# Projet 9 : Prédire la consommation électrique

# Sommaire

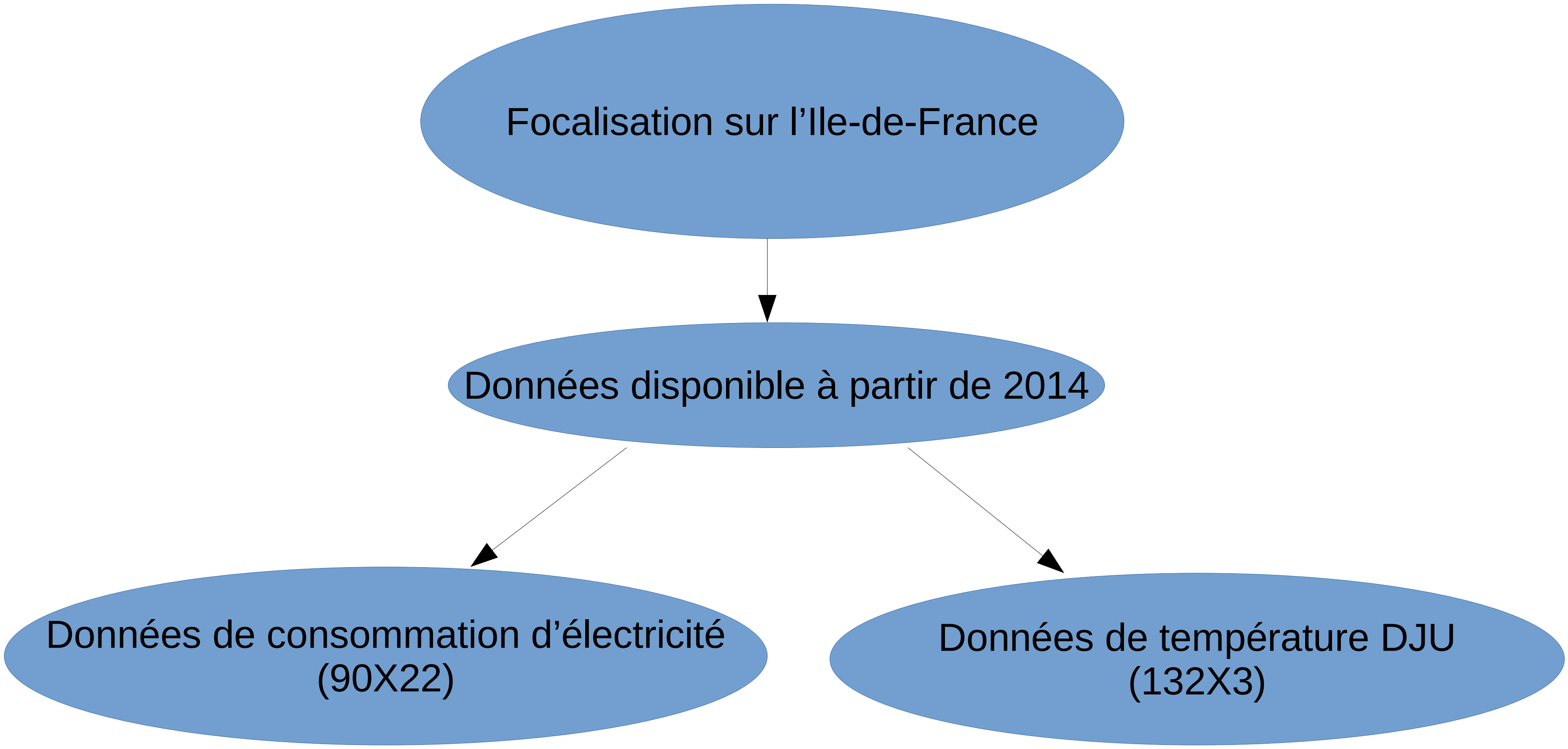
# Introduction

# Introduction

La prévision est une nécessité pour l’équilibre entre l’offre et la demande électrique

