

**Objectif :** modélisation de séries temporelles



# Prédisez la demande en électricité !

# Glossaire

## Données consolidées

Données basées sur les comptages de Rte et d'Enedis (données disponibles en M+1)

## Données définitives

Données basées sur l'ensemble des comptages de Rte et des ELD\* (données disponibles en A+1)

\* Entité Locale de Distribution (Enedis, régies municipales, ...)

## DJU

Valeur représentative de l'écart entre la température d'une journée donnée et un seuil de température préétabli (18 °C )

## Energie

Seule l'énergie électrique sera considérée (en TWh)

## Correction des données

De l'effet température uniquement dues au chauffage électrique

## CVS

Correction des Variations Saisonnières

# Enercoop **spécialiste** des énergies **renouvelables** !

La plupart des énergies renouvelables sont intermittentes d'où des difficultés de prévision des capacités de production d'électricité car la demande varie en fonction :

- *de la temporalité*
- *de paramètres tels que la météo (température...)*

Le challenge : mettre en adéquation l'offre et la demande

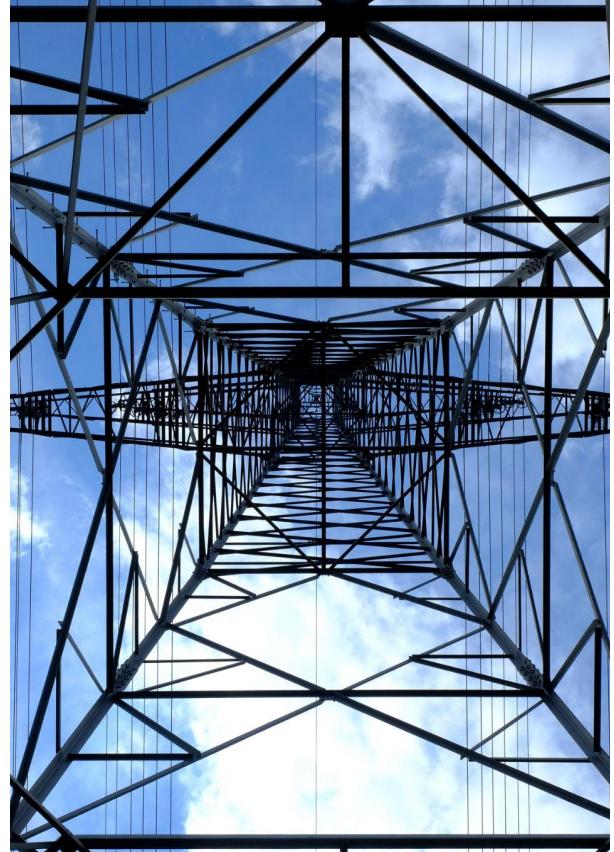
# Sommaire !

6 principales parties :

- *Présentation des données*
- *Détail du programme implémenté (python)*
- *Correction par régression linéaire*
- *Désaisonnalisation par moyennes mobiles*
- *Prévision par la méthode Holt Winters*
- *Prévision par SARIMA*

# 1. Présentation des données

Données de consommation **mensuelle** en  
électricité en **énergie**  
Données **météo**



# Consommation totale mensuelle en énergie

Source : **Rte**

<https://www.rte-france.com/eco2mix/telecharger-les-indicateur>



**Consommation totale mensuelle en énergie (TWh) du 01.01.2012 au 01.10.2021 1390 observations et 21 variables** (colonnes)

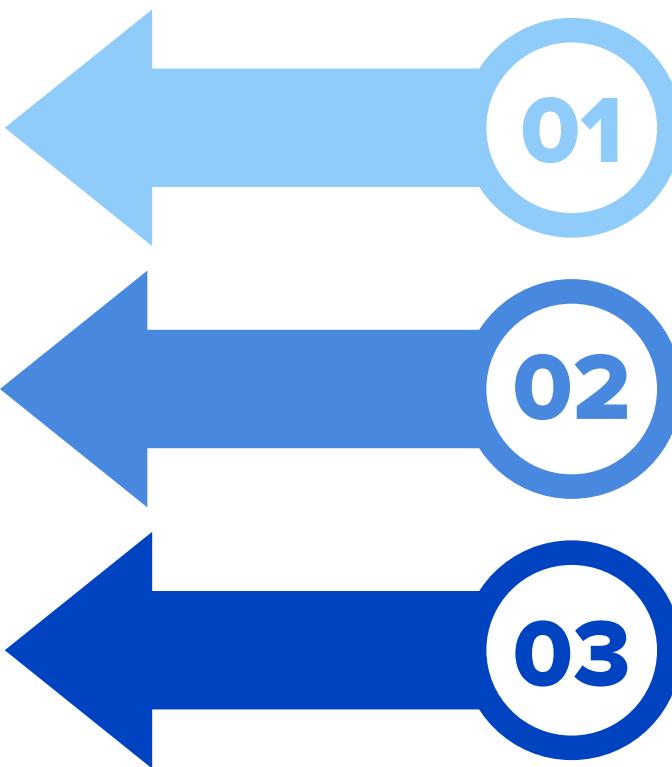
Mois	Qualité	Territoire	Production totale	Production nucléaire	Production thermique totale	Production thermique charbon	Production thermique fioul	Production thermique gaz	Production hydraulique	Production éolien	...	Production bioénergies	Consommation totale
12-01-01	Données définitives	France	57177	42811.0	5399	741.0	691	3967	6741	1587	...	514	51086
12-02-01	Données définitives	France	54419	38661.0	8721	2511.0	1309	4901	5244	1123	...	454	54476
12-03-01	Données définitives	France	48583	37549.0	5276	1435.0	666	3175	3894	977	...	531	43156
12-04-01	Données définitives	France	44192	33100.0	3484	1655.0	486	1343	5302	1572	...	405	40176
12-05-01	Données définitives	France	40433	29058.0	1772	854.0	368	549	7706	974	...	469	35257
12-06-01	Données définitives	France	37518	26562.0	2219	960.0	384	875	6823	994	...	437	33219
12-07-01	Données définitives	France	37910	28818.0	2221	1110.0	437	674	4901	912	...	513	34141
12-08-01	Données définitives	France	36339	28314.0	2533	1165.0	413	956	3563	907	...	486	32247
12-09-01	Données définitives	France	38035	30130.0	2450	1382.0	396	671	3641	931	...	470	33269
12-10-01	Données définitives	France	42987	32848.0	3518	2039.0	456	1022	4534	1355	...	450	38628
12-11-01	Données définitives	France	49228	37119.0	5189	1821.0	524	2845	4914	1309	...	514	44260
12-12-01	Données définitives	France	54715	39912.0	5319	1714.0	548	3057	6521	2290	...	525	49602

# Dataset d'étude de consommation totale

Mois	Qualité	Territoire	Consommation totale
2012-01-01	Données définitives	France	51086
2012-02-01	Données définitives	France	54476
2012-03-01	Données définitives	France	43156
2012-04-01	Données définitives	France	40176
2012-05-01	Données définitives	France	35257
2012-06-01	Données définitives	France	33219
2012-07-01	Données définitives	France	34141
2012-08-01	Données définitives	France	32247
2012-09-01	Données définitives	France	33269
2012-10-01	Données définitives	France	38628
2012-11-01	Données définitives	France	44260
2012-12-01	Données définitives	France	49602

