



PRÉDISEZ LA DEMANDE EN ÉLECTRICITÉ

PARCOURS DATA ANALYST – OPENCLASSROOMS

NAYESCHA GOMES



SOMMAIRE

- I. Le jeu des données et son traitement**
- II. Correction des données de consommation mensuelles de l'effet température**
- III. Désaisonnalisation de la consommation**
- IV. Prédiction de la consommation**
- V. Conclusion**

The background features a stylized, light-colored illustration of a sustainable landscape. It includes solar panels on the left, wind turbines on the right, and a small house in the foreground. The sky is light blue with a sun and clouds. The entire scene is framed by decorative circuit-like lines with circular nodes in the corners.

SOMMAIRE

I. Le jeu des données et son traitement

II. Correction des données de consommation mensuelles de l'effet température

III. Désaisonnalisation de la consommation


IV. Prédiction de la consommation

V. Conclusion

LE JEU DES DONNÉES ET SON TRAITEMENT

Pour ce projet, nous avons utilisé deux bases de données pour estimer la demande énergie.

- La première base contient les données mensuelles de consommation totale d'électricité en énergie en France et par région, de 2012 à novembre 2021
- Elle a été téléchargée directement sur [le site de la RTE](#) (Réseau de Transport d'Electricité).



Télécharger les indicateurs

Téléchargements des données en puissances

France ▼

Journée unique :

En-cours mensuel temps réel

En-cours annuel consolidé

Annuel définitif 2012

Téléchargements des données en énergie

Données mensuelles

	Mois	Qualité	Territoire	Production totale	Production nucléaire	Production thermique totale	Production thermique charbon	Production thermique fioul	Production thermique gaz	Production hydraulique	...	Production bioénergies	Consommation totale
0	2012-01-01	Données définitives	France	57177	42811.0	5399	741.0	691	3967	6741	...	514	51086
1	2012-02-01	Données définitives	France	54419	38661.0	8721	2511.0	1309	4901	5244	...	454	54476

