lissage exponentiel :

➡ Préparation des données



Entraînement du modèle :
ExponentialSmoothing.fit (train)

➡ Prédiction

Données

Mission 1

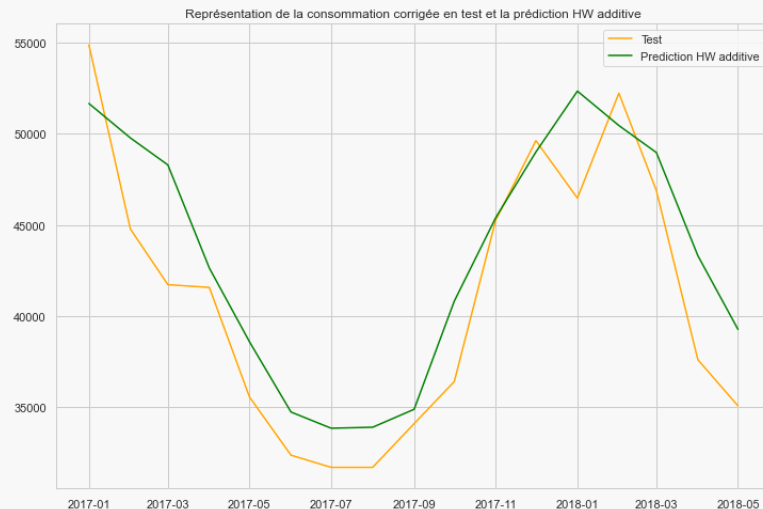
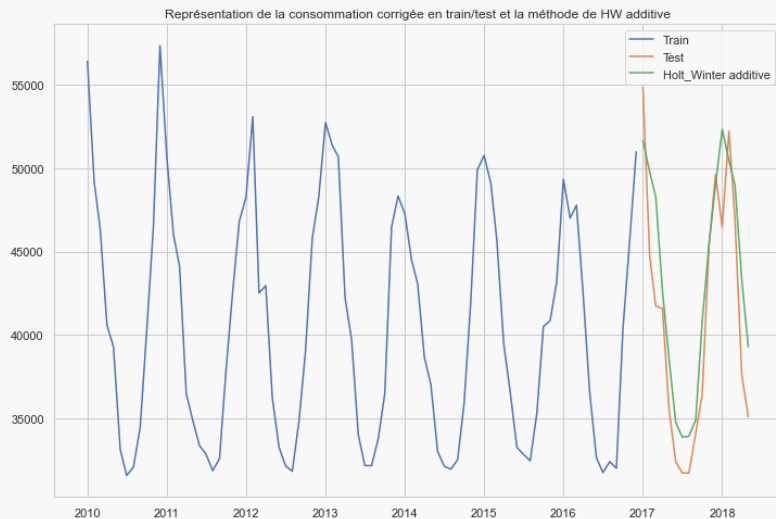
Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec le lissage exponentiel :