## Dataset propre

Pas de valeurs manquantes



01

## 12 régions et France

La Corse, les régions d'Outre-Mer ne sont pas représentées

02

## Consommations nulles

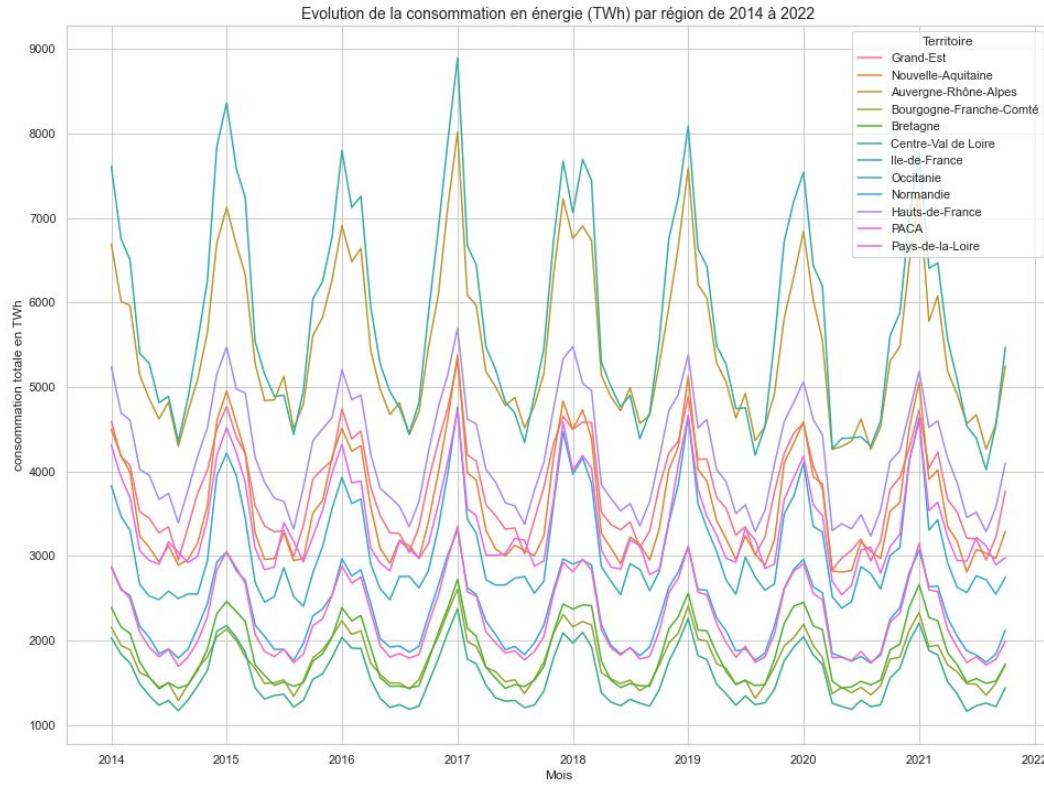
Suppression de ces valeurs pour les régions concernées en 2013

03

# Consommation régionale en énergie

Seules les régions du continent sont représentées.

Tendance **linéaire** et **saisonnalité** pour chaque région.

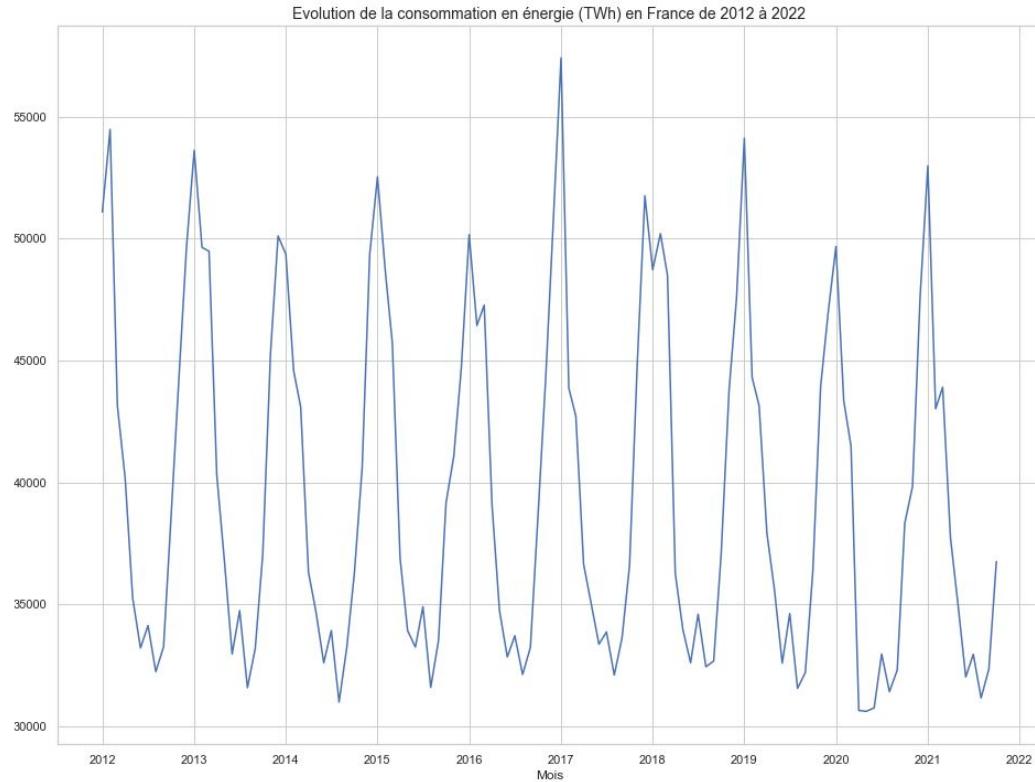


# Consommation en France en énergie

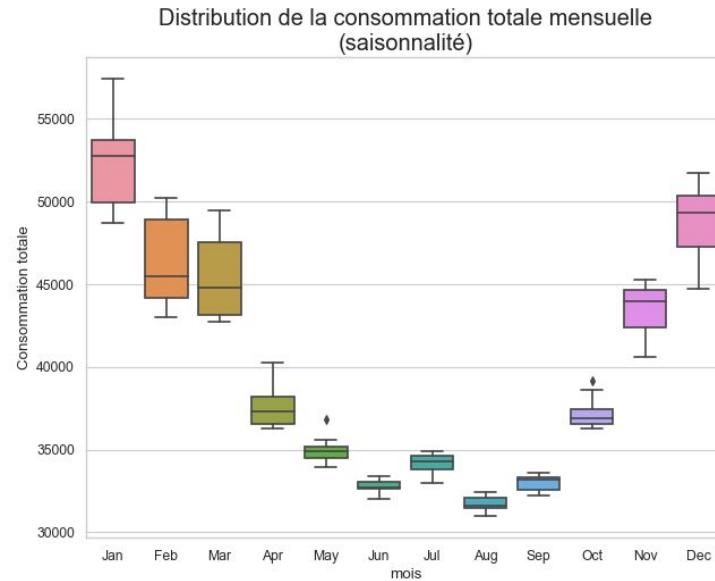
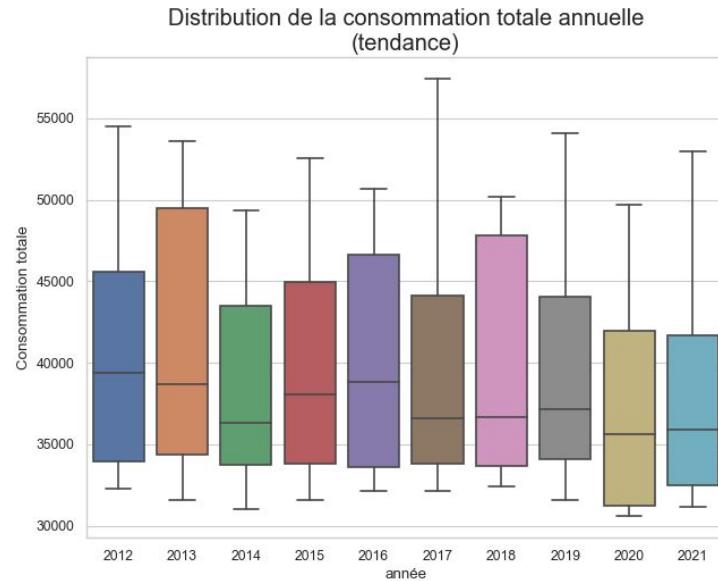
Tendance **linéaire** et **saisonnalité**.

2017 : forte consommation due à une vague de froid.

2020 : faible consommation due à la crise sanitaire.