1403 rows x 22 columns

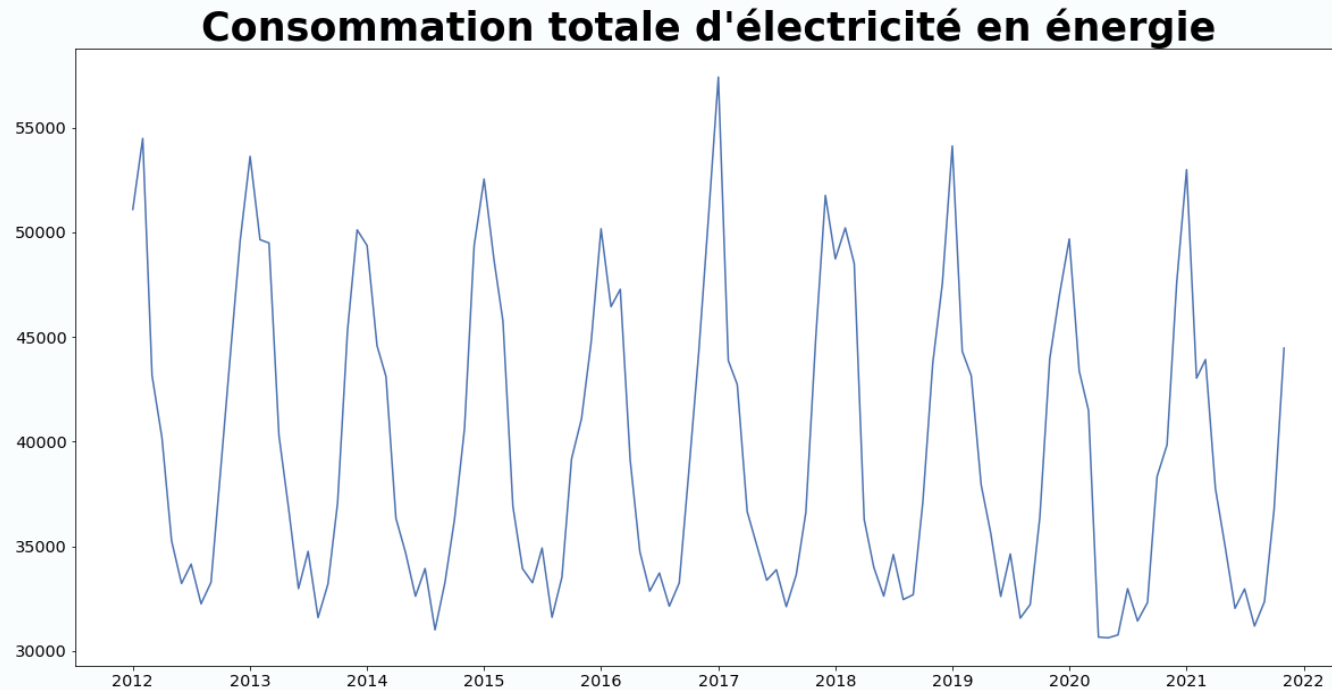
LE JEU DES DONNÉES ET SON TRAITEMENT

Traitement et correction des données :

Pour la base des données de **consommation des ménages français**, nous avons fait la vérification de l'existence de valeurs nulles, des doublons et des valeurs aberrantes. Nous avons décidé de garder que les colonnes « mois » et « consommation totale » pour la suite du projet.

mois	consommation
2012-01-01	51086
2012-02-01	54476
2012-03-01	43156
2012-04-01	40176
2012-05-01	35257
...	...
2021-07-01	32959
2021-08-01	31184
2021-09-01	32350
2021-10-01	36762
2021-11-01	44458

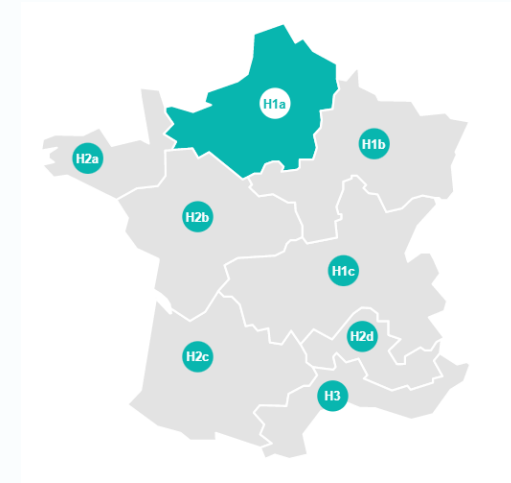
119 rows × 1 columns



LE JEU DES DONNÉES ET SON TRAITEMENT

La deuxième base contient les données météo qui ont été utilisées pour corriger l'effet de la température sur la consommation :

- Ces données sont issues du [site de la CEGIBAT](#), le centre d'expertise efficacité énergétique de la GRDF (Gaz Réseau Distribution France).
- L'outil, réalisé en partenariat avec Météo France, permet de calculer les degrés jour (DJ ou DJU) chauffage ou climatisation sur une période, une station météo et un seuil de température donnés.
- Pour ce projet la station météo retenue est celle à Rossy, la méthode de calcul « météo », le type d'usage « chauffage » pour une température de référence de 18°C ont été sélectionnés.
- La période disponible va de janvier 2009 à mi-juin 2021.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	NaN	JAN	FÉV	MAR	AVR	MAI	JUN	JUL	AOÛ	SEP	OCT	NOV	DÉC	Total
1	2021.0	421.1	324.5	301.6	274.5	162.7	3.7	0	0	0	0	0	0	1487.8
2	2020.0	367	270.1	297.9	111.7	94.3	27.3	3.9	6.5	39.7	166.2	246.3	353.1	1983.8
3	2019.0	433.4	288.5	259.8	197.6	147.1	31.2	3.2	2.9	37.7	146.5	295.5	347.7	2190.9
4	2018.0	324.9	453.6	333.6	137	80	15.2	0.1	5.2	50.4	138.4	297	346	2181.3
5	2017.0	499.6	298.6	225.6	224.6	88.8	14	1.8	12.2	72.8	113.2	302.9	390.4	2244.2
6	2016.0	384.6	341.8	352.1	243.6	115.4	38.6	9.8	6.2	16.8	197.2	310.3	418.6	2434.9
7	2015.0	414	388.3	307.2	178.5	117.8	33.6	12.3	8.7	94.2	200.5	211	256.1	2221.9
8	2014.0	341.4	292.9	253.9	164.8	126.5	39.7	15.7	26.2	31.8	101.6	236.6	396.2	2026.9
9	2013.0	455.2	435.7	412.5	234.9	184.4	63.6	6.2	9.2	55.6	118	327.7	363.5	2666.2
10	2012.0	361.5	468.6	241.4	254.3	104.6	50.7	23.1	4.2	77.3	169.5	313.3	362.7	2430.9
11	2011.0	414.2	324.6	275.7	107.6	71.2	44.6	32.5	19.7	30.8	142.2	248	331.8	2042.6
12	2010.0	531.8	390.7	321.9	202.7	169.1	39.6	0	22.5	72	194.3	329.3	551.1	2824.8
13	2009.0	513.8	386.7	322.2	154.8	108.5	54.2	4.6	2.3	43.1	170.6	232.6	438.8	2432

