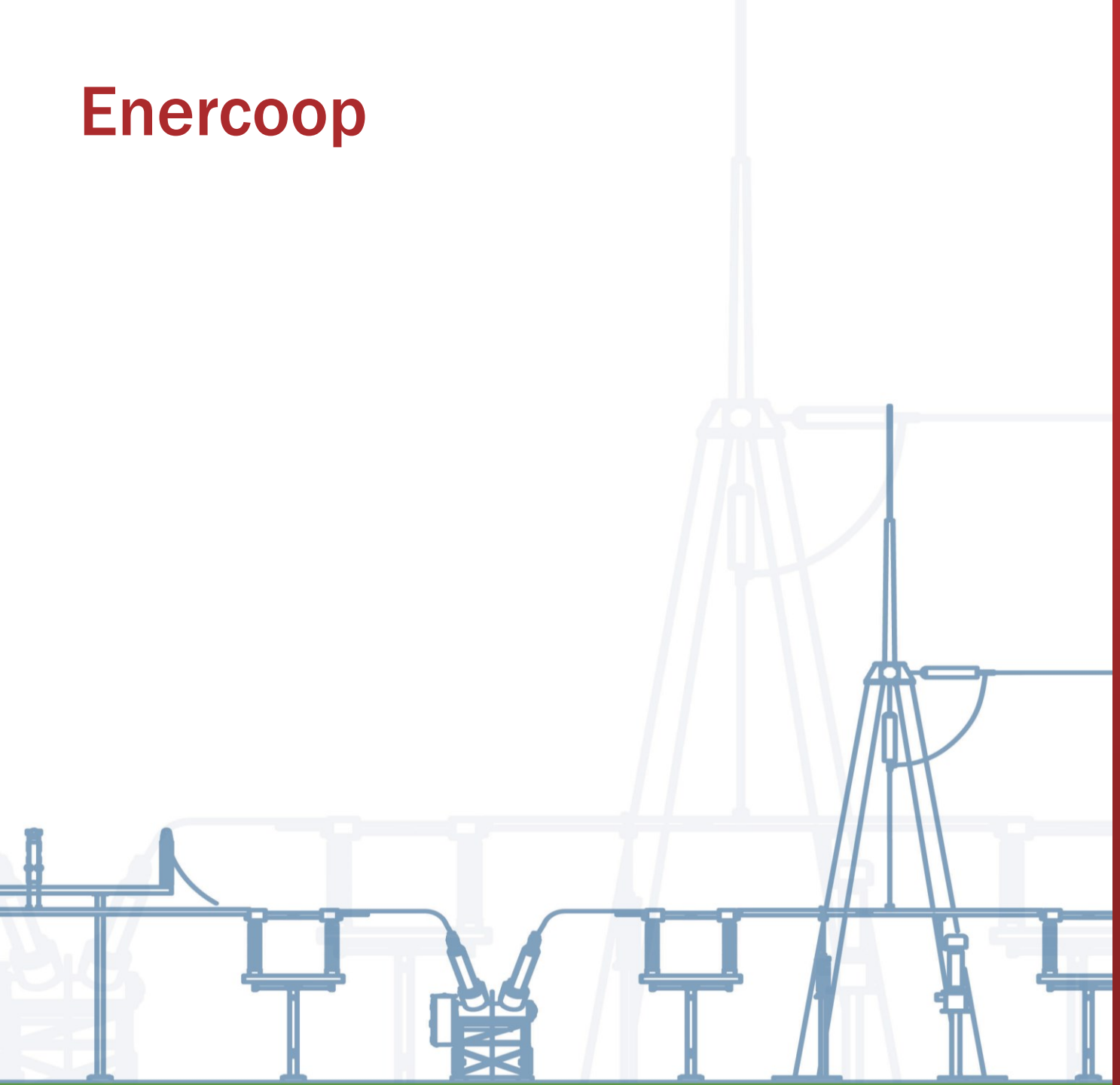
Enercoop



Prédire la demande en électricité.

Projet 9 – DA – Marc SELLAM - 01/2021

Sommaire :

Mission

- Enercoop est une société coopérative qui s'est développée grâce à la libéralisation du marché de l'électricité en France.
- La demande en électricité des utilisateurs varie au cours du temps et dépend de paramètres comme la météo (température, luminosité, etc.) .
- Tout le challenge est de mettre en adéquation l'offre et la demande, en créant un model de prévision de la consommation électrique totale.

Les données fournies

- Téléchargement des données mensuelles de consommation totale d'électricité en énergie sur https://www.rte-france.com/eco2mix/telecharger-les-indicateurs.
  - Consommation totale ile de France en GWh , données mensuelles

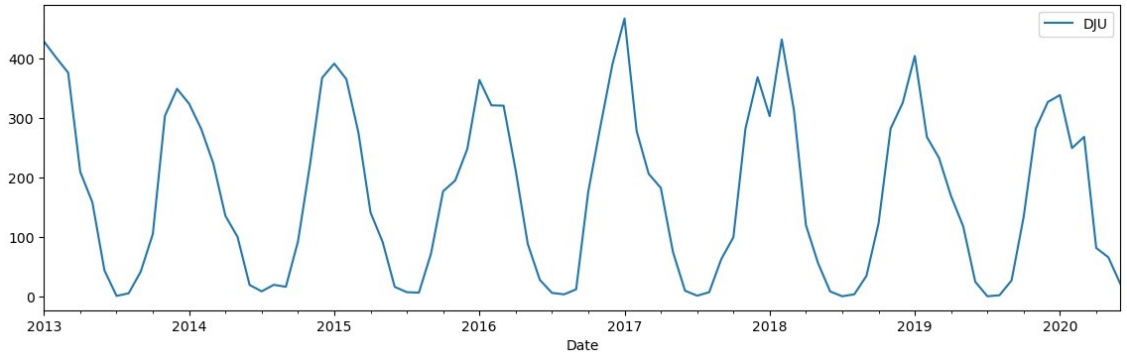
    (unité de mesure d'énergie qui correspond à la puissance d'un gigawatt actif pendant une heure)