# Une consommation qui dépend des saisons



# Météo (Degré Jour Unifié)

Source : **cegibat**

<https://cegibat.grdf.fr/simulateur/calcul-dju>



# DJU pour l'effet température

**CEGIBAT**  
Centre d'études et de gestion du gaz

Réglementation gaz ▾ Efficacité énergétique ▾ Actualités ▾ Simulateurs

Q Recherche

**CALCUL DES DJU**

Étape 1 Étape 2 Résultats

**1. Indiquez la station météo**

Sélectionnez la station météo

- 01 - AMBERIEU
- 02 - ST QUENTIN
- 03 - VICHY-CHARMEIL
- 04 - ST AUBAN
- 05 - EMBRUN
- 06 - CANNES

Version 2.0.0

## DJU

Écart entre la T° de référence (18°C) et la T° médiane du jour

01

## Stations météo

1 station sélectionnée pour chaque territoire

02

## DJU finaux

Moyenne des 8 stations météo sélectionnées

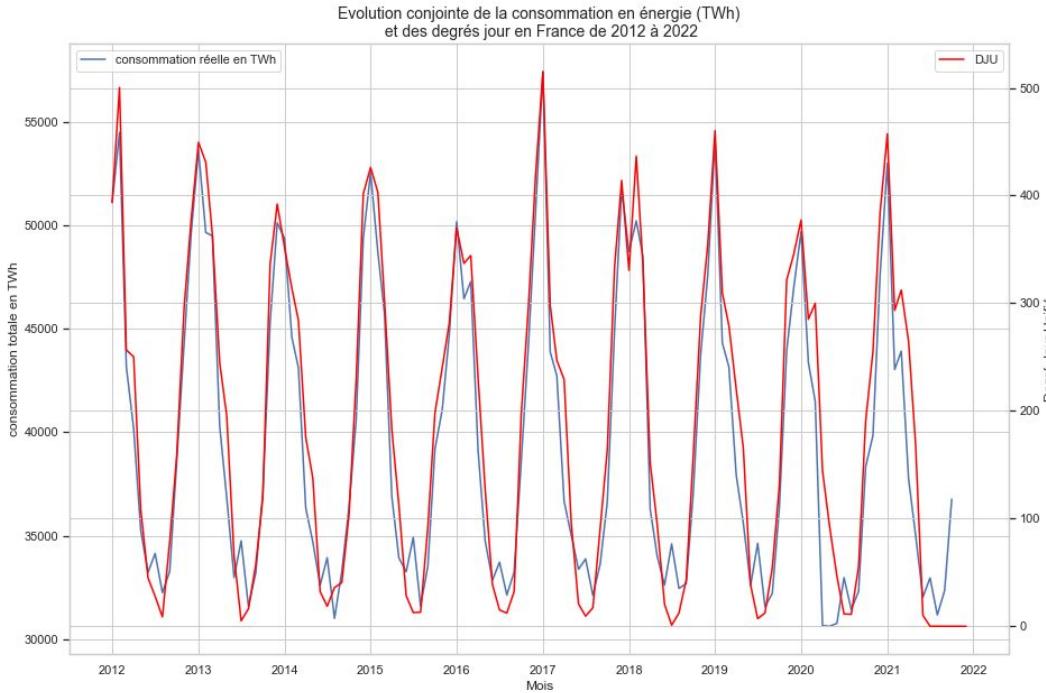
03

**DJU** : 8 stations sélectionnées : Ambérieu, Bourges, Carcassonne, Charleville-Méz,  
Mende, Rennes, St Quentin, Toulouse

Mois	dju_zone_h2b	dju_zone_h1c	dju_zone_h1b	dju_zone_h2d	dju_zone_h2a	dju_zone_h1a	dju_zone_h2c	dju_zone_h3	dju_mean
2012-01-01	384.1	444.6	430.9	479.0	332.0	398.1	348.6	332.3	393.7000
2012-02-01	491.2	574.1	531.8	576.1	409.3	497.8	472.7	449.4	500.3000
2012-03-01	237.1	253.4	313.3	309.6	252.0	274.3	207.3	205.9	256.6125
2012-04-01	244.8	211.2	294.1	344.9	259.0	291.4	184.1	172.5	250.2500
2012-05-01	94.1	90.3	128.4	169.9	127.2	137.7	57.4	64.8	108.7250
2012-06-01	31.8	19.5	86.4	69.4	58.3	82.6	6.3	5.1	44.9250
2012-07-01	13.0	15.2	53.6	52.5	39.3	43.3	3.4	3.0	27.9125
2012-08-01	2.5	3.8	21.4	16.0	11.9	13.9	0.1	0.0	8.7000
2012-09-01	64.6	63.8	147.9	114.0	91.4	109.1	16.6	21.6	78.6250
2012-10-01	142.0	146.6	239.1	223.8	151.0	202.8	85.1	83.9	159.2875
2012-11-01	298.3	293.4	360.0	344.3	292.4	338.0	229.0	210.1	295.6875
2012-12-01	363.0	413.0	422.5	489.9	312.8	394.4	311.3	315.2	377.7625

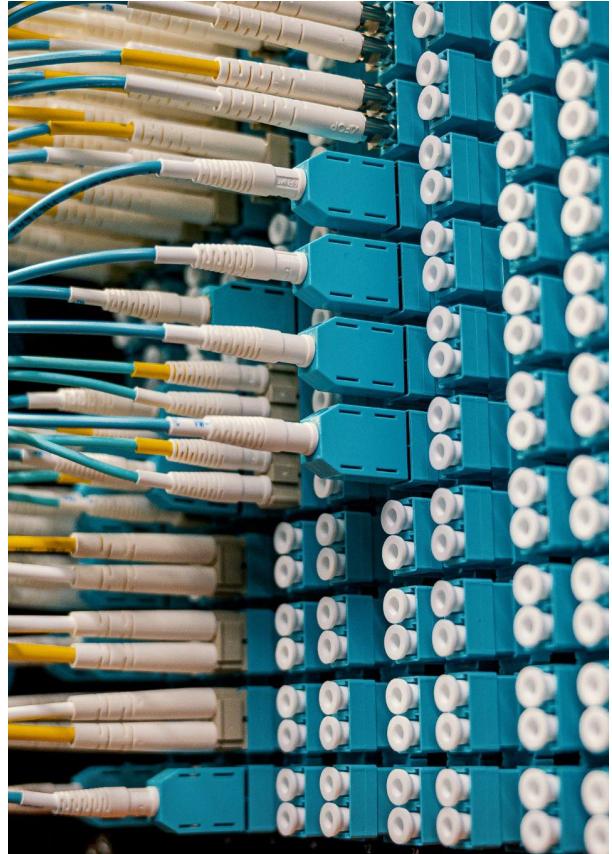
# Evolution DJU et consommation en énergie

La consommation énergétique en France varie de manière similaire aux DJU (écart de température, 18°C) avec une [saisonnalité](#). Correction des données de consommation mensuelles de l'effet température pour une meilleure comparaison.



## **2. Le programme en détail**

Les dessous du code python



# Série temporelle... les points forts du code

## datetimeIndex

```
pd.read_csv('energie_mensuelle.csv',
            parse_dates=True,
            index_col='Mois',
            encoding=ENCODAGE)
```

## Indexing-Slicing

```
df_cons_hw.loc['2019-01-01':'2019-12-01']
```

## Log pour une saisonnalité additive

```
train_cons_ln = np.log(train_cons)
test_cons_ln = np.log(test_cons)
```

## Autocorrelation

```
sgt.plot_pacf(df_cons_sar_ln_dif_6_12);
sgt.plot_acf(df_cons_sar_ln_dif_6_12);
```

## Décomposition saisonnière

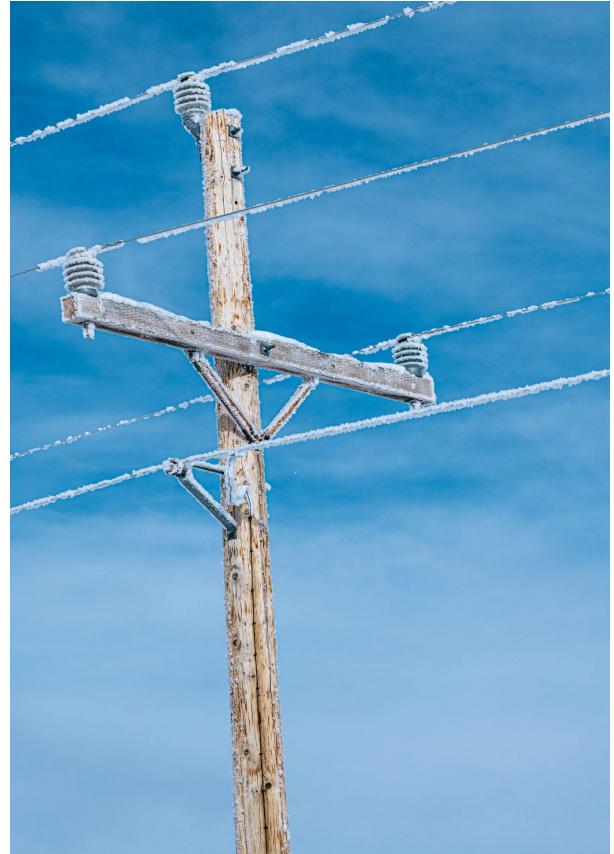
```
seasonal_decompose(df_cons_mob['Consommation_totale_corrigee'],
                     model='additive',
                     extrapolate_trend='freq')
```