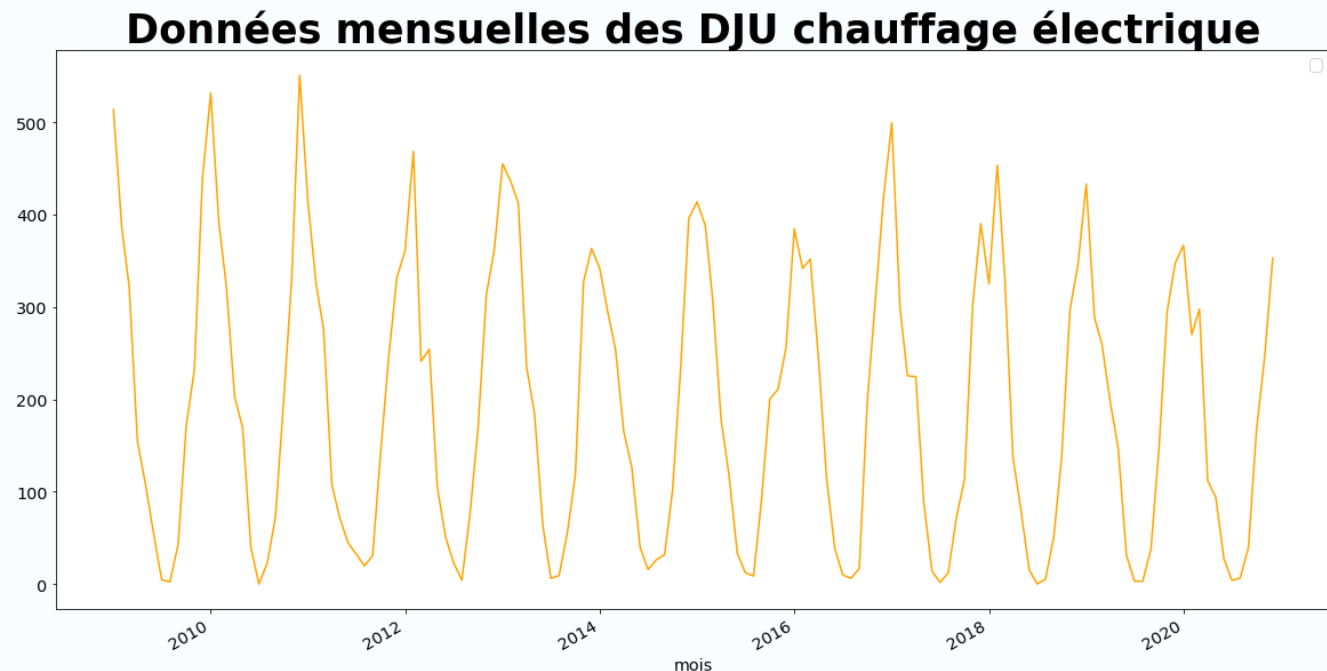
LE JEU DES DONNÉES ET SON TRAITEMENT

Traitement et correction des données :