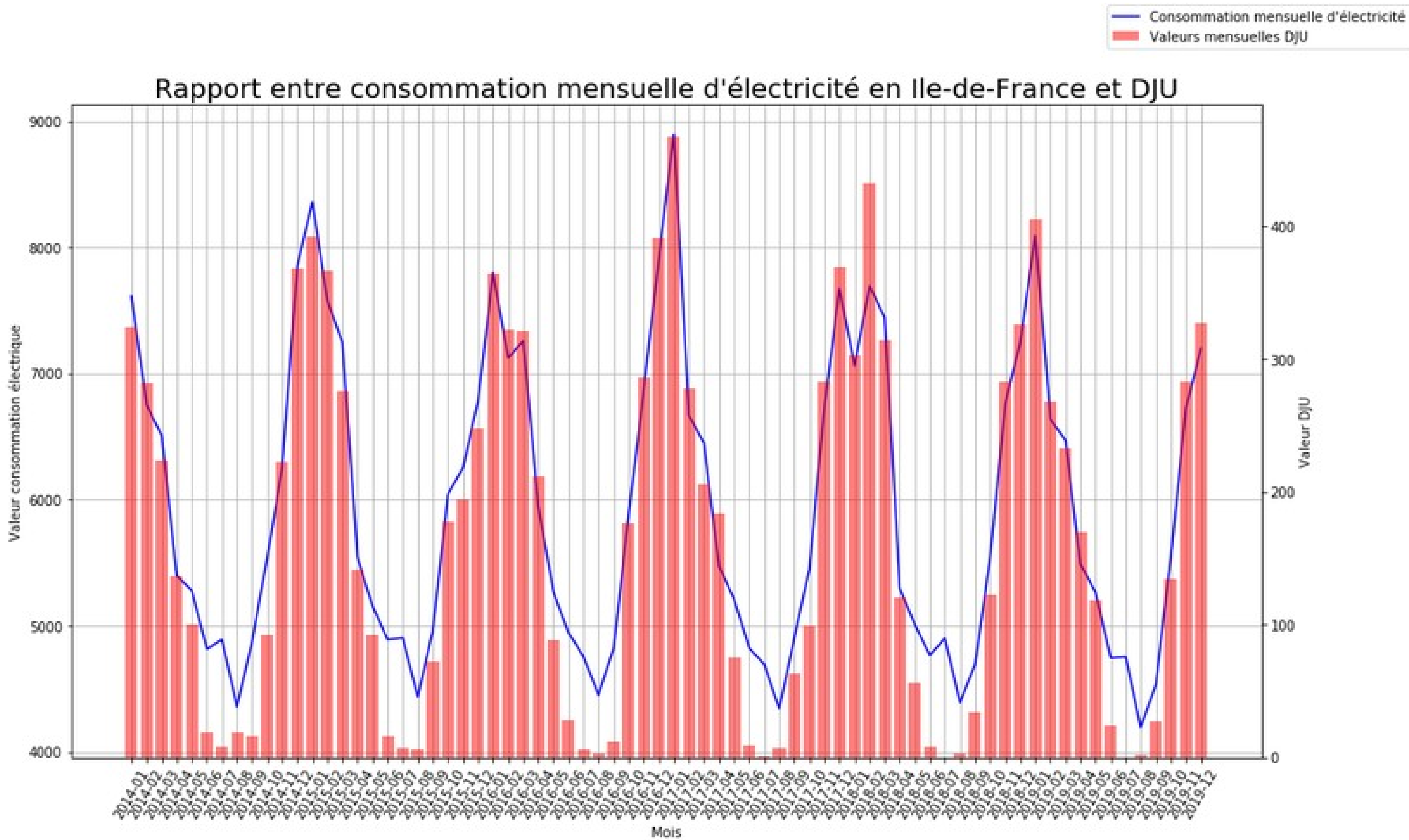
- Sous-évaluation
  - Moyens de production peu flexibles
  - Risque de coupures tournantes
  - Perte d’argent
- Sur-évaluation
  - Perte d’électricité car peu stockable
  - Perte d’argent

# Les données


Focalisation sur l’Ile-de-France
Données disponible à partir de 2014
Données de consommation d’électricité (90X22)
Données de température DJU (132X3)

# L’effet température

# Analyse des variables




En superposant les variables, on constate qu'elles ont des variations corrélées
Par ailleurs, on constate que plus les températures sont basses et plus la consommation est forte