## SARIMA

```
SARIMAX(np.asarray(df_cons_sarima_ln_tronc['Conso']),
        order=(1, 0, 1),
        seasonal_order=(1, 0, 1, 12))
```

### **3. Correction par régression linéaire**

Consommations mensuelles  
corrigées de l'effet température  
(dues au chauffage électrique)



# Série temporelle... quelques notions

## Saisonnalité

Comportement, motif périodique dans la série temporelle (période ici égale à 12)

## Désaisonnalisation

Identifier et estimer les effets saisonniers selon le temps et élimination de la série temporelle

## Modèle

### Additif

$$X(t) = T(t) + S(t) + \epsilon(t)$$

### Multiplicatif

$$X(t) = T(t)(1+S(t))(1+\epsilon(t))$$

## Régression linéaire

Relation linéaire forte entre 2 variables du type

$$Y = aX + b + \epsilon$$

## Méthode des moindres carrés

Minimiser la somme au carré des écarts

## Correction des effets saisonniers

Recherche des coefficients a et b

# Régression linéaire par Statsmodels

```
OLS Regression Results
=====
Dep. Variable: Consommation_totale R-squared:      0.933
Model:           OLS Adj. R-squared:     0.932
Method:          Least Squares F-statistic:    1467.
Date: Fri, 14 Jan 2022 Prob (F-statistic): 6.57e-64
Time: 15:37:35 Log-Likelihood: -964.61
No. Observations: 108 AIC:             1933.
Df Residuals:    106 BIC:             1939.
Df Model:        1
Covariance Type: nonrobust
=====
            coef    std err        t      P>|t|      [0.025      0.975]
Intercept  3.066e+04   296.353   103.450      0.000   3.01e+04   3.12e+04
dju         46.2979    1.209    38.301      0.000    43.901    48.694
=====
Omnibus:                 6.168 Durbin-Watson:       1.411
Prob(Omnibus):            0.046 Jarque-Bera (JB):    6.203
Skew:                   -0.396 Prob(JB):        0.0450
Kurtosis:                  3.866 Cond. No.       409.
=====
```

**R<sup>2</sup>**

La variable consommation est expliquée à 93 %

01

**Modèle significatif**

Prob (F-statistic) : 6.57e-64

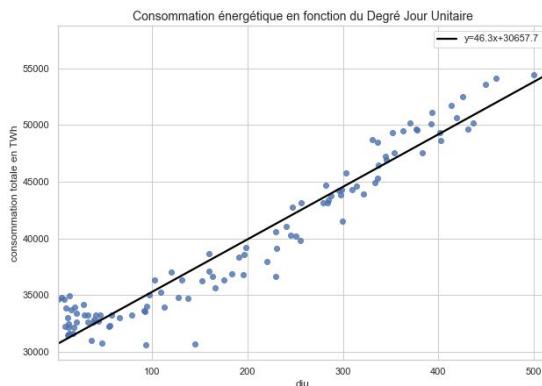
02

**Variable DJU**

Significative au seuil de 5%

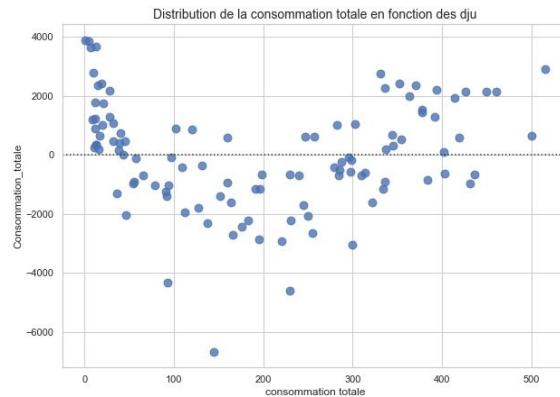
03

# La régression linéaire en image



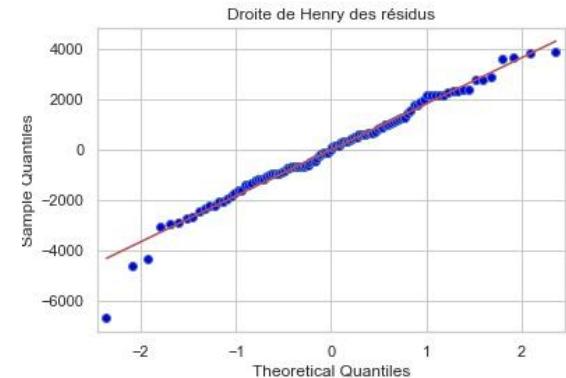
## Droite de régression

$$y = ax + b + \epsilon$$
$$a = 46.29$$
$$b = 30657.74$$



## Résidus

Test de Breusch-Pagan: HO d'homoscédaticité des résidus ne peut être rejetée (bruit blanc avec variance constante)

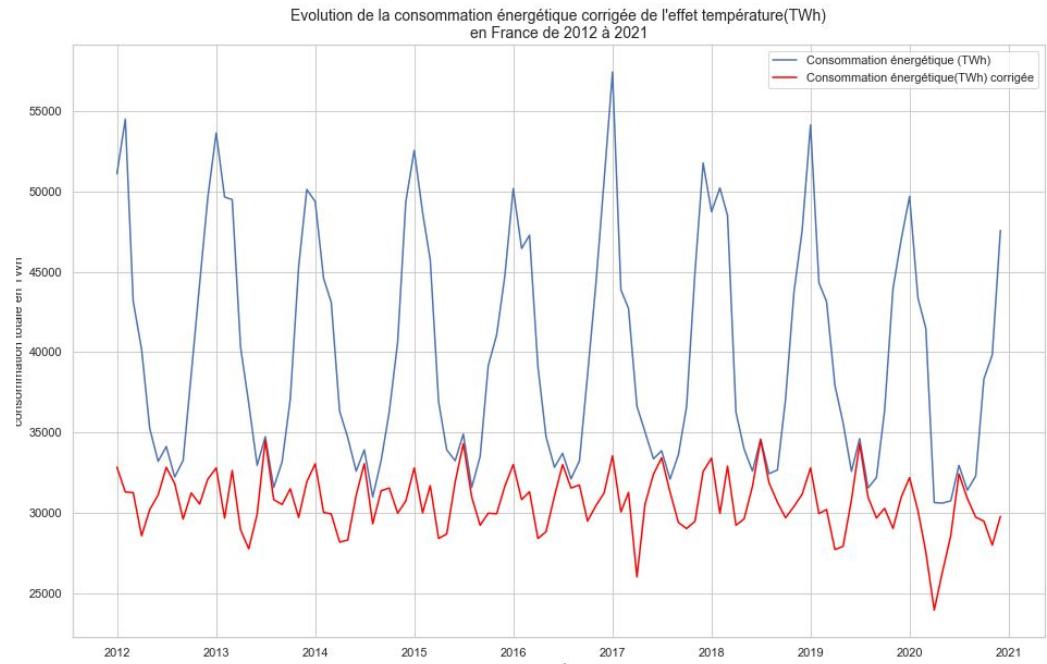


## Courbe de Henry

Les résidus suivent une loi normale. Test de Kolmogorov-Smirnov  $p\_value : 0.83$  HO de normalité ne peut être rejetée

# Consommation corrigée de l'effet température

Dans l'ensemble, les données ont été lissées mais il reste une saisonnalité...



```
df_cons_dju['Consommation_totale'] - (df_cons_dju['dju'] * slope)
```