Pour la base des données qui continent **les données météo**, nous avons de ne pas prendre en compte les données de l'année 2021 car ils s'arrêtent au mois de juillet et cela peut compromettre notre analyse. La base n'a pas d'autres valeurs manquantes. Un traitement de forme a été fait pour faciliter l'intégration des données dans le cadre d'une jointure des deux bases.

dju_chauffage	
mois	
2020-01-01	367.0
2020-02-01	270.1
2020-03-01	297.9
2020-04-01	111.7
2020-05-01	94.3
...	...
2009-08-01	2.3
2009-09-01	43.1
2009-10-01	170.6
2009-11-01	232.6
2009-12-01	438.8

144 rows × 1 columns



LE JEU DES DONNÉES ET SON TRAITEMENT

A partir des deux précédentes sources de données, un échantillon a été constitué à partir de la méthode « merge inner » :

- Années : 2012 à 2020
- Nombre d'observations : 108
- Aucune valeur aberrante
- Aucune donnée manquante
- Aucun doublon

	consommation	dju_chauffage
mois		
2012-01-01	51086	361.5
2012-02-01	54476	468.6
2012-03-01	43156	241.4
2012-04-01	40176	254.3
2012-05-01	35257	104.6
...
2020-08-01	31428	6.5
2020-09-01	32307	39.7
2020-10-01	38332	166.2
2020-11-01	39840	246.3
2020-12-01	47565	353.1

108 rows × 2 columns



SOMMAIRE

I. Le jeu des données et son traitement

II. Correction des données de consommation mensuelles de l'effet température

III. Désaisonnalisation de la consommation

IV. Prédiction de la consommation

V. Conclusion

CORRECTION DES DONNÉES DE CONSOMMATION MENSUELLES DE L'EFFET TEMPÉRATURE

L'effet température (due au chauffage électrique) sera corrigé par **Régression linéaire** :