➔ **Lissage exponentiel triple : Holt Winter additif**



Indicateurs d'écart

- Comparer les observations avec les prédictions

Erreur quadratique moyenne

12727394 Gwh²

Erreur moyenne absolue

3009 Gwh

Erreur moyenne en pourcentage

7.6 %

Données

Mission 1

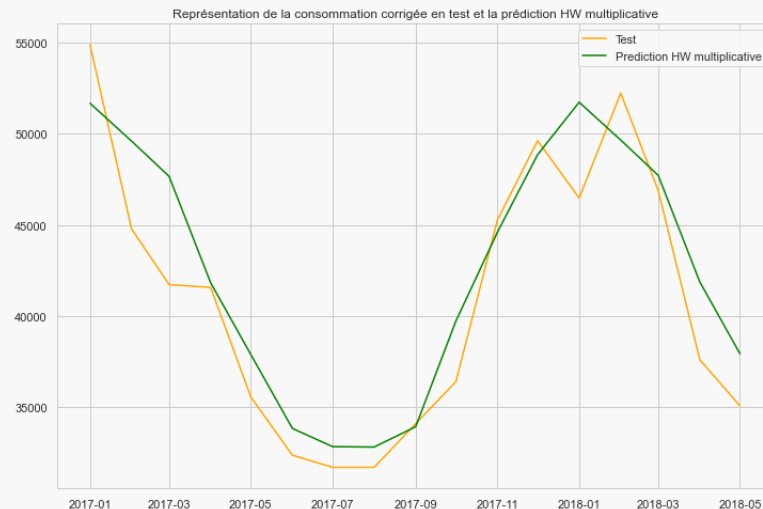
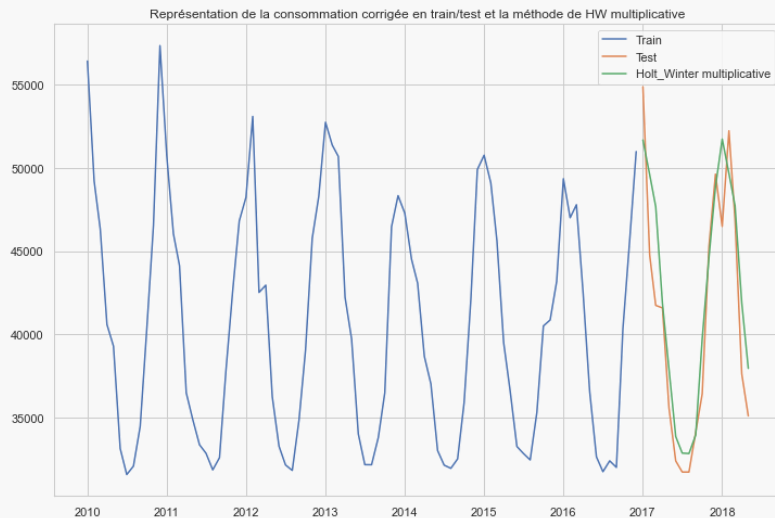
Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec le lissage exponentiel :

➔ **Lissage exponentiel triple : Holt Winter multiplicatives**



Erreur quadratique moyenne

8969770 Gwh²

Erreur moyenne absolue

2411 Gwh

Erreur moyenne en pourcentage

5.9 %

Données

Mission 1

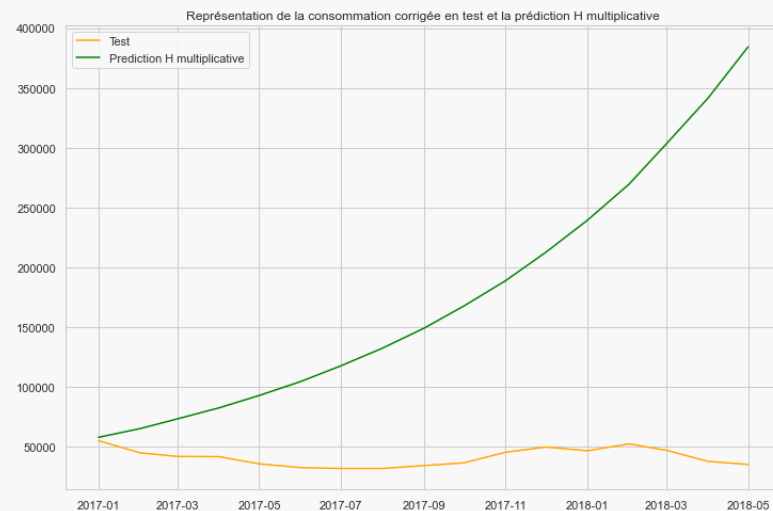
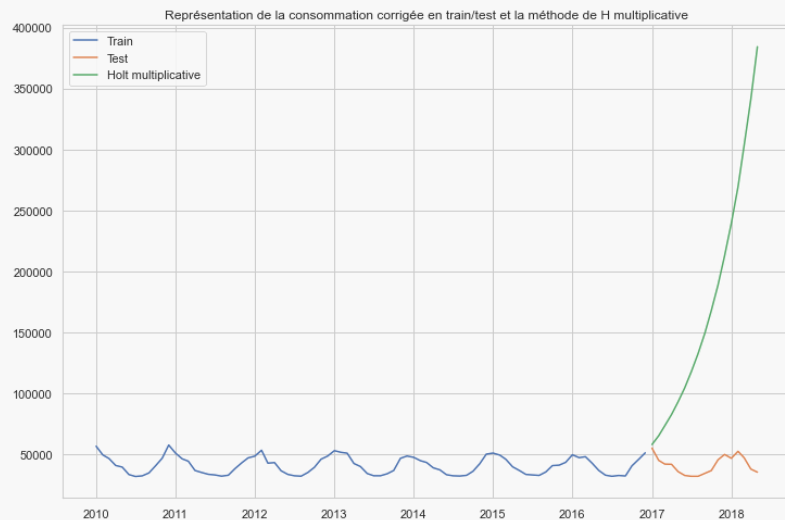
Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec le lissage exponentiel :