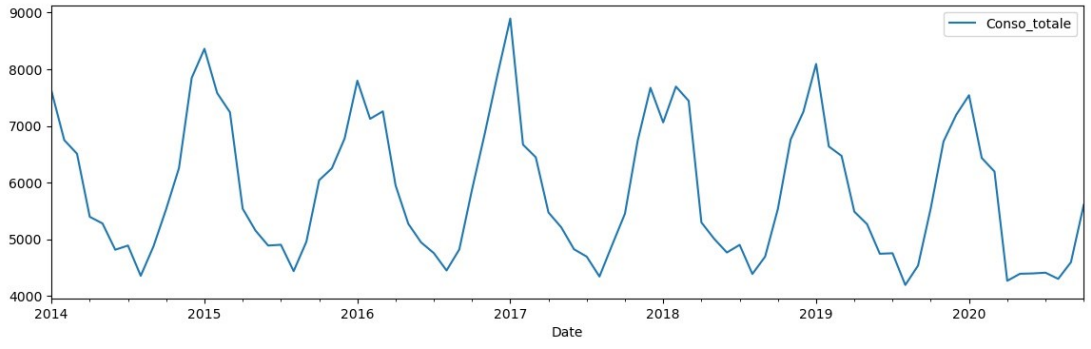
- Données météo utilisées pour corriger les données de l'effet température téléchargées sur https://cegibat.grdf.fr/simulateur/calcul-dju.
  - Dju paris Montsouris, données mensuelles

    (degré Jour Unifié est l'écart entre la température d'une journée donnée et un seuil de température établi à 18 °C)

Les données fournies

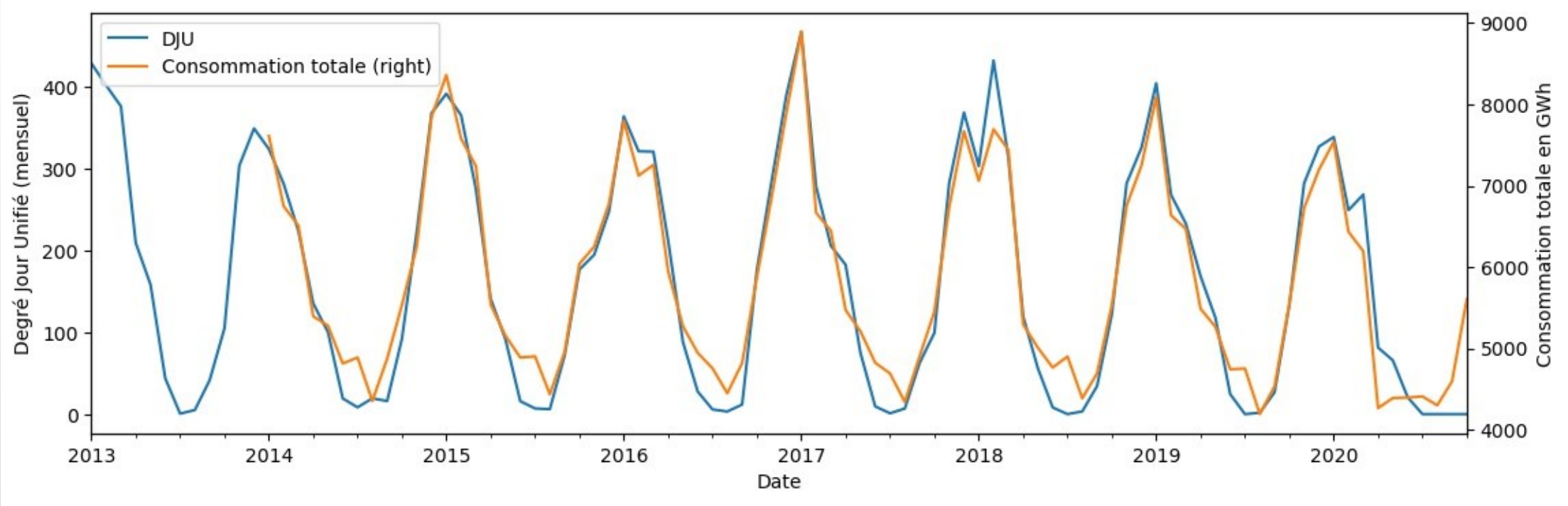
	Année	JAN	FÉV	MAR	AVR	MAI	JUN	JUI	AOÛ	SEP	OCT	NOV	DÉC	Total
11	2020	339	249.6	268.6	81.4	65.7	20.6	0	0	0	0	0	0	1024.9
12	2019	404.9	268.3	233.1	168.5	117.9	24.4	0	1.7	26.7	133.7	282.6	327.3	1989
13	2018	303.4	432.6	314.3	119.7	55.9	8.1	0	3.3	34.3	122.4	282.5	325.9	2002.2
14	2017	467.9	278.4	206.1	182.6	75	9.4	1	6.8	62.6	99.4	282.6	369	2040.6
15	2016	364.4	321.6	321.1	212.1	88.1	27.5	5.7	3.2	11.7	176	285.6	390.8	2207.3
16	2015	392	365.7	275.5	141.1	91.5	15.8	6.9	6.1	71.9	176.9	195	248.1	1986.2
17	2014	324.4	281.9	223.9	135.5	100.2	19.1	8.3	19.3	16	92.3	222.6	368.2	1811.5
18	2013	429.2	402.2	376.6	209.5	158.4	43.6	0.6	5	41.5	105	303.9	349.5	2424.8
19	2012	336	435.9	201.9	230.3	83.3	35	12.4	2.4	58	154.6	296.2	345.9	2191.5
20	2011	392	304.8	243.1	77.6	43.4	31.4	15	11.9	23.2	127.6	226.6	312.7	1809
21	2010	499.2	371.4	294.5	165.3	140.9	22.6	0	11.1	52.3	172.2	310	512	2551.1
22	2009	486.8	365.7	293.2	135.1	82.2	39.8	3.1	0.9	26.9	149.6	224.7	411.8	2219.7





	Date	Conso_totale
19	2013-01	0
32	2013-02	0
45	2013-03	0
58	2013-04	0
71	2013-05	0
...	...	...
1176	2020-06	4397
1189	2020-07	4410
1202	2020-08	4301
1215	2020-09	4595
1228	2020-10	5605

Préparation des données



- Apres concaténation des données :
  - 6 années de données exploitables.
  - Similitude des 2 courbes.