# Analyse de l’effet température

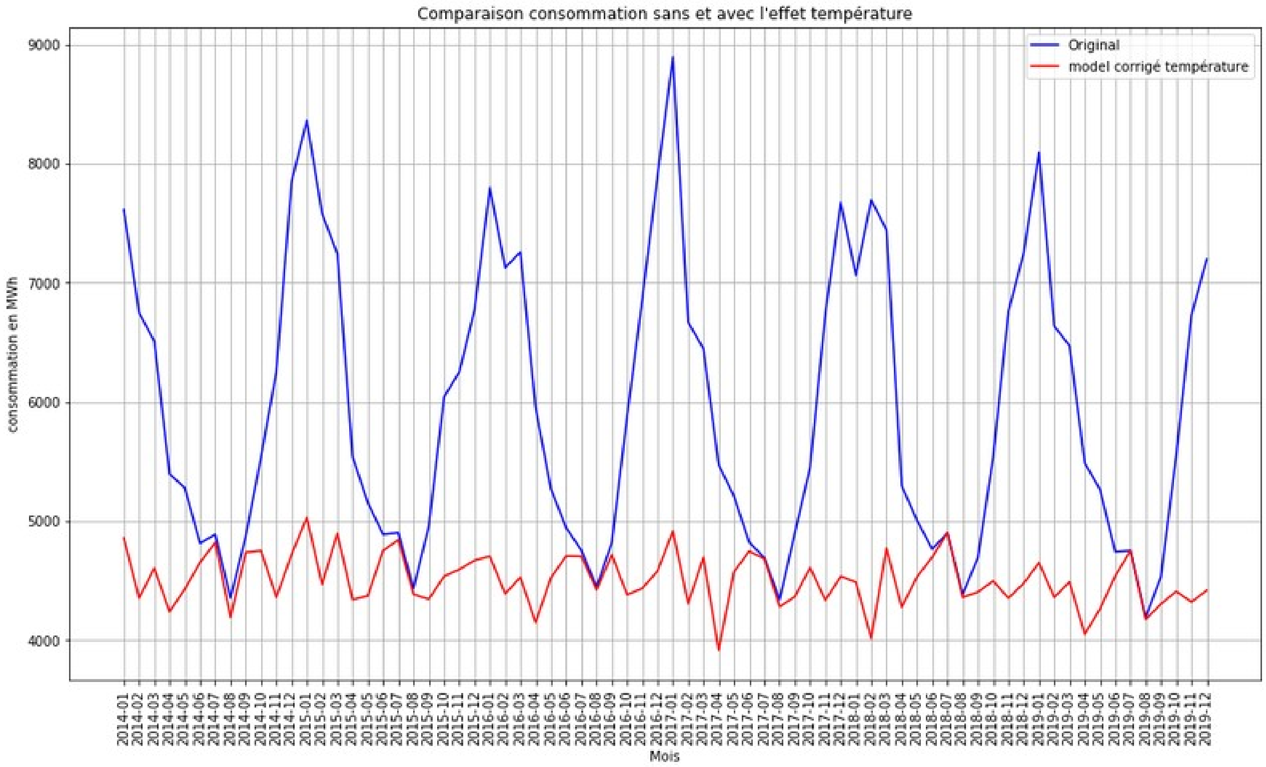
```
                        OLS Regression Results
==============================================================================
Dep. Variable:     Consommation totale   R-squared:                       0.964
Model:                             OLS   Adj. R-squared:                  0.964
Method:                  Least Squares   F-statistic:                     1888.
Date:                 Fri, 04 Dec 2020   Prob (F-statistic):           2.23e-52
Time:                         16:12:12   Log-Likelihood:                -492.48
No. Observations:                   72   AIC:                             989.0
Df Residuals:                       70   BIC:                             993.5
Df Model:                            1
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       4510.3590     42.417    106.334      0.000    4425.761    4594.957
DJU            8.4947      0.195     43.454      0.000       8.105       8.885
==============================================================================
Omnibus:                        0.180   Durbin-Watson:                   2.185
Prob(Omnibus):                  0.914   Jarque-Bera (JB):                0.340
Skew:                          -0.094   Prob(JB):                        0.844
Kurtosis:                       2.720   Cond. No.                         341.
==============================================================================
```

Via la régression linéaire on confirme la forte corrélation des variables :
DJU explique 96 % de la variance

# Suppression de l’effet température

Pour supprimer l’effet température on soustrait à la consommation la variable DJU multiplié par son coefficient de régression

Représentation graphique du modèle original vs le modèle corrigé de la température :