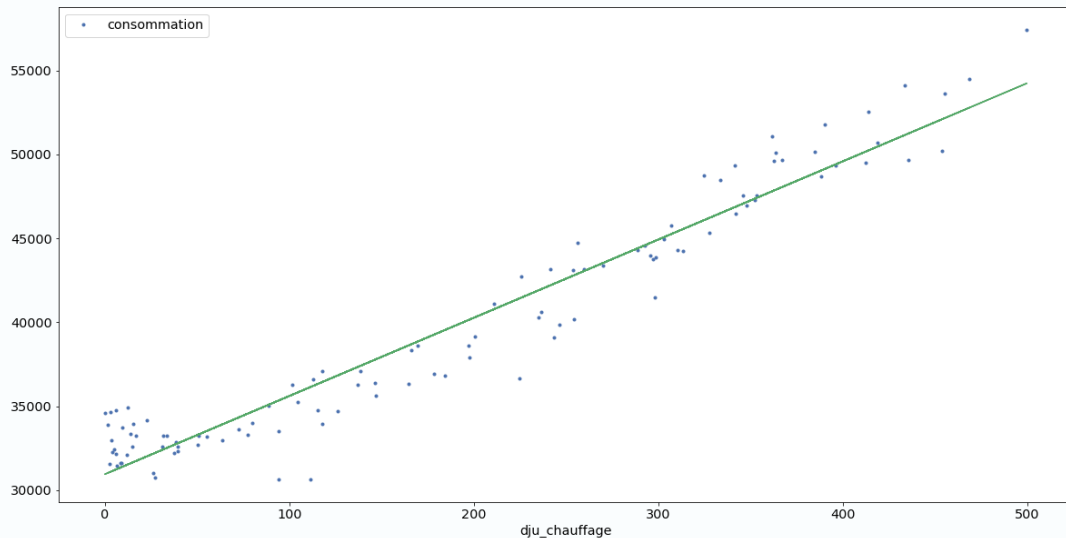
- La modélisation a été faite à partir de Statsmodels, mais également par Sckit-learn à des fins de comparaison.
- La qualité du modèle est estimée à 93% (coefficient de détermination R^2).

```
OLS Regression Results
=====
Dep. Variable:      consommation      R-squared:      0.933
Model:              OLS               Adj. R-squared:  0.932
Method:             Least Squares     F-statistic:    1476.
Date:               Wed, 09 Feb 2022   Prob (F-statistic): 4.86e-64
Time:               11:42:55          Log-Likelihood: -964.31
No. Observations:   108              AIC:            1933.
Df Residuals:       106              BIC:            1938.
Df Model:           1
Covariance Type:    nonrobust
=====
                    coef    std err          t      P>|t|      [0.025      0.975]
-----