Préparation des données

```
#data_tr =  data_i de 2014 à 2017, annees 'd'entrainement'
data_tr = data_i.loc['2014':'2017'].copy()
#data_te =  data_i de 2018, annee 'adjust'
data_te = data_i.loc['2018'].copy()
#data_te2 =  data_i de 2019, annee 'test'
data_te2 = data_i.loc['2019'].copy()
```

- 6 années :

Les quatre premières pour l'entrainement

La cinquième pour ajuster le modèle automatisé SARIMA et vérifier la qualité des prédictions du modèle manuel .

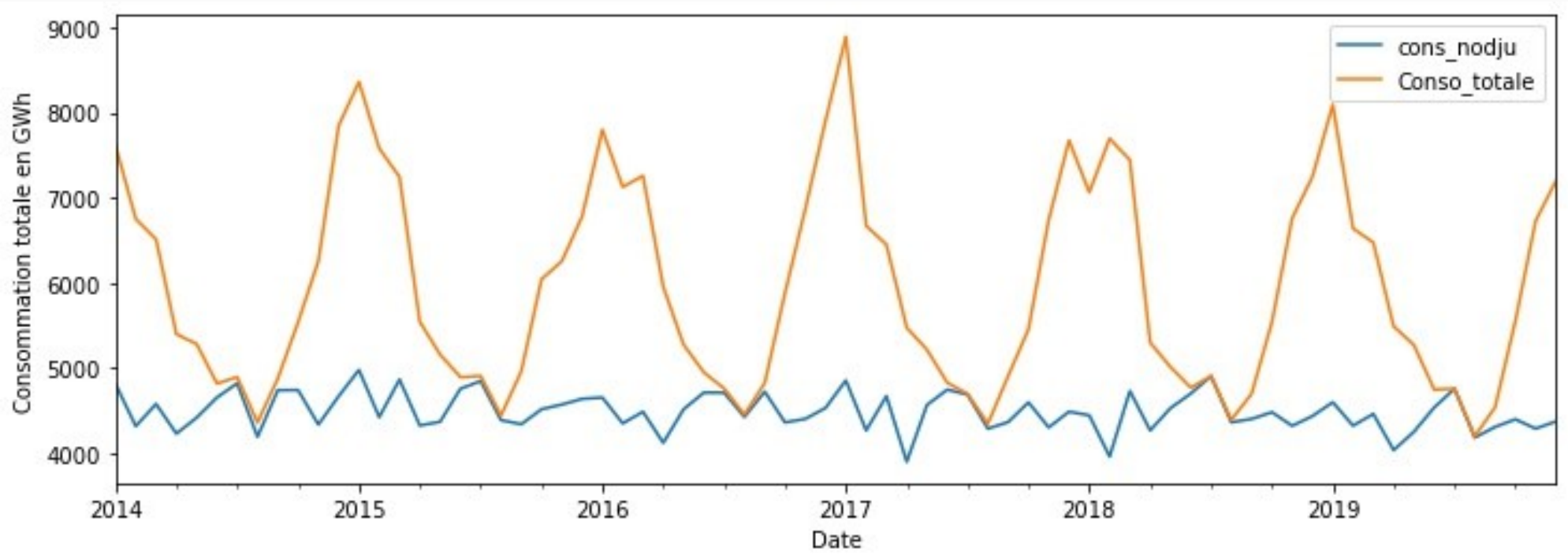
La sixième année pour comparer la qualité des 2 prédictions un an après afin de vérifier leur efficacité à long terme.

Date	DJU	Conso_totale
2014-01-01	324.4	7612
2014-02-01	281.9	6749
2014-03-01	223.9	6509
2014-04-01	135.5	5396
2014-05-01	100.2	5279
...	...	...
2017-08-01	6.8	4342
2017-09-01	62.6	4902
2017-10-01	99.4	5452
2017-11-01	282.6	6740
2017-12-01	369.0	7672

Date	DJU	Conso_totale
2018-01-01	303.4	7062
2018-02-01	432.6	7694
2018-03-01	314.3	7442
2018-04-01	119.7	5297
2018-05-01	55.9	5008
...	...	...
2018-08-01	3.3	4387
2018-09-01	34.3	4694
2018-10-01	122.4	5535
2018-11-01	282.5	6758
2018-12-01	325.9	7248

Date	DJU	Conso_totale
2019-01-01	404.9	8093
2019-02-01	268.3	6637
2019-03-01	233.1	6471
2019-04-01	168.5	5487
2019-05-01	117.9	5266
...	...	...
2019-08-01	1.7	4193
2019-09-01	26.7	4535
2019-10-01	133.7	5549
2019-11-01	282.6	6726
2019-12-01	327.3	7197

Correction des données de consommation mensuelles de l'effet température (dues au chauffage électrique)