➔ **Lissage exponentiel double : Holt multiplicative**



Erreur quadratique moyenne

27703914871 Gwh²

Erreur moyenne absolue

134306 Gwh

Erreur moyenne en pourcentage

336 %



Données

Mission 1

Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec la méthode SARIMA :

➔ Modèle ARIMA

Integrated - value denoted by d

Auto Regressive - value
denoted by p

Moving Average - value
denoted by q

AR I MA

Modèle ARIMA

- Modèles de prévision qui visent à décrire les autocorrélations dans la série

SARIMA

Extension d'AMIRA qui prend en charge la composante saisonnière

Données

Mission 1

Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec la méthode SARIMA :

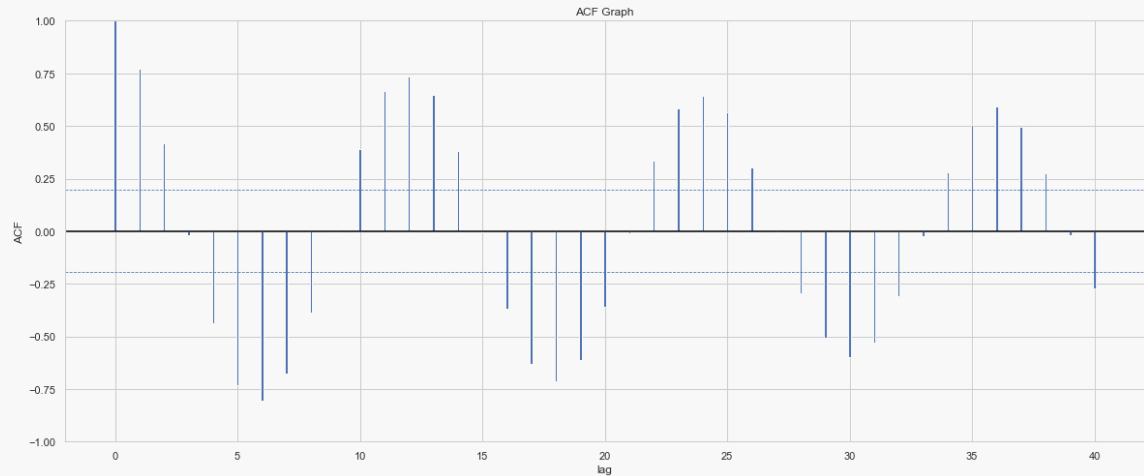
➔ Détermination de l'ordre de différenciation

ACF

- **Fonction d'autocorrélation :** description de l'autocorrélation entre une observation et une observation à un temps antérieur

Différenciation

- Remplacement de la série originale par la série des différences de points adjacents