const             3.094e+04    289.622    106.831     0.000     3.04e+04     3.15e+04
dju_chauffage     46.6095         1.213     38.418     0.000      44.204      49.015
=====
Omnibus:                 1.868    Durbin-Watson:           1.468
Prob(Omnibus):            0.393    Jarque-Bera (JB):         1.370
Skew:                    -0.250    Prob(JB):                 0.504
Kurtosis:                 3.231    Cond. No.                  390.
=====

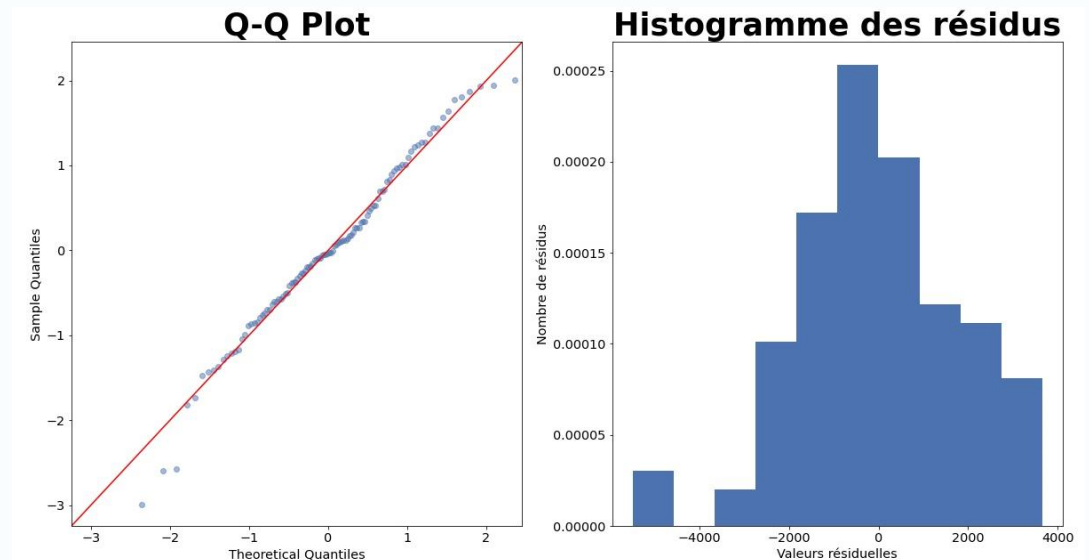
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

CORRECTION DES DONNÉES DE CONSOMMATION MENSUELLES DE L'EFFET TEMPÉRATURE

- Le modèle de régression linéaire illustre la relation dont nous parlons.
- Visuellement nous pouvons distinguer la corrélation entre nos deux variables continues.

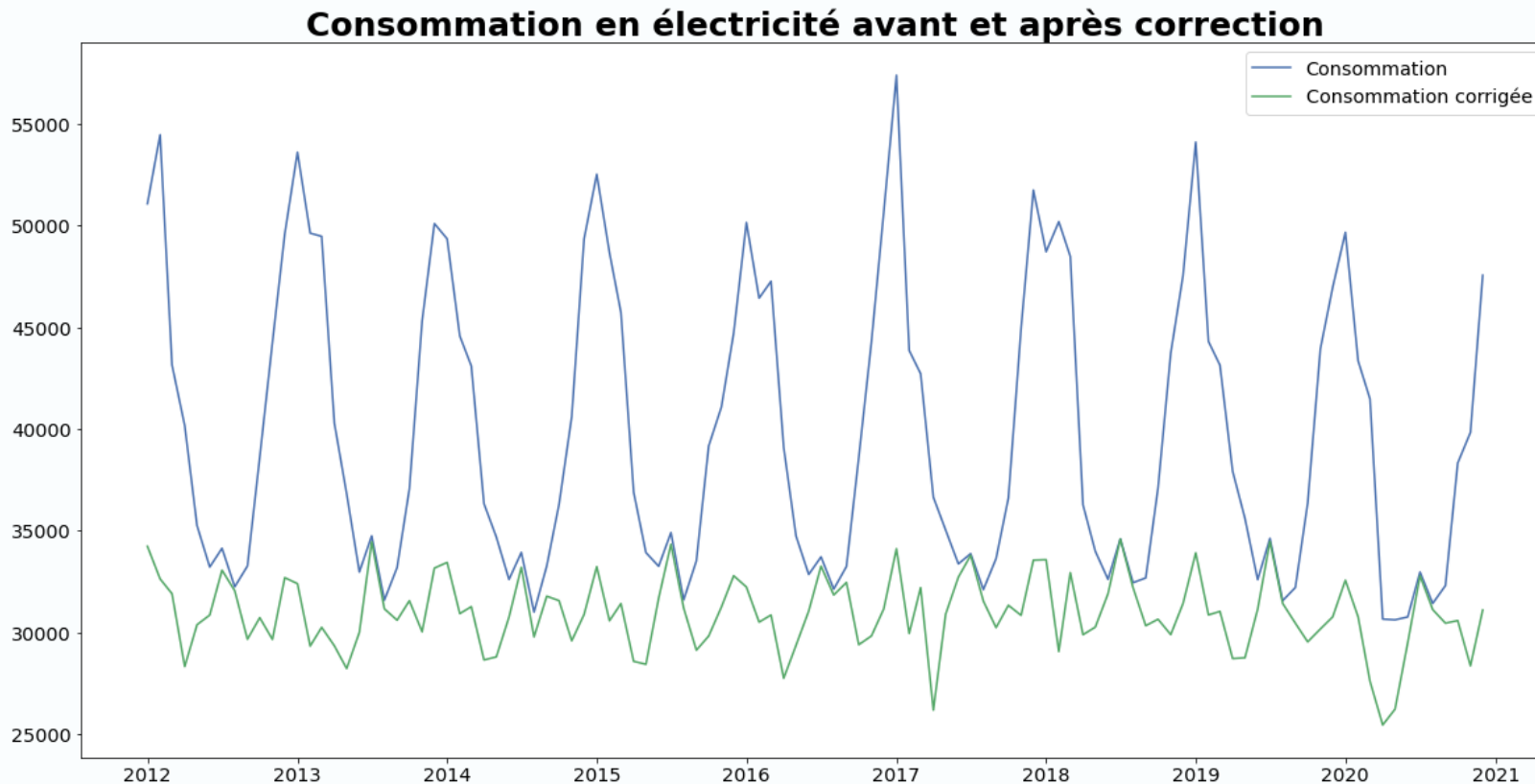


- Notre test statistique du modèle nous amène à penser que les résidus issus de la modélisation sont gaussiens