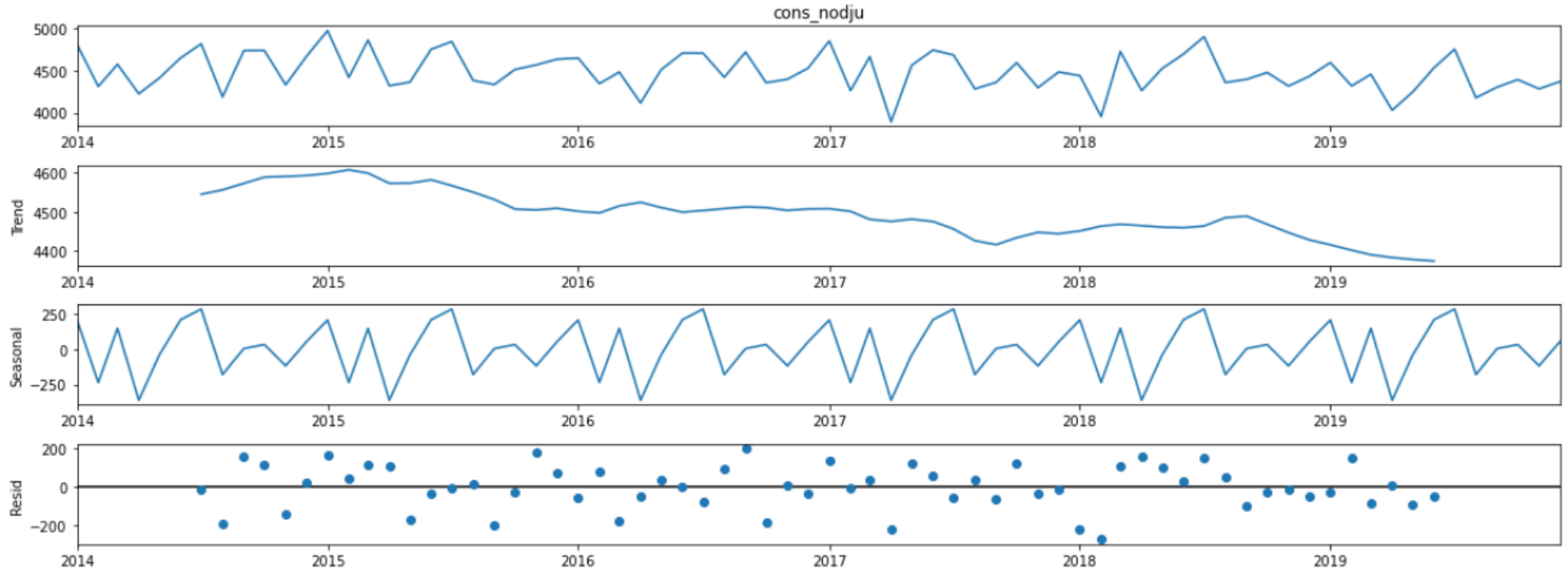

```
x = data_tr[['DJU']].copy()
y = data_tr[['Conso_totale']].copy()
```

```
import statsmodels.api as sm
import statsmodels.formula.api as smf

reg = smf.ols('y~x', data=data_tr).fit()
```

```
coef = reg.params[1]
data_tr['co_dju'] = data_tr['DJU'] * coef
data_tr['cons_nodju'] = data_tr['Conso_totale'] - data_tr['co_dju']
data_te['co_dju'] = data_te['DJU'] * coef
data_te['cons_nodju'] = data_te['Conso_totale'] - data_te['co_dju']
data_te2['co_dju'] = data_te2['DJU'] * coef
data_te2['cons_nodju'] = data_te2['Conso_totale'] - data_te2['co_dju']
data_all = data_tr.copy().append(data_te.copy(), sort=False)
data_all = data_all.copy().append(data_te2.copy(), sort=False)
data_all
```

Désaisonnalisation de la consommation obtenue après correction, grâce aux moyennes mobiles.

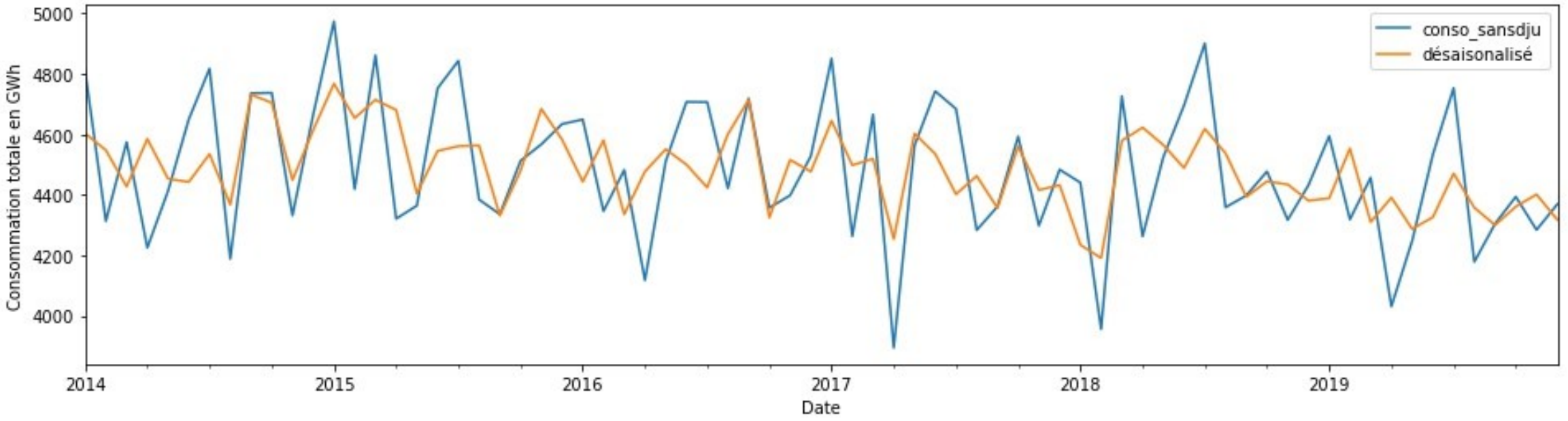
```
from statsmodels.tsa.seasonal import seasonal_decompose
import matplotlib.pyplot as plt
plt.rcParams["figure.figsize"]=[16,6]
decomp_x = seasonal_decompose(data_all['cons_nodju'],  model='additive', period = 12)
fig = decomp_x.plot()
#plt.savefig("2_0.jpg",bbox_inches ="tight")
plt.show()
```



Désaisonnalisation de la consommation obtenue après correction, grâce aux moyennes mobiles.

```
#desaisonalisation :
data_all['cons_nodjusea'] = data_all['cons_nodju'] - decomp_x.seasonal
data_all['cons_nodju'].plot(label='conso_sansdju',figsize=(16,4))
data_all['cons_nodjusea'].plot(label = 'désaisonalisé')
plt.legend()
plt.ylabel("Consommation totale en GWh",fontsize=10)
plt.savefig("P9_03_graphiques.jpg",bbox_inches ="tight")
```