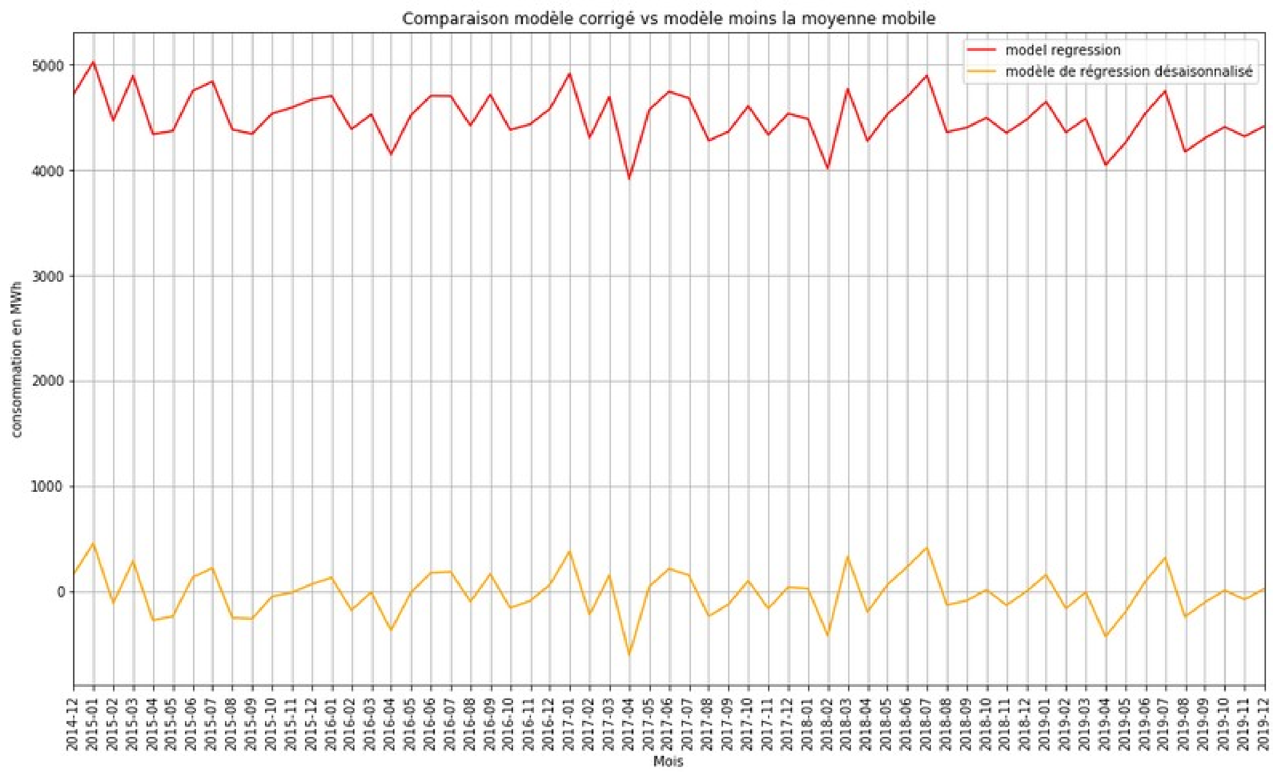



Sans l’effet de température, la consommation d’électricité est bien plus stable dans le temps

# Stationnarité

# Correction par les moyennes mobiles

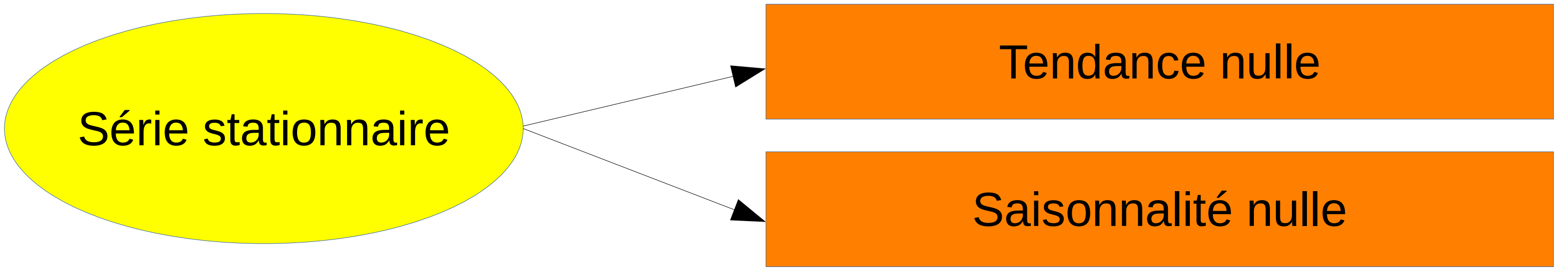
Nous tentons de corriger la saisonnalité via soustraction de la moyenne mobile sur 12 mois

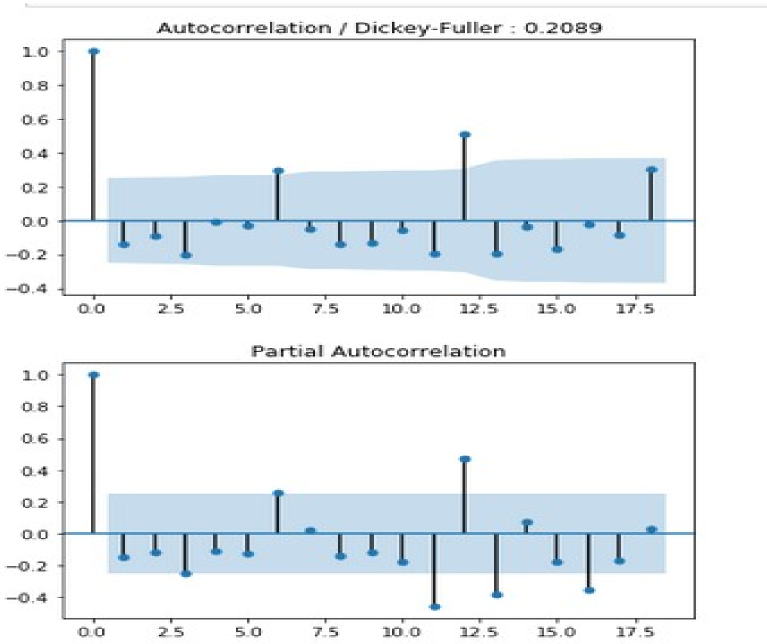



En réalité on stabilise la moyenne à 0, ceci afin de tenter de supprimer l’effet de tendance

# Notion de stationnarité

Soustraction de la moyenne mobile : méthode pour tenter de stationnariser une série temporelle