**Autocorrélation pour un grand nombre de décalages :
nécessite une différenciation**

Données

Mission 1

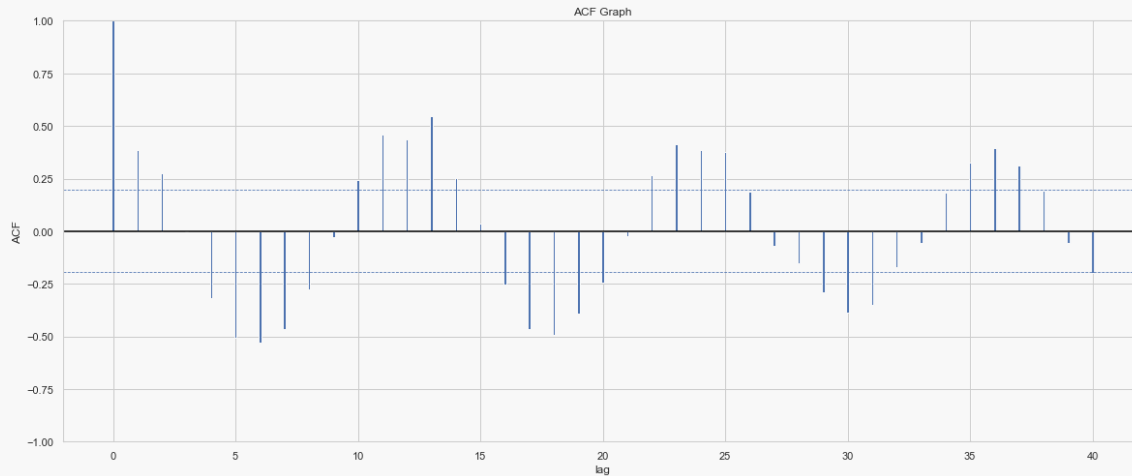
Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec la méthode SARIMA :

➡ Différenciation I – B



Encore une autocorrélation pour un grand nombre de décalages : différenciation

Données

Mission 1

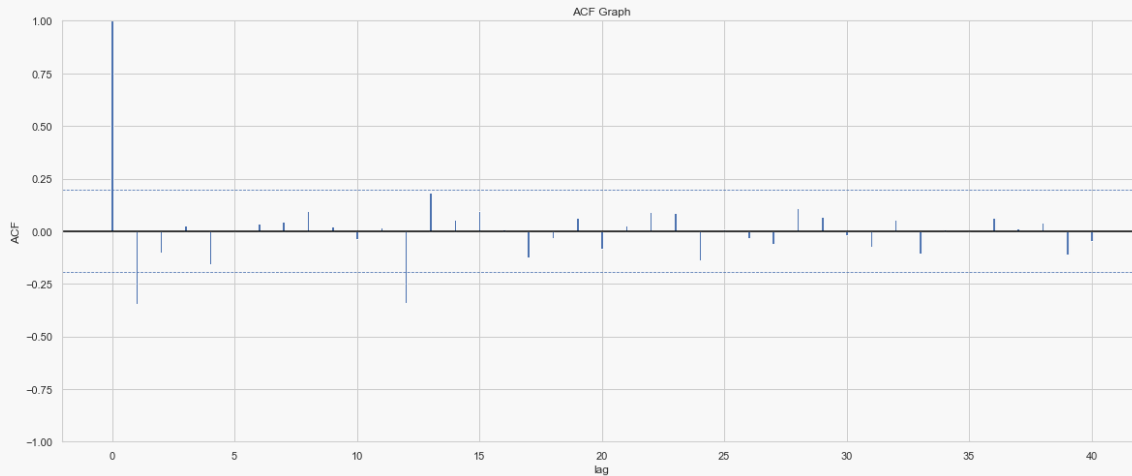
Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec la méthode SARIMA :