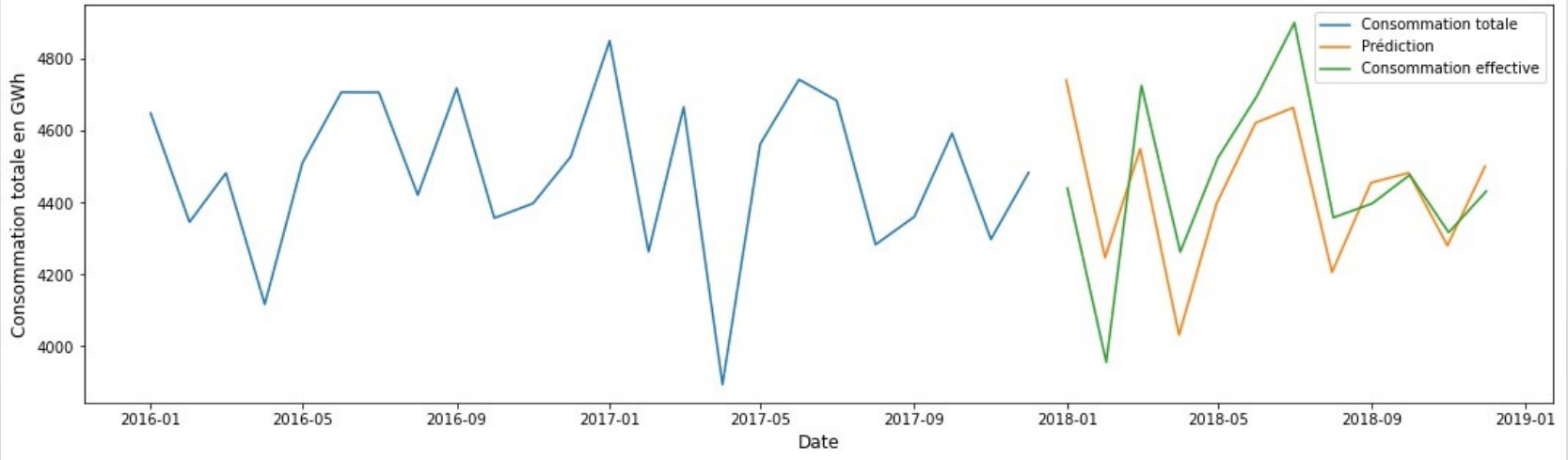
- Désaisonnalisation appliquée aux 6 années



Prévision de la consommation (corrigée de l'effet température) sur un an, en utilisant la méthode de Holt Winters (lissage exponentiel).

```
y = data_tr[['cons_nodju']].copy()
```

```
from statsmodels.tsa.api import ExponentialSmoothing
hw = ExponentialSmoothing(np.asarray(y), seasonal_periods=12, trend='add', seasonal='add').fit()
hw_pred = hw.forecast(12)
plt.figure(figsize=(18,5))
plt.plot(y.loc['2016':], label='Consommation totale')
plt.plot(pd.date_range(y.index[len(y)-1], periods=12, freq='M'), hw_pred, label='Prédiction')
plt.plot(data_all['cons_nodju'].loc['2018'], label='Consommation effective')
plt.ylabel("Consommation totale en GWh",fontsize=12)
plt.xlabel("Date",fontsize=12)
plt.legend()
plt.savefig("P9_04_1graphiques.jpg",bbox_inches ="tight")
plt.show()
```



Prévision de la consommation (corrigée de l'effet température) sur un an, en utilisant la méthode SARIMA sur la série temporelle.

Vérification de stationnarités et différenciations

Sur les données de test :
Consommation non stationnaire (1)

Apres différenciation :

Consommation avec 1 décalage(2) : stationnaire.

Consommation avec 12 décalage(saisonier)(3) : stationnaire.

```
adf_check(data_tr['cons_nodju'])
```
(1)

```
Test de Dickey_fuller Augmenté (adf)
ADF Test Statistic:-1.9695593414436445
p-value:0.3000776501879813
#lags used:5
Number of observations used:42
Faible évidence contre l'hypothèse nulle
Echoue à rejetter l'hypothèse nulle
Les données ont une racine unitaire et sont donc non stationnaires
```

```
#data_tr['First_difference'] = data_tr['cons_nodju'] - data_tr['cons_nodju'].shift(1)
adf_check(data_tr['First_difference'].dropna())#perte 1ere valeur
```
(2)

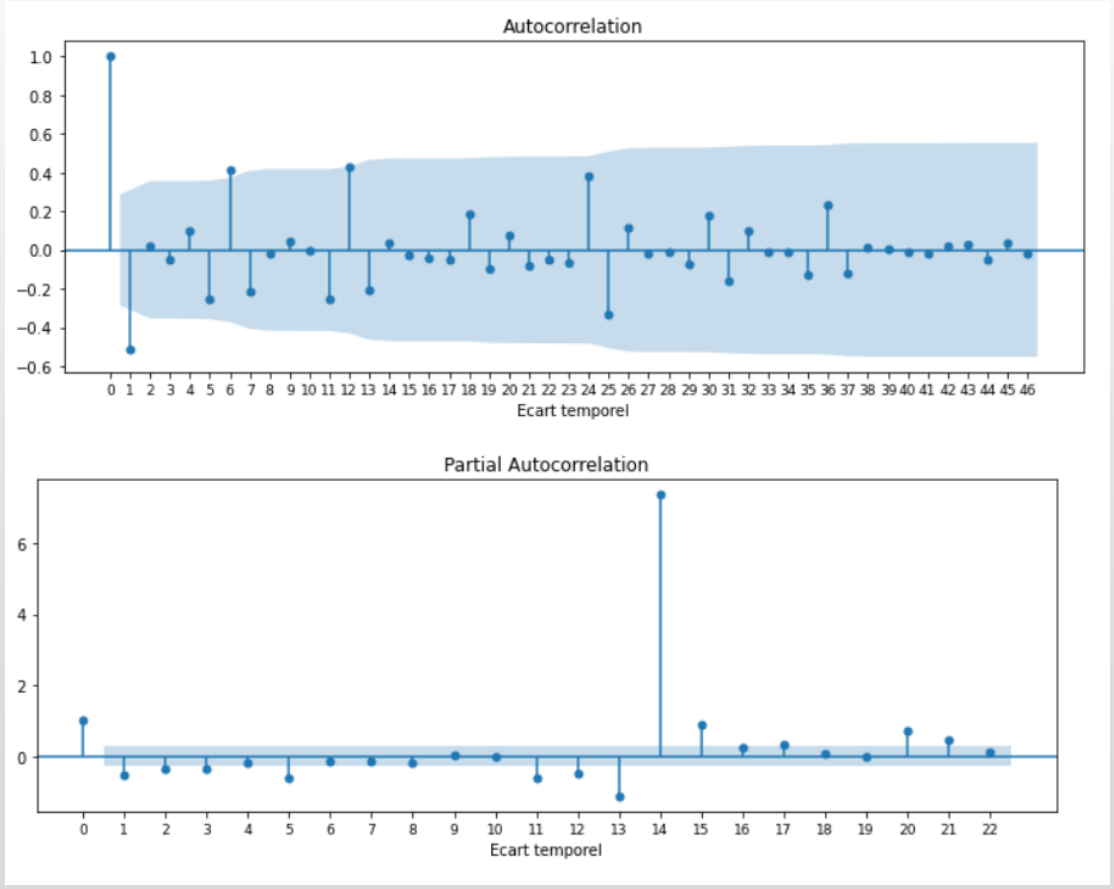
```
Test de Dickey_fuller Augmenté (adf)
ADF Test Statistic:-3.075395467104454
p-value:0.028428432997530888
#lags used:10
Number of observations used:36
Forte preuve contre l'hypothèse nulle
Rejette l'hypothèse nulle
Les données n'ont pas de racine unitaire et sont donc stationnaires
```

```
#data_tr['Seasonal_difference'] = data_tr['cons_nodju'] - data_tr['cons_nodju'].shift(12)
adf_check(data_tr['Seasonal_difference'].dropna())
```
(3)