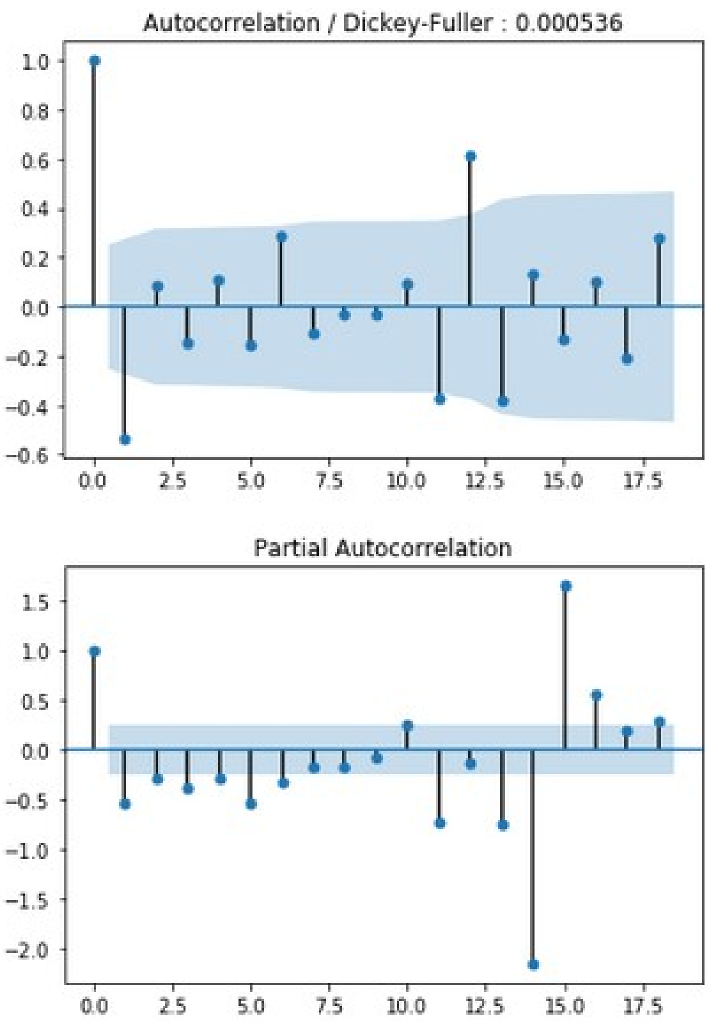




Test Dickey-Fuller : pvalue > 0.05

Saisonnalité présente en multiple de 6

Par conséquent, l’hypothèse de stationnarité n’est pas vérifiée

# Solution stationnaire




**On applique une différenciation à la série pour tenter de la stationnariser**

Test Dickey-Fuller : pvalue < 0.05

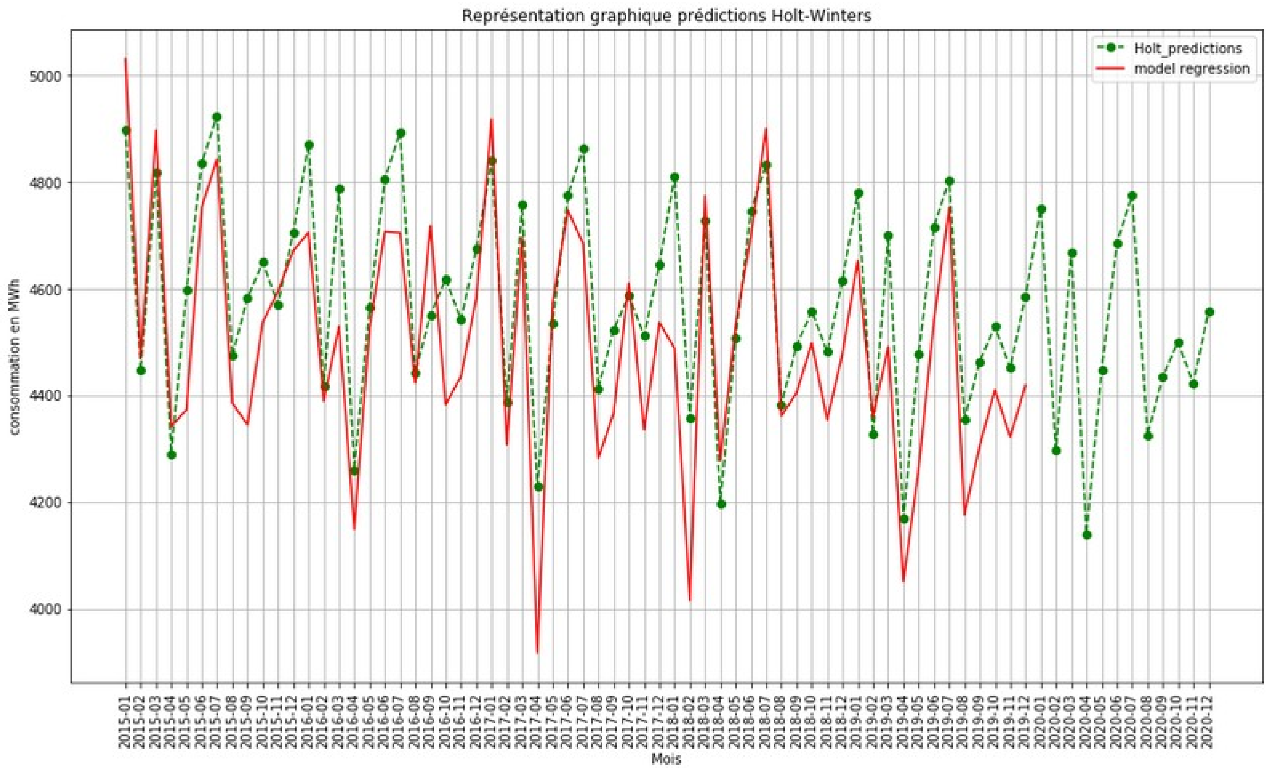
Pas de décroissance vers 0 sur l'auto corrélation, qui aurait indiqué la présence d'une tendance