➔ Différenciation I – B12



Test d'adfuller

p-value = 8.85e-11

H0 : La série comporte une racine unitaire

Ha : Les séries ne comportent pas de racine unitaire (stationnaire)

p value < 0.05 La série est stationnaire

Autocorrélation décline rapidement vers 0 : Série stationnaire

Données

Mission 1

Mission 2

Mission 3

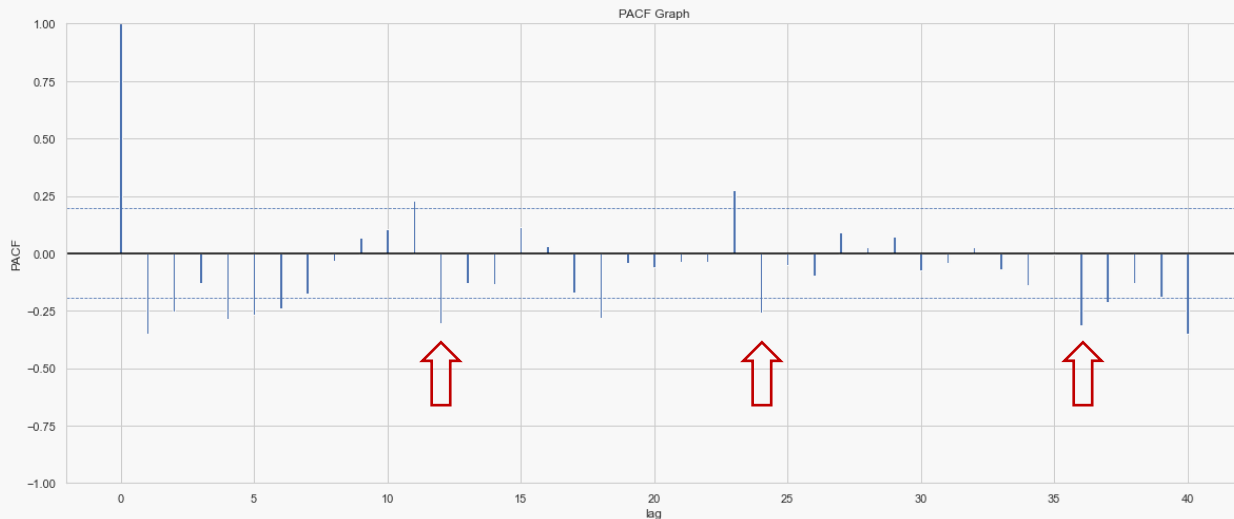
Conclusion

Prévision de la consommation corrigée avec la méthode SARIMA :

➔ Différenciation I – B12

PACF

- **Fonction d'autocorrélation partielle** : description de la relation directe entre une observation et son décalage



Autocorrélation significative à 12, 24, 36 mois : période saisonnière de 12 mois



Données

Mission 1

Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec la méthode SARIMA :

→ Détermination du meilleur modèle SARIMA

```
smodel = pm.auto_arima(train_conso['conso_corrige'],  
                        test='adf',  
                        m=12,  
                        seasonal=True,  
                        d=0, D=1, trace=True,  
                        error_action='ignore',  
                        suppress_warnings=True,  
                        stepwise=True)
```

```
smodel.summary()
```



ARIMA(1,0,2)(0,1,1)[12]	: AIC=1349.086, Time=0.51 sec
ARIMA(1,0,2)(0,1,0)[12]	: AIC=1379.345, Time=0.09 sec
ARIMA(1,0,2)(1,1,1)[12]	: AIC=1355.942, Time=0.64 sec
ARIMA(1,0,2)(0,1,2)[12]	: AIC=1356.057, Time=1.10 sec
ARIMA(1,0,2)(1,1,0)[12]	: AIC=1359.889, Time=0.31 sec
ARIMA(1,0,2)(1,1,2)[12]	: AIC=1356.239, Time=1.27 sec
ARIMA(0,0,2)(0,1,1)[12]	: AIC=1366.459, Time=0.09 sec
ARIMA(1,0,3)(0,1,1)[12]	: AIC=1354.601, Time=0.28 sec
ARIMA(0,0,3)(0,1,1)[12]	: AIC=1367.950, Time=0.11 sec
ARIMA(2,0,3)(0,1,1)[12]	: AIC=1370.288, Time=0.36 sec

Best model: **ARIMA(1,0,2)(0,1,1)[12]**
Total fit time: 30.035 seconds

auto.arima ()