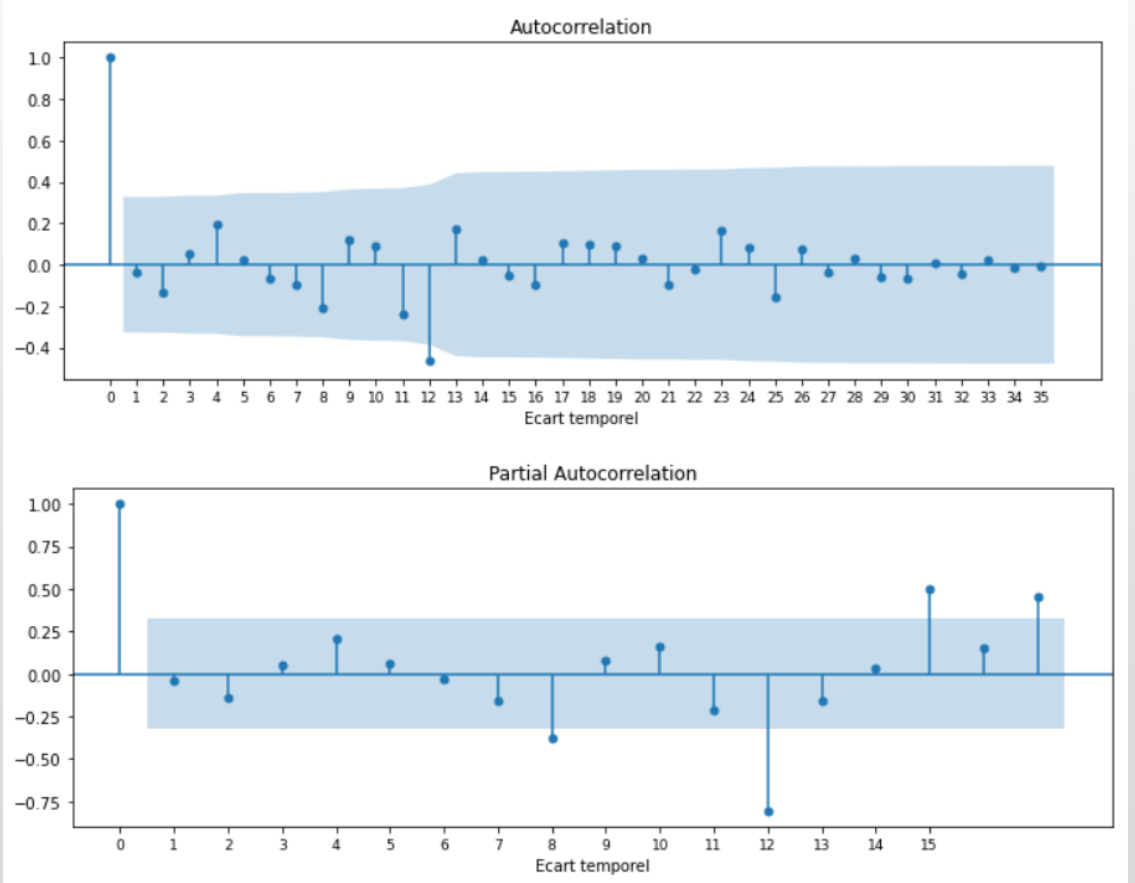
```
Test de Dickey_fuller Augmenté (adf)
ADF Test Statistic:-6.039356380390106
p-value:1.3576731873513718e-07
#lags used:0
Number of observations used:35
Forte preuve contre l'hypothèse nulle
Rejette l'hypothèse nulle
Les données n'ont pas de racine unitaire et sont donc stationnaires
```

Prévision de la consommation (corrigée de l'effet température) sur un an, en utilisant la méthode SARIMA sur la série temporelle.

Acf et Pacf pour paramétrage manuel SARIMA



1 décalage : paramétrage manuel (p,d,q) : (1,1,1)



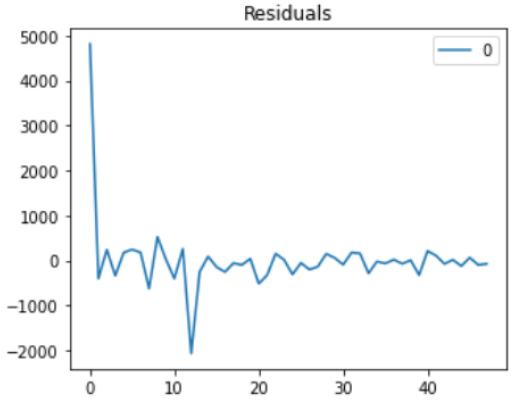
Décalage saisonnier : paramétrage manuel (P,D,Q) : (1,1,1)

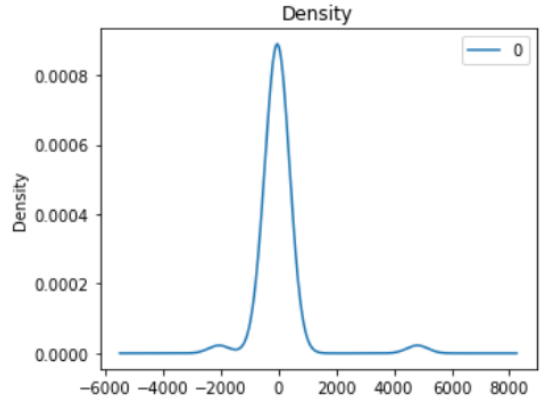
Prévision de la consommation (corrigée de l'effet température) sur un an, en utilisant la méthode SARIMA sur la série temporelle.