## **4. Moyennes mobiles : désaisonnalisation**

Désaisonnalisation de la  
consommation corrigée de l'effet  
température



# Moyennes mobiles, des conditions ?

## MOYENNES MOBILES

C'est une combinaison linéaire d'instants passés et futurs de la série temporelle

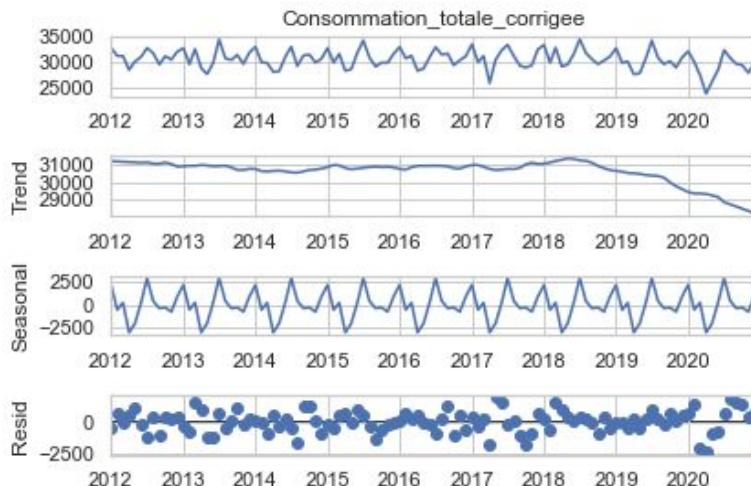
## OBJECTIF

Mettre en évidence l'allure de la tendance en supprimant la composante saisonnière et en atténuant le bruit

## MODELE ADDITIF

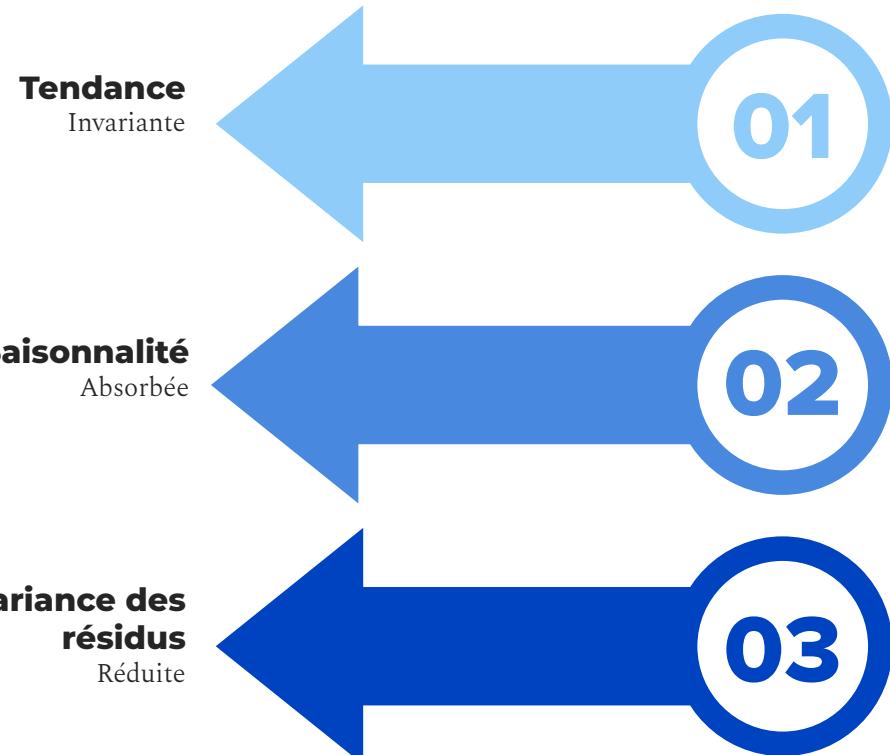
L'amplitude de la composante saisonnière  $S(t)$  et du bruit  $\varepsilon(t)$  reste constante au cours du temps.

# La moyenne mobile idéale



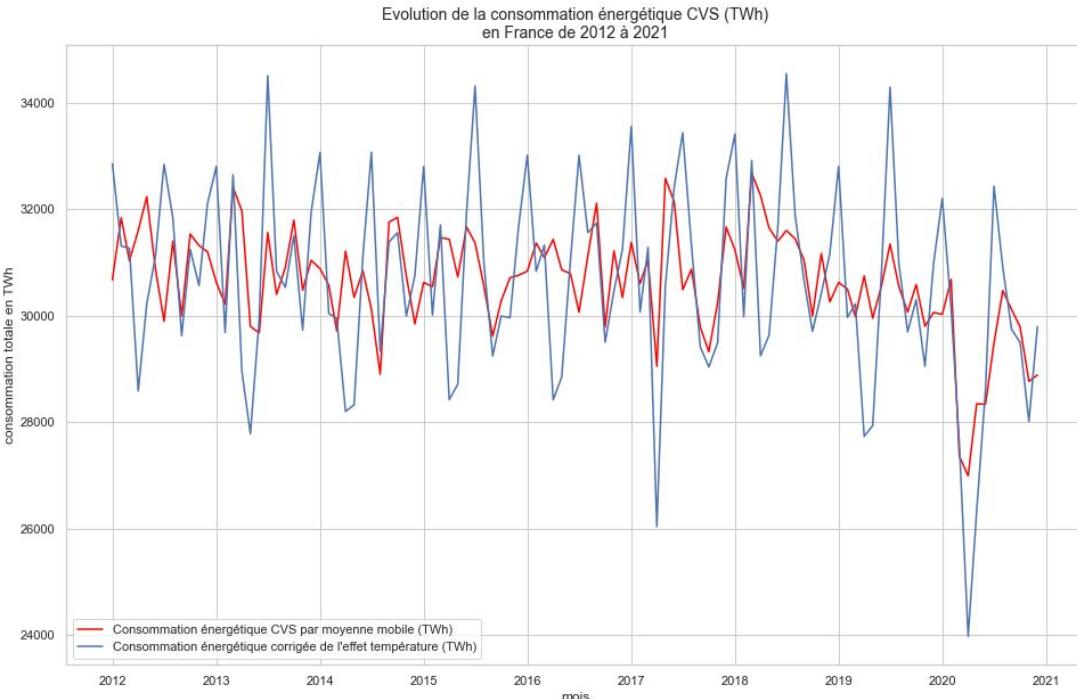
Décomposition saisonnière

```
df_cons_mob['Saisonnalité'] = decomp_df_cons_mo.seasonal
```



# Désaisonnaliser grâce aux moyennes mobiles

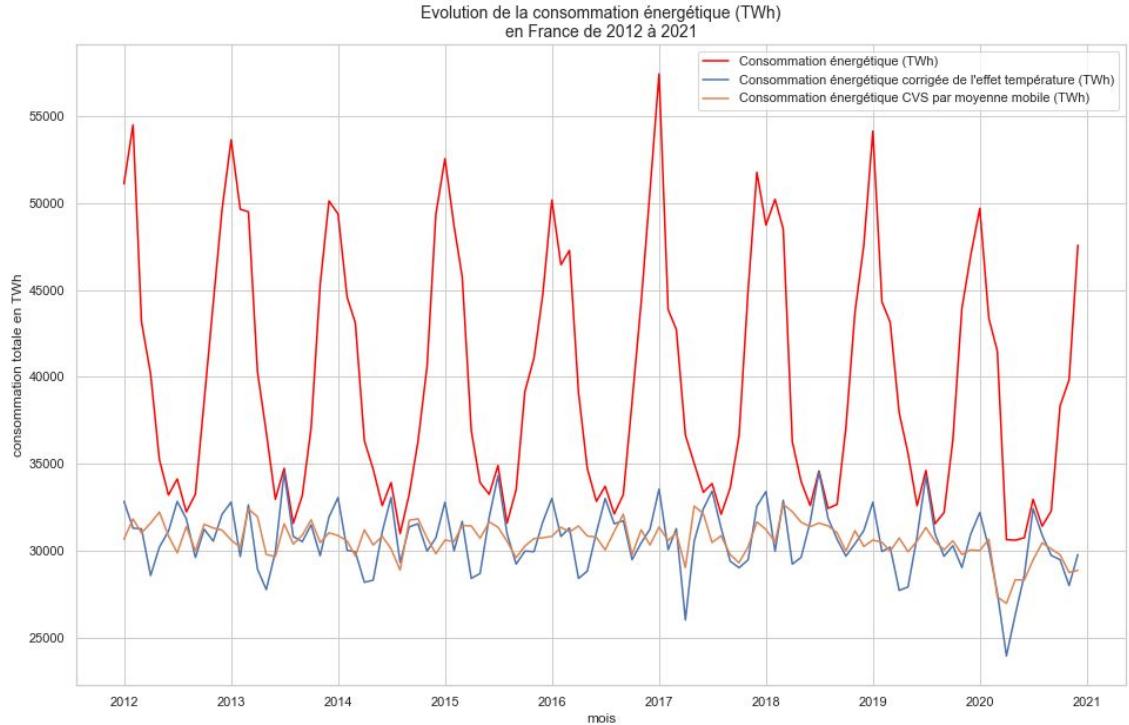
La désaisonnalisation par les moyennes mobiles donne le meilleur résultat.



```
df_cons_mob['Consommation_totale_corrigeé'] - df_cons_mob['Saisonalité']
```