- Approche par étape pour rechercher plusieurs combinaisons des paramètres p,d,q

AIC

- **Akaike's Information Criterion** : estime le modèle par une méthode de maximum de vraisemblance (pénalise les modèles comportant trop de variables)

Sélection du modèle avec le AIC le plus faible

Données

Mission 1

Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec la méthode SARIMA :

➔ Modèle SARIMA

SARIMAX Results

```

=====
Dep. Variable:                  y      No. Observations:          84
Model:          SARIMAX(1, 0, 2)x(0, 1, [1], 12)  Log Likelihood          -669.543
Date:                   Sat, 21 Aug 2021      AIC          1349.086
Time:                   11:09:07      BIC          1360.469
Sample:                   0      HQIC          1353.618
                             - 84
Covariance Type:            opg
=====

```

➔ AIC le plus faible

```

=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.L1          0.9513      0.012      79.071      0.000      0.928      0.975
ma.L1         -0.8717      0.069     -12.676      0.000     -1.006     -0.737
ma.L2         -0.0183      0.076      -0.241      0.809     -0.167      0.131
ma.S.L12      -0.4237      0.073      -5.820      0.000     -0.566     -0.281
sigma2         6.742e+06      1.86e-09      3.62e+15      0.000      6.74e+06      6.74e+06
=====
Ljung-Box (L1) (Q):                3.70      Jarque-Bera (JB):                1.09
Prob(Q):                          0.05      Prob(JB):                0.58
Heteroskedasticity (H):            0.84      Skew:                    -0.28
Prob(H) (two-sided):              0.66      Kurtosis:                 3.20
=====

```

Données

Mission 1

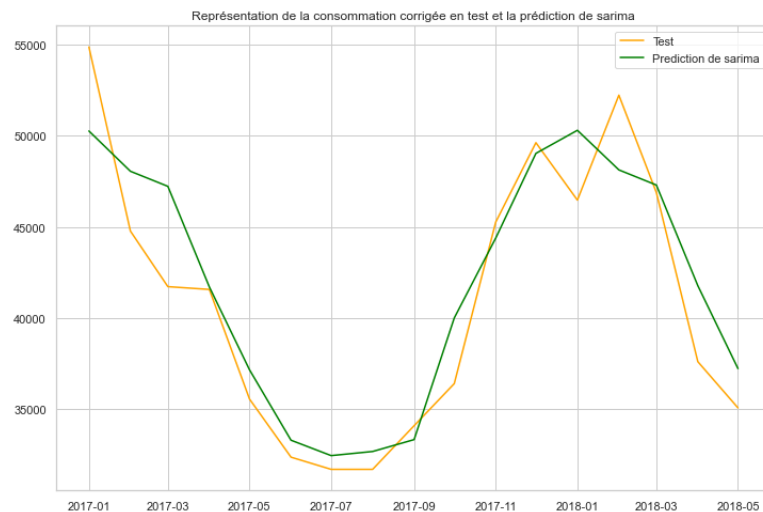
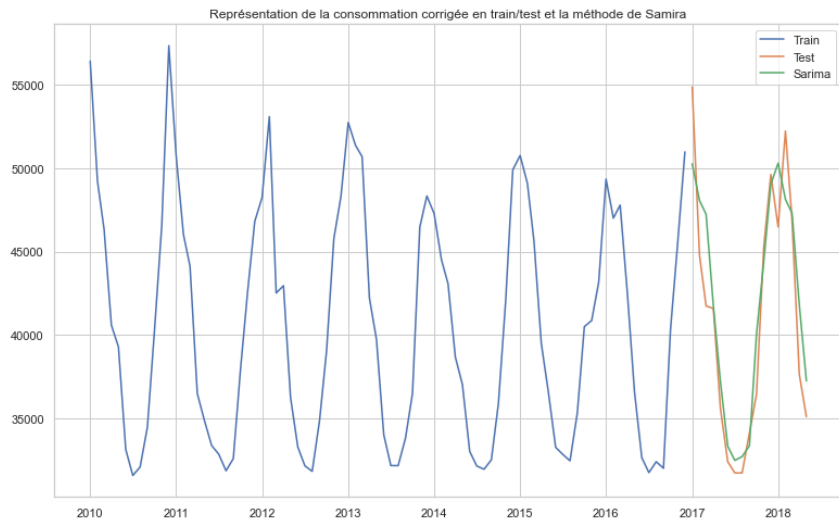
Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec la méthode SARIMA :