Création d'un model SARIMA manuel:

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
model1 = SARIMAX(np.asarray(data_tr['cons_nodju']), order=(1,1,1),seasonal_order = (1,1,1,12))
result1 = model1.fit()
print(result1.summary())
#sarimax_manuel
```

Test des résidus du modèle:






```
                0
count    48.000000
mean      5.324517
std     796.703006
min   -2063.068865
25%    -214.645235
50%     -56.708852
75%     117.694273
max    4808.852608
coefficient d'asymetrie: 4.454061293382172
coefficient d'aplatissement: 29.86983833331907
p-value :  6.147813161327953e-12 ,la p-value est inférieure à un niveau alpha choisi (0.05), alors
l'hypothèse nulle est rejetée (i.e. il est improbable d'obtenir de telles données en supposant qu'ell
es soient normalement distribuées).
```

Prévision de la consommation (corrigée de l'effet température) sur un an, en utilisant la méthode SARIMA sur la série temporelle.

Test de blancheur :

```
from statsmodels.stats.diagnostic import acorr_ljungbox
print('Retard : p-value')
for elt in [6, 12, 18, 24, 30, 36]:
    print('{} : {}'.format(elt, acorr_ljungbox(result1.resid, lags=elt)[1].mean()))
```

```
Retard : p-value
6 : 0.8602530370258208
12 : 0.8833036761383047
18 : 0.853473764402951
24 : 0.8790753004384936
30 : 0.9021222217857379
36 : 0.91835274989351
```

Test Ljung-Box : les p valeurs ne sont pas inferieur au niveau test de 5%, nous ne rejetons pas l hypothèse nulle qui est que le résidu suit un bruit blanc.

Prévision de la consommation (corrigée de l'effet température) sur un an, en utilisant la méthode SARIMA sur la série temporelle.

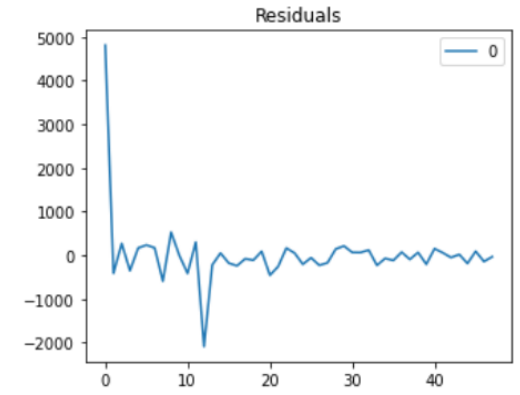
Création d'un model SARIMA automatisé:

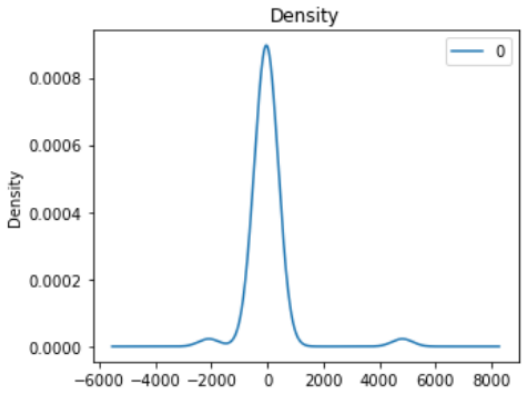
```
train,test = np.array(data_tr['cons_nodju']),np.array(data_te['cons_nodju'])

for p in p_values:
    for d in d_values:
        for q in q_values:
            der = (p,d,q)
            for P in P_values:
                for D in D_values:
                    for Q in Q_values:

                        try:

                            model = SARIMAX(train, order = der , seasonal_order=(P,D,Q,12))

                            model_fit = model.fit()

                            pred_y = model_fit.forecast(12)

                            error = np.sqrt(mean_squared_error(test,pred_y))

                            print('SARIMA%s RMSE = %.2f'% (der,error),P,D,Q)
                            if error < error_min:

                                error_min = error
                                ord = der
                                P_min = P
                                D_min = D
                                Q_min = Q


                        except:

                            print('erreur p d q :', der,P,D,Q, error)
                            continue

print('valeurs trouvées :',ord,P_min,D_min,Q_min,'RMSE:',error_min)
```

```
valeurs trouvées: (6, 1, 4) -- 0 1 1 12
```

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
model2 = SARIMAX(np.asarray(data_tr['cons_nodju']), order=ord,seasonal_order = (P_min,D_min,Q_min,12))
result2 = model2.fit()
print(result2.summary())
#sarimax auto
```

Test des résidus du modèle :






```
                0
count    48.000000
mean     14.385965
std     795.716913
min   -2087.800514
25%    -205.273010
50%     -43.420842
75%      98.482538
max    4808.852608
coefficient d'asymetrie: 4.427195756230769
coefficient d'aplatissement: 29.86317764491501
p-value :  5.392743109705611e-12 ,la p-value est inférieure à un niveau alpha choisi (0.05), alors
l'hypothèse nulle est rejetée (i.e. il est improbable d'obtenir de telles données en supposant qu'ell
es soient normalement distribuées).
```

Prévision de la consommation (corrigée de l'effet température) sur un an, en utilisant la méthode SARIMA sur la série temporelle.