# CVS

Les moyennes mobiles ont permis une désaisonnralisation pour un modèle prédictif plus performant



## **5. Prévision par la méthode Holt Winters**

Rajoute une saisonnalité.  
L'observation récente a plus de  
poids



# Lissage exponentiel, Pourquoi?

## OBJECTIF

Prévoir à court terme

## COMMENT ?

Moyennes pondérées d'observations antérieures, les poids décroissant de manière exponentielle au fil du temps

## LISSAGE SIMPLE

Renvoie une constante.  
Prévision de données sans tendance ni saisonnalité à court terme.

## LISSAGE DOUBLE

Renvoie une droite. Prévision de données avec tendance à plus long terme que LES

## HOLT-WINTERS

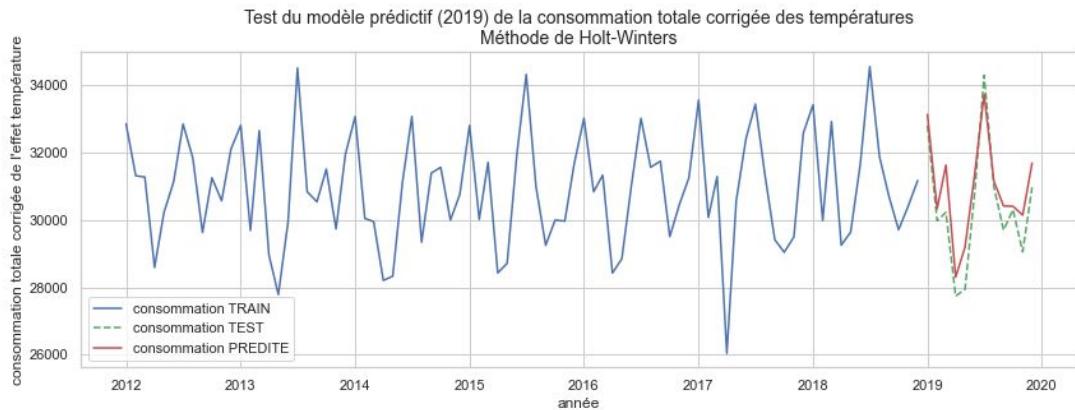
Renvoie un “motif”. Prévision de données avec éventuelle tendance et saisonnalité

## MODELE

Additif pour “glisser” dans le temps

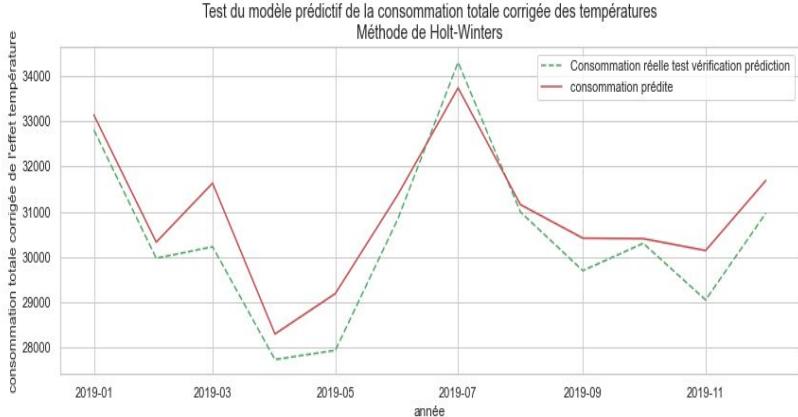
# Holt-Winters prévision 2019

Les prévisions sont satisfaisantes mais perfectibles



```
ExponentialSmoothing(np.asarray(df_cons['Conso']),
                      seasonal_periods=12,
                      trend='add',
                      seasonal='add'
                     ).fit()
```

# Zoom sur les données



## MODELE

Additif. Les données sont transformées en log

01

## RMSE

Root Mean Square Error : **762.13**

02

## MAPE

Mean Average Percentage Error : **0.021**

03

## 6. Prévision par SARIMA

Seasonal Auto Regressive  
Integrated Moving Average



# AR, MA, SARIMA, Pour quoi?

## MODÈLE AR( $p$ )

AR : AutoRégressif  
S'il y a une relation linéaire entre  
 $X(t)$  et les  $p$  derniers instants de  $X(t)$

## CORRÉOGRAMME AR

ACF : décroît rapidement vers 0  
(exponentielle ou sinusoïdale amortie)  
ACFP : s'annule à partir du rang  $p+1$

## MODÈLE MA( $q$ )

MA : Moving Average  
Une valeur est une combinaison  
linéaire de  $q$  erreurs qui se suivent

## CORRÉOGRAMME MA

ACF : s'annule à partir du rang  $q+1$   
ACFP : décroît rapidement vers 0

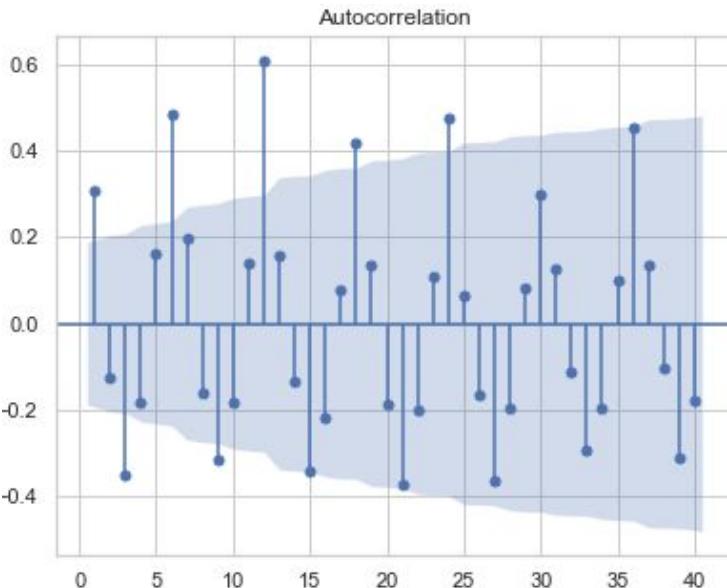
## DIFFÉRENCIATION (I)

$d$  est le nombre de différenciations  
nécessaire à la stationnarisation de la  
série temporelle condition essentielle  
pour la modélisation

## SARIMA

Seasonal AutoRegressive Integrated  
Moving Average  
Inclut un opérateur de différentiation  
(D) et de saisonnalité (S)

# Stationnarisation de la série temporelle



## STATIONNAIRE

Moyenne, variance et covariance constantes et ne dépendent pas du temps

01

## MODELE ADDITIF

Transformation en log pallier l'accroissement de la saisonnalité

02

## TEST DE DICKEY FULLER

p value = 0.99

03

# Différenciation de la série temporelle

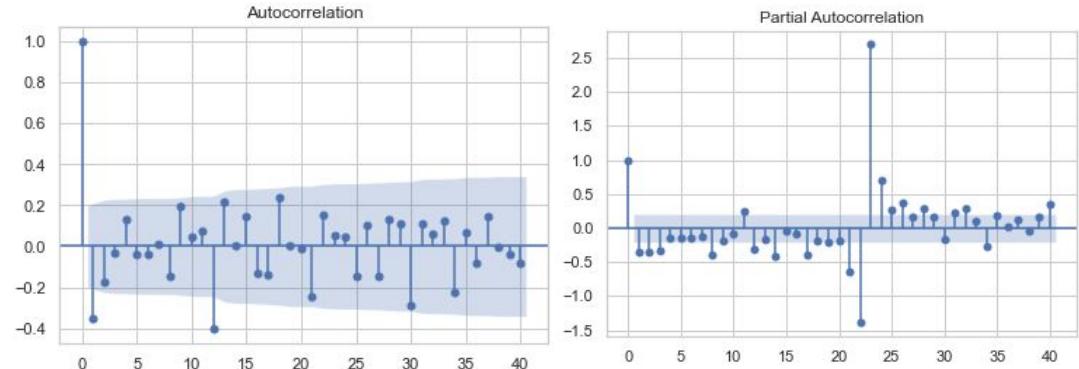
Différenciation 1-12

Test de Dickey Fuller :

p value = 2e-4

La série temporelle est  
stationnaire.