Par conséquent, l'hypothèse de stationnarité est vérifiée

# Modèle de prédiction

# Holt-Winters

Double lissage exponentiel
Plus les valeurs sont anciennes moins leur poids est important




```
The Mean Squared Error is 22577.75
The Root Mean Squared Error is 150.26
The Mean Absolute Percentage Error is 3.21%
```

MAPE = 3,21 % (bonne efficacité)

# SARIMA

S : Saisonnalité – AR : Autocorrélation – I : Intégré – MA : Moyenne mobile

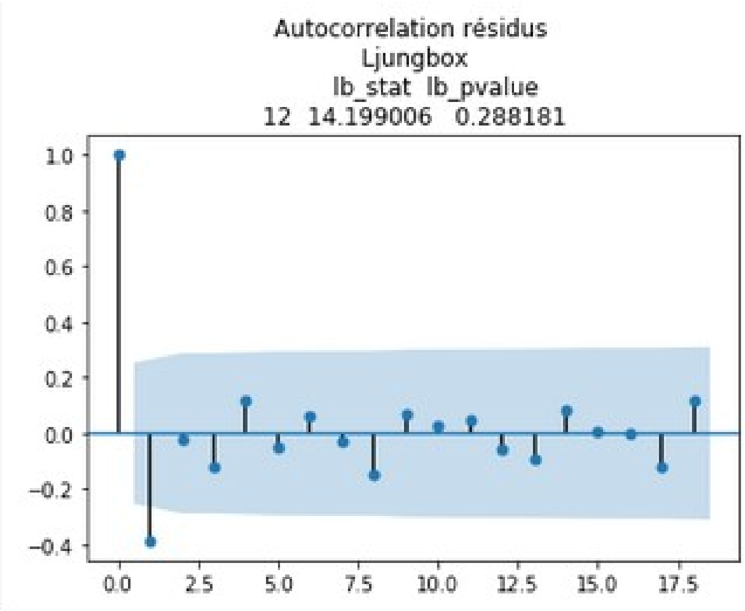
```
3]: p = range(0, 3)
d = range(0, 2)
q = range(0, 2)
pdq = list(itertools.product(p, d, q))
seasonal_pdq = [(x[0], x[1], x[2], 12) for x in list(itertools.product(p, d, q))]
best_aic = 0
for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = sm.tsa.statespace.SARIMAX(y,order=param,seasonal_order=param_seasonal,enforce_stationarity=False,er
            results = mod.fit()
            if ((best_aic == 0) | (best_aic > results.aic)) & (results.pvalues.max() <= 0.05)\
            & (sm.stats.diagnostic.acorr_ljungbox(results.resid,lags=[12], return_df=False)[1]>0.05) :
                best_aic = results.aic
                best_pvalue = results.pvalues.max()
                best_param = param
                best_param_seasonal = param_seasonal
        except:
            continue
print('Meilleurs paramètres : SARIMAX{}x{} - AIC:{}'.format(best_param,best_param_seasonal,best_aic))

mod = sm.tsa.statespace.SARIMAX(y,
                                order=(best_param),
                                seasonal_order=(best_param_seasonal),
                                enforce_stationarity=False,
                                enforce_invertibility=False)
results = mod.fit()
print(results.summary())
```

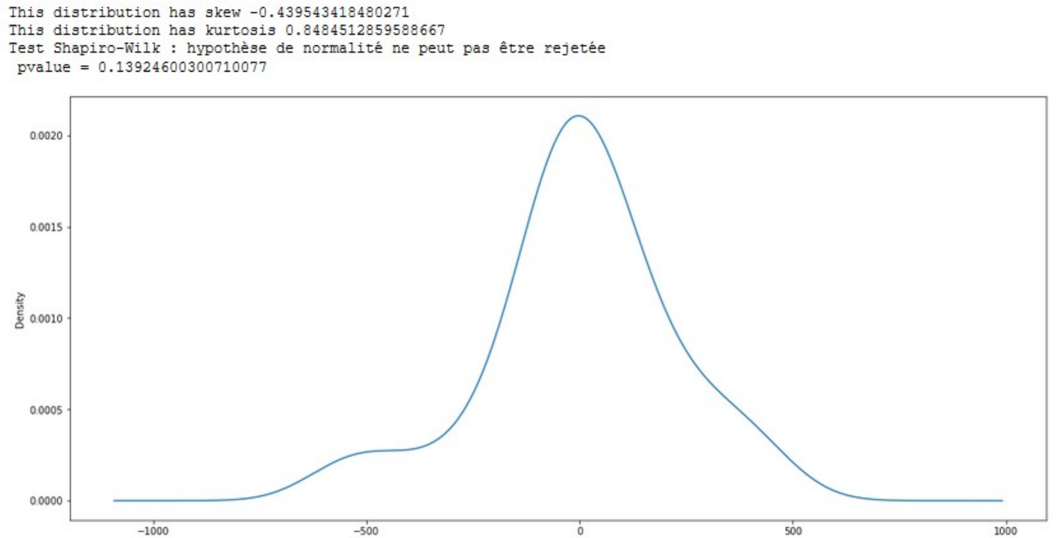
```
Meilleurs paramètres : SARIMAX(0, 0, 1)x(2, 1, 0, 12) - AIC:313.62856999320286
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                       Mois, Consommation   No. Observations:            60
Model:             SARIMAX(0, 0, 1)x(2, 1, [], 12)   Log Likelihood         -152.814
Date:                              Thu, 17 Dec 2020   AIC                     313.629
Time:                                      18:35:53   BIC                     318.341
Sample:                                  01-01-2015   HQIC                    314.879
                                       - 12-01-2019
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.5163      0.231     -2.234      0.025      -0.969      -0.063
ar.S.L12      -0.7797      0.184     -4.230      0.000      -1.141      -0.418
ar.S.L24      -0.4620      0.206     -2.245      0.025      -0.865      -0.059
sigma2       1.97e+04   5269.011      3.739      0.000    9374.450       3e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.35   Jarque-Bera (JB):                 3.17
Prob(Q):                              0.55   Prob(JB):                         0.21
Heteroskedasticity (H):               0.38   Skew:                             0.84
Prob(H) (two-sided):                  0.19   Kurtosis:                         3.60
```