Visualisation du modèle SAMIRA



Erreur quadratique moyenne

7976462 Gwh²

Erreur moyenne absolue

2252 Gwh

Erreur moyenne en pourcentage

5.4 %

Données

Mission 1

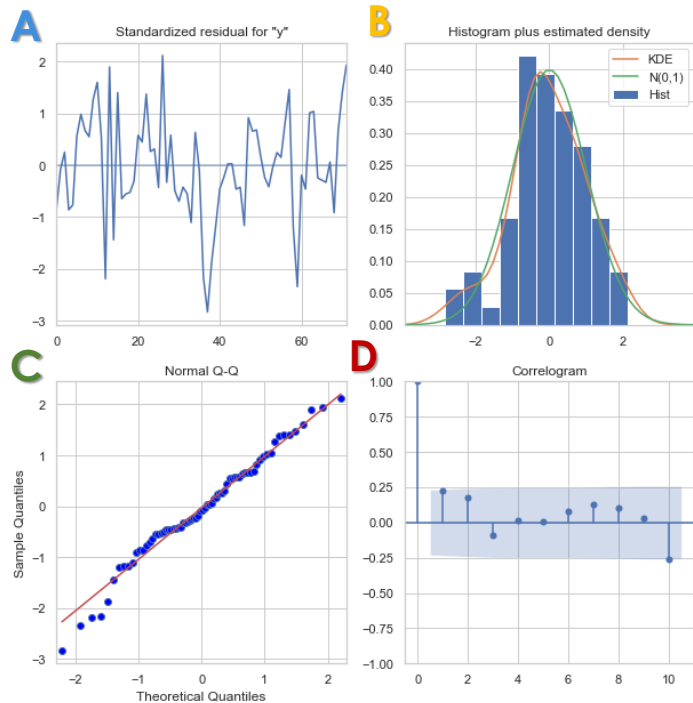
Mission 2

Mission 3

Conclusion

Prévision de la consommation corrigée avec la méthode SARIMA :

Analyse des résidus



A

Les résidus fluctuent autour d'une moyenne de 0

B

C

Les résidus suivent une loi normale

D

Il n'y a pas d'autocorrélation des résidus

Données

Mission 1

Mission 2

Mission 3

Conclusion

- ❑ La **régression** linéaire a permis de **corriger l'effet température** **Série non stationnaire**
- ❑ Les **moyennes mobiles** ont permis la **désaisonnalisation** de la série corrigée **Série stationnaire**
- ❑ Le **lissage exponentiel** a permis la **prévision** de la consommation en énergie
- ❑ Le **modèle SARIMA** a permis la **prévision** de la consommation et a nécessité une **différenciation** :
 - Série non différenciée **Série non stationnaire**
 - Série différenciée 1-B **Série non stationnaire**
 - Série différenciée 1-B12 **Série stationnaire**

Données

Mission 1

Mission 2

Mission 3

Conclusion

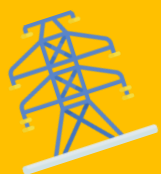
❑ Récapitulatif des différents modèles

	Holt multiplicative	Holt-Winters additif	Holt-Winters multiplicative	SARIMA
Erreur moyenne en pourcentage	336 %	7.6 %	5.9 %	5.4 %
Graphique				

Meilleur modèle : SARIMA



Prévision de la consommation en énergie d'une année pour
l'adéquation entre l'offre et la demande de **ENERCOOP**



Merci pour votre attention