SAMIRA (0, 1, 1) (0, 1, 1, 12)



# Modèle SAMIRA $(1, 0, 1)(1, 0, 1, 12)$

Les coefficients sont tous significatifs au seuil de 5%.

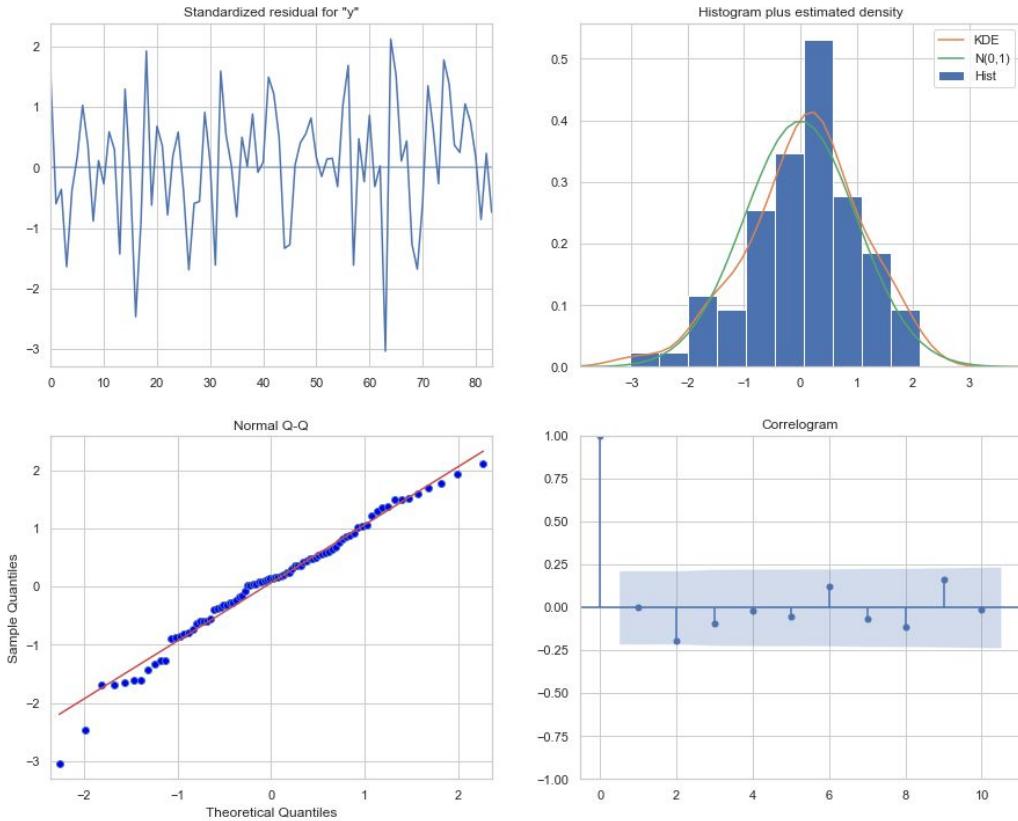
Faible AIC.

Le résidu conforme à un bruit blanc (p values du test de Ljung-Box supérieures à 5%).

SARIMAX Results							
Dep. Variable:	y	No. Observations:	84	Model:	SARIMAX(1, 0, 1)x(1, 0, 1, 12)	Log Likelihood	156.457
Date:	Mon, 17 Jan 2022	AIC	-302.914	Time:	14:55:59	BIC	-290.759
Sample:	0	HQIC	-298.028		- 84		
Covariance Type:	opg						
	coef	std err	z	P> z	[0.025	0.975]	
ar.L1	1.0000	8.82e-05	1.13e+04	0.000	1.000	1.000	
ma.L1	-0.9433	0.057	-16.451	0.000	-1.056	-0.831	
ar.S.L12	0.9936	0.017	58.666	0.000	0.960	1.027	
ma.S.L12	-0.8072	0.236	-3.425	0.001	-1.269	-0.345	
sigma2	0.0009	0.000	4.554	0.000	0.001	0.001	
Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	3.15				
Prob(Q):	0.98	Prob(JB):	0.21				
Heteroskedasticity (H):	1.38	Skew:	-0.45				
Prob(H) (two-sided):	0.40	Kurtosis:	3.31				
Warnings:							
[1] Covariance matrix calculated using the outer product of gradients (complex-step).							
Retard : p-value							
6 : 0.9937646356260735							
12 : 0.9968823173673043							
18 : 0.9979215449115362							
24 : 0.9984411586836522							
30 : 0.9987529269469219							
36 : 0.9989607724557682							
48 : 0.999220579341826							

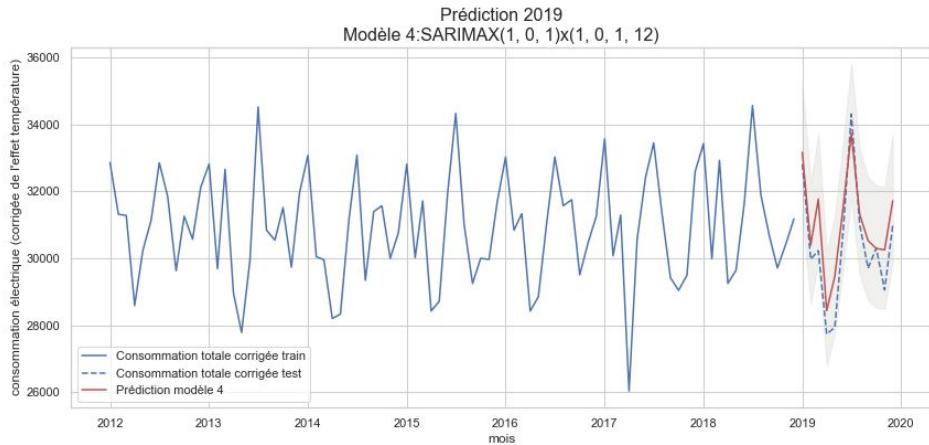
# Bruit blanc du modèle SAMIRA **(1, 0, 1)(1, 0, 1, 12)**

Le bruit blanc semble gaussien et aléatoire.  
Corrélogramme conforme à un bruit blanc.  
Moyenne proche de zéro.  
Variance constante