Méthode choix du modèle SARIMA :

Minimisation AIC
Critère de vraisemblance < 0.05
Blancheur des résidus > 0.05

# SARIMA : Analyse des résidus




Test blancheur des résidus : 0,29

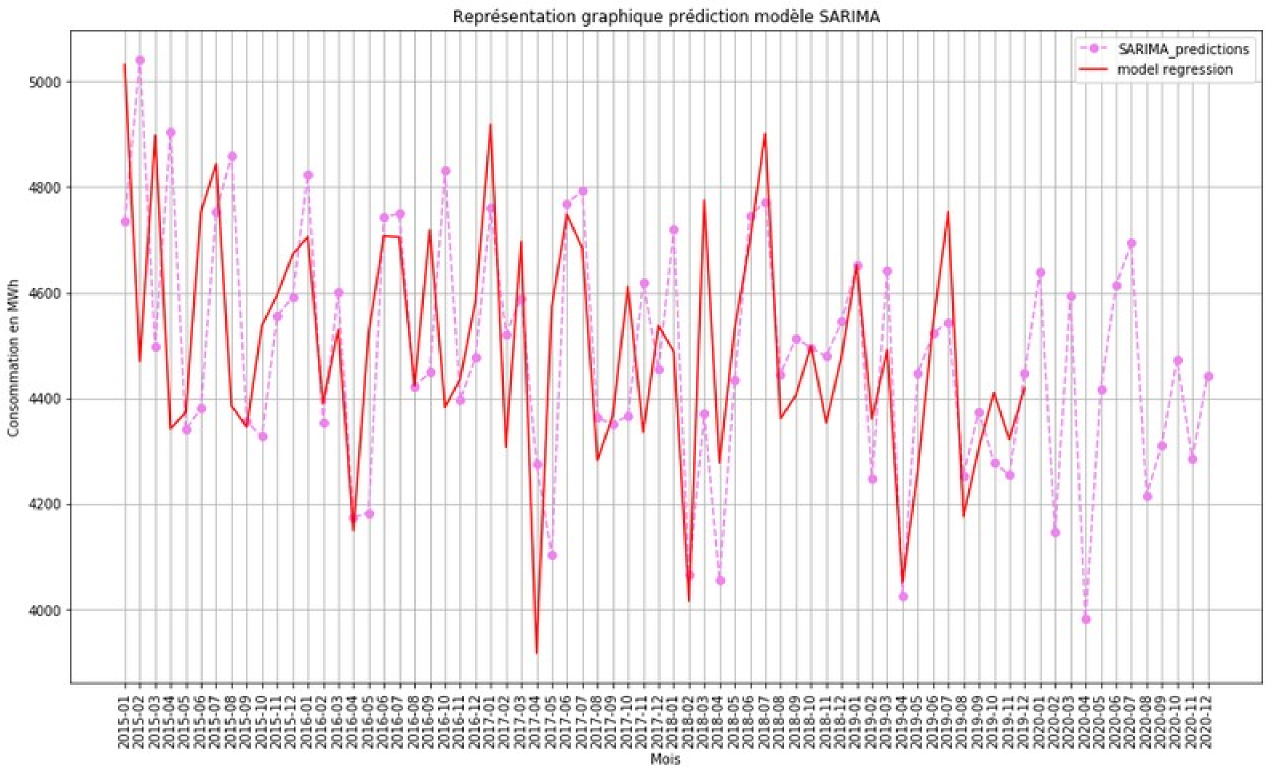
Démontre que les résidus sont indépendants

Test de normalité des résidus :

Shapiro-Wilk = 0,14 (hypothèse de normalité vérifié)