Test de blancheur :

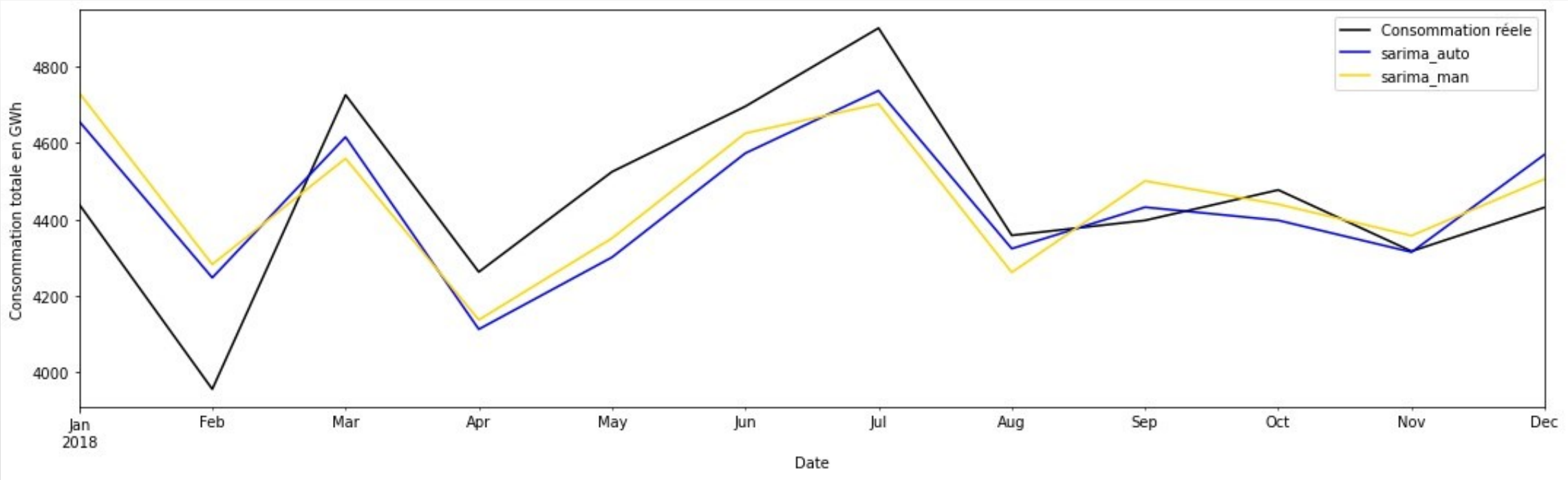
```
print('Retard : p-value')
for elt in [6, 12, 18, 24, 30, 36]:
    print('{} : {}'.format(elt, acorr_ljungbox(result2.resid, lags=elt)[1].mean()))
#valeurs pour les ordres
```

```
Retard : p-value
6 : 0.8345852398298366
12 : 0.8660079405258937
18 : 0.8286196759466289
24 : 0.8575619310117456
30 : 0.8845944406210319
36 : 0.9037139190140571
```

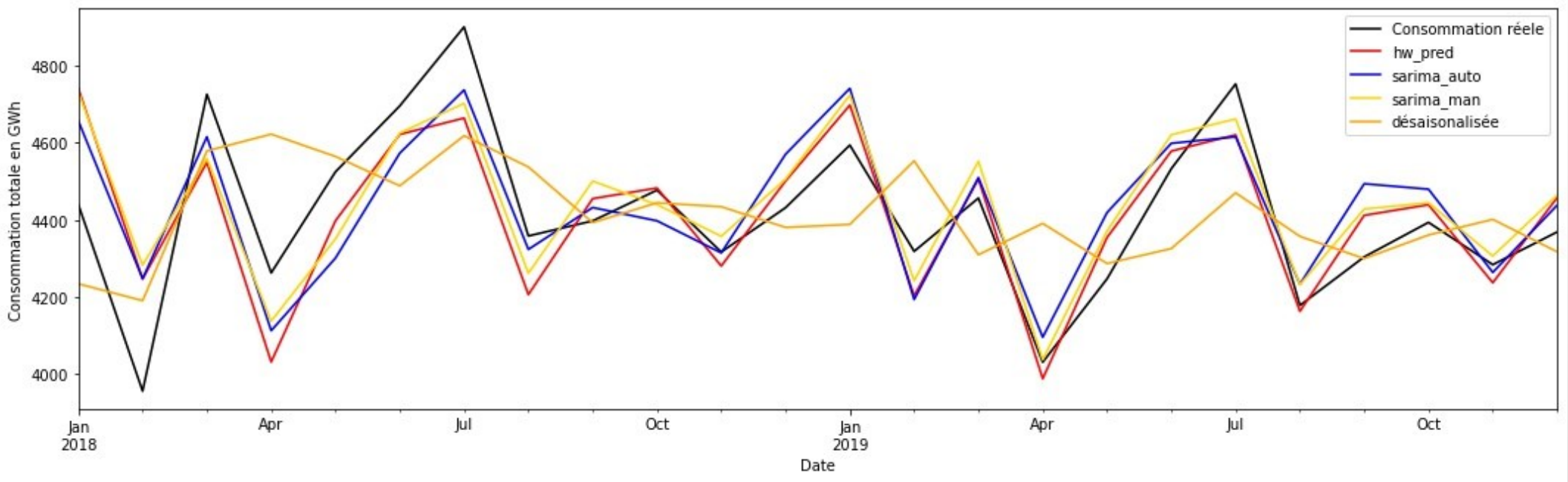
Test Ljung-Box : les p valeurs ne sont pas inferieur au niveau test de 5%, nous ne rejetons pas l hypothèse nulle qui est que le résidu suit un bruit blanc.

Prévision de la consommation (corrigée de l'effet température) sur un an, en utilisant la méthode SARIMA sur la série temporelle.

Prévisions des modèles SARIMA et consommation réelle



Comparatif de tous les modèles



Prévisions 2018:

```
cons_nodjusea rmse : 187.6638159029349
hw_pred rmse : 175.6146919573403
sarima_man rmse : 167.93075819287859
sarima_auto rmse : 154.5727576116403
```

Prévisions 2019:

```
cons_nodjusea rmse : 187.66381590293494
hw_pred rmse : 83.36540450639818
sarima_man rmse : 88.3982588404183
sarima_auto rmse : 111.32578011987924
```

Conclusion

La méthode SARIMA a fourni les prédictions les plus proches de la réalité en comparaison des autres modèles, .