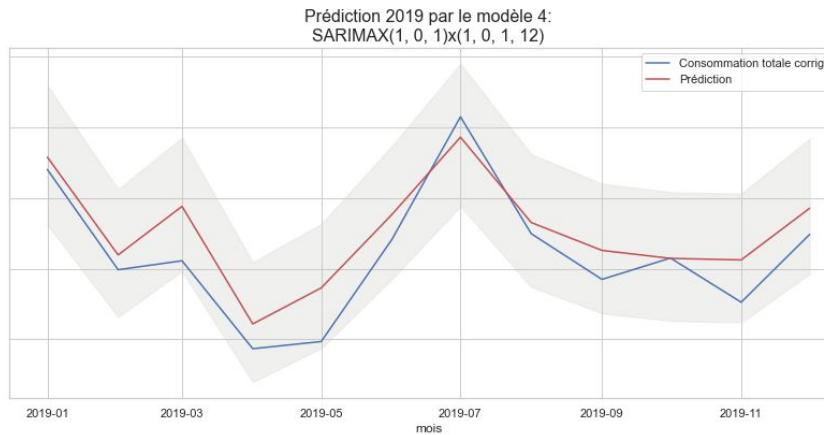


# SARIMA **(1,0,1)(1,0,1,12)** prévision 2019

Les prévisions sont satisfaisantes mais perfectibles



# Zoom sur les données



**SEUIL DE CONFIANCE**  
5 %

**RMSE**  
Root Mean Square Error : **868.29**

**MAPE**  
Mean Average Percentage Error : **0.024**

01

02

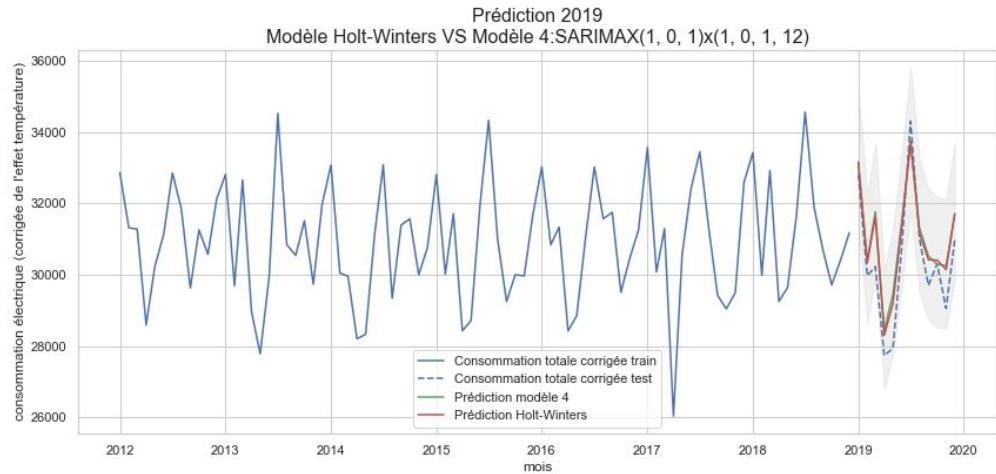
03

# SARIMA $(1, 0, 1)(1, 0, 1, 12)$

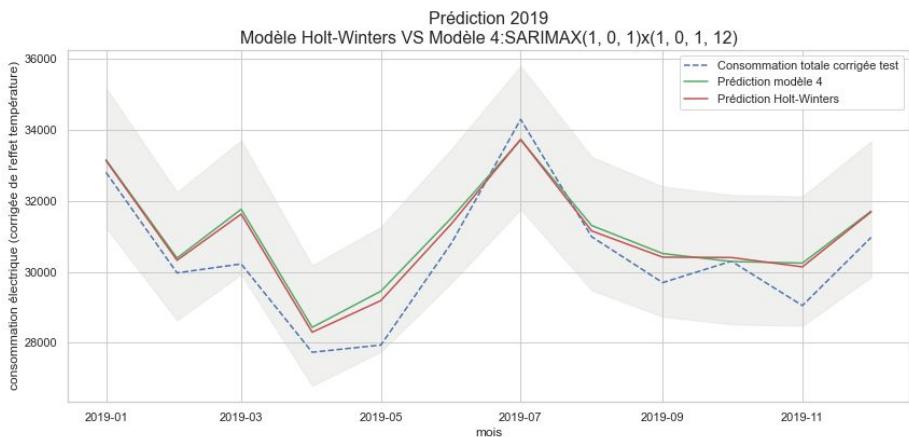
## VS

# Holt- Winters

Les prévisions semblent sensiblement identiques



# Zoom sur les données



**SEUIL DE CONFIANCE**  
5 %

**RMSE**  
HW : 762.13  
SAMIRA : 868.29

**MAPE**  
HW : 0.021  
SAMIRA : 0.024

01

02

03

# Modèle SAMIRA **(1, 0, 0)(1, 0, 1, 12)**

Les coefficients sont tous significatifs au seuil de 5%.

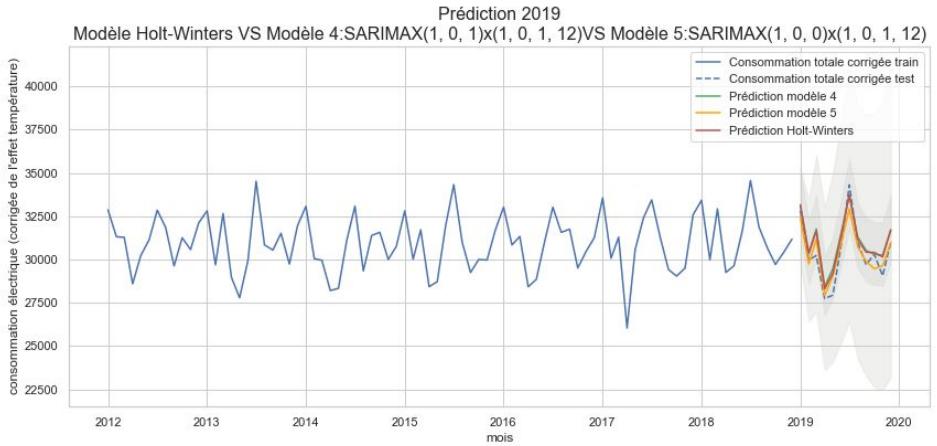
AIC plus élevé (-254 au lieu de -302)

Le résidu conforme à un bruit blanc

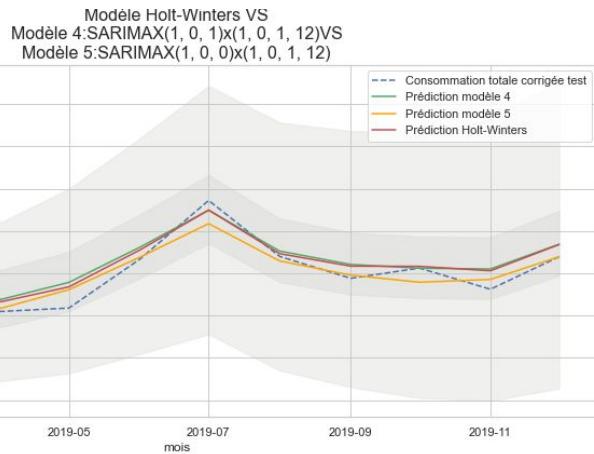
```
SARIMAX Results
=====
Dep. Variable:                      y      No. Observations:      84
Model:             SARIMAX(1, 0, 0)x(1, 0, [1], 12)   Log Likelihood   131.062
Date:                Wed, 19 Jan 2022   AIC                  -254.123
Time:                      20:07:47     BIC                  -244.400
Sample:                           0      HQIC                 -250.215
                                         - 84
Covariance Type:            opg
=====
              coef    std err        z   P>|z|    [0.025    0.975]
-----
ar.L1      0.9999    0.000   4323.632      0.000      0.999    1.000
ar.S.L12   0.9737    0.043    22.430      0.000      0.889    1.059
ma.S.L12   -0.6959   0.237    -2.939      0.003     -1.160   -0.232
sigma2     0.0018    0.000     7.112      0.000      0.001    0.002
=====
Ljung-Box (L1) (Q):                  12.45    Jarque-Bera (JB):       7.07
Prob(Q):                            0.00    Prob(JB):           0.03
Heteroskedasticity (H):               1.23    Skew:                 0.46
Prob(H) (two-sided):                  0.59    Kurtosis:            4.09
=====
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
Retard : p-value
6 : 0.9929638039738028
12 : 0.9964819016036152
18 : 0.9976546010690767
24 : 0.9982409508018075
30 : 0.998592760641446
36 : 0.9988273005345383
48 : 0.9991204754009037
```

# SARIMA $(1,0,0)(1,0,1,12)$ Meilleur modèle ?

Prédiction sur l'année  
2019



# Zoom sur les données



**SEUIL DE CONFIANCE**  
5 %

01

**RMSE**  
HW : 762.13  
SAMIRA : 668.41

02

**MAPE**  
HW : 0.021  
SAMIRA : 0.016

03



# Différenciation de la série temporelle

Test de Dickey Fuller :  
p value = 1.07e-5  
La série temporelle est  
stationnaire.