# SARIMA : Analyse des résidus

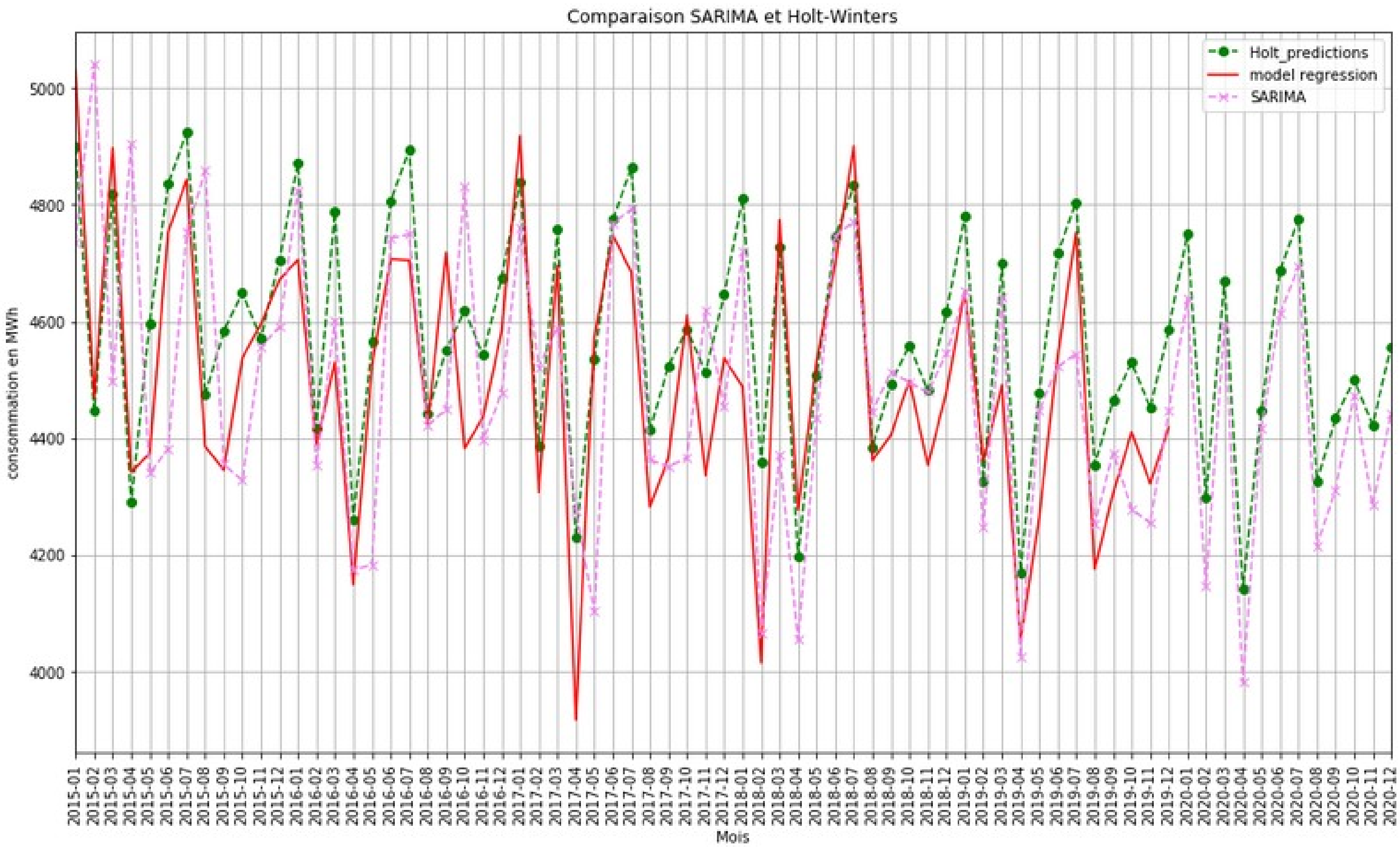
S : Saisonnalité – AR : Autocorrélation – I : Intégré – MA : Moyenne mobile




```
The Mean Squared Error is 12261.81
The Root Mean Squared Error is 110.73
The Mean Absolute Percentage Error is 2.03%
```

MAPE = 2,03% (bonne efficacité)

# Comparaison Hotl-Winters et SARIMA




Holt-Winters prédit une tendance à la hausse
SARIMA à la baisse plus corrélée à la tendance de la série

SARIMA améliore de 1,18 % le MAPE

# Conclusion

Dans le domaine de l'électricité la prédiction est d'une importance capitale, afin d'adapter au mieux les moyens de production et d'éviter des coupures.

Comme on a pu le constater la température est fortement corrélée à la consommation électrique, seulement pour prédire la consommation il faut la soustraire aux données en déterminant sont coefficient de régression.

Mais l'étape la plus importante est de stationnariser notre série temporelle afin d'avoir le modèle de prédiction le plus efficace possible et éviter l'influence d'une tendance ou d'une saisonnalité, plusieurs méthodes existes dans notre cas la méthode de différenciation a été le plus efficace.

On a pu utiliser ensuite deux méthodes pour prédire notre futur consommation : Holt-Winters et son double lissage exponentiel et SARIMA.
Au final le modèle SARIMA est le plus performant avec un pourcentage moyen d'erreur absolu à 1,82 %, une amélioration de 1,18 % par rapport à la méthode Holt-Winters.