- La distribution des résidus est alignée avec la distribution théorique d'une loi normale



CORRECTION DES DONNÉES DE CONSOMMATION MENSUELLES DE L'EFFET TEMPÉRATURE

A partir de notre Régression Linéaire nous pouvons créer notre série temporelle de la consommation en électricité corrigée de l'effet température



The background features a stylized, light-colored illustration of a sustainable landscape. It includes rolling green hills, several solar panels mounted on poles, three wind turbines, and two small houses with red roofs. The sky is light blue with a few white clouds and a faint sun. The entire scene is framed by decorative circuit-like lines with circular nodes in the corners.

SOMMAIRE

I. Le jeu des données et son traitement

II. Correction des données de consommation mensuelles de l'effet température

III. Désaisonnalisation de la consommation

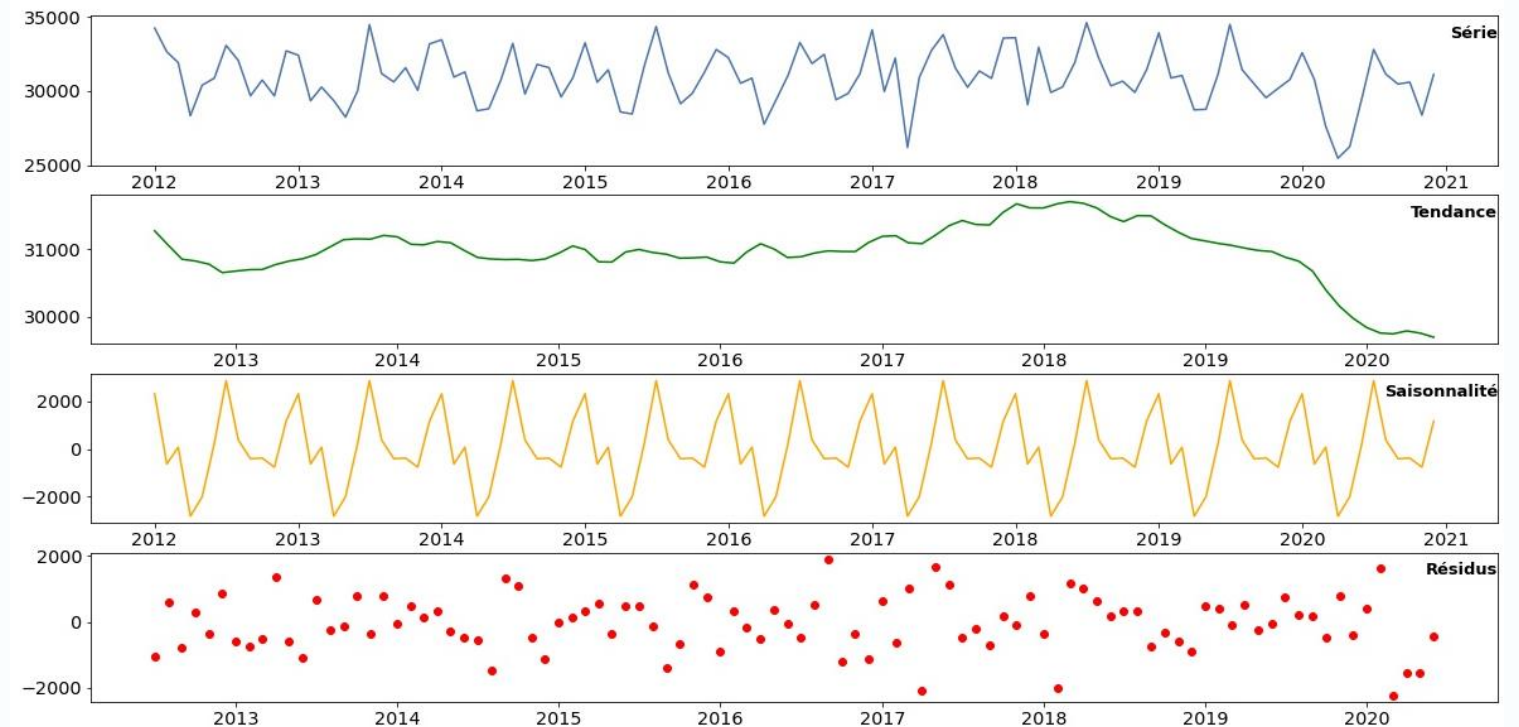
IV. Prévission de la consommation

V. Conclusion

DÉSAISONNALISATION DE LA CONSOMMATION

Pour la désaisonnalisation après la correction de la consommation, nous avons utilisées les moyennes mobiles pour transformer le signal temporel :

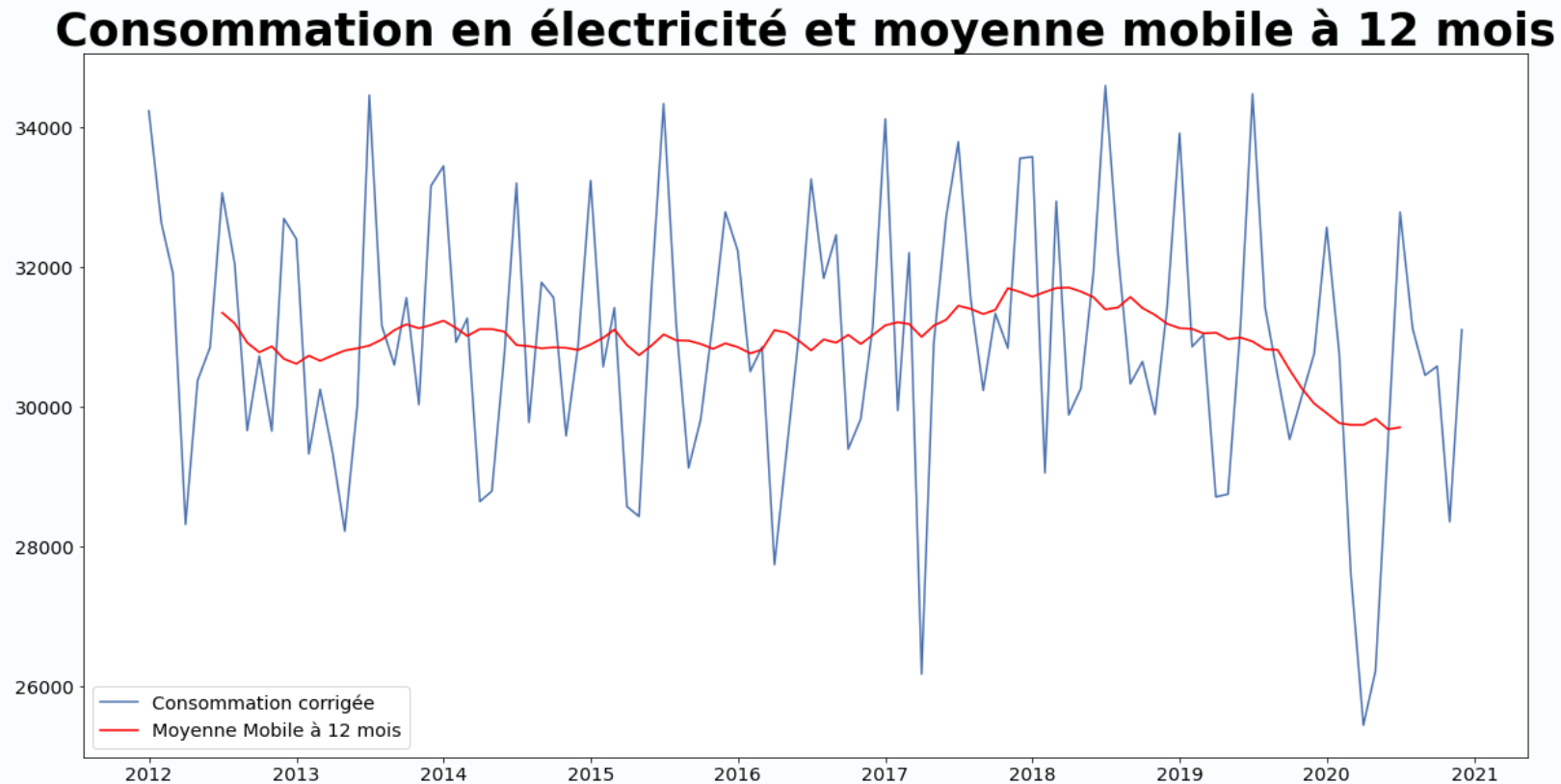
- La méthode « `seasonal_decompose` » de `statsmodels` permet de visualiser la tendance, la saisonnalité et le résidu du signal.
- La décomposition de la série nous permet de retrancher la saisonnalité



DÉSAISONNALISATION DE LA CONSOMMATION

MOYENNE MOBILE SUR 12 MOIS :

La visualisation de la consommation en électricité corrigée comparé à la moyenne mobile à 12 mois pour chaque année nous affiche une pré-représentation graphique de la tendance de la série temporelle désaisonnalisée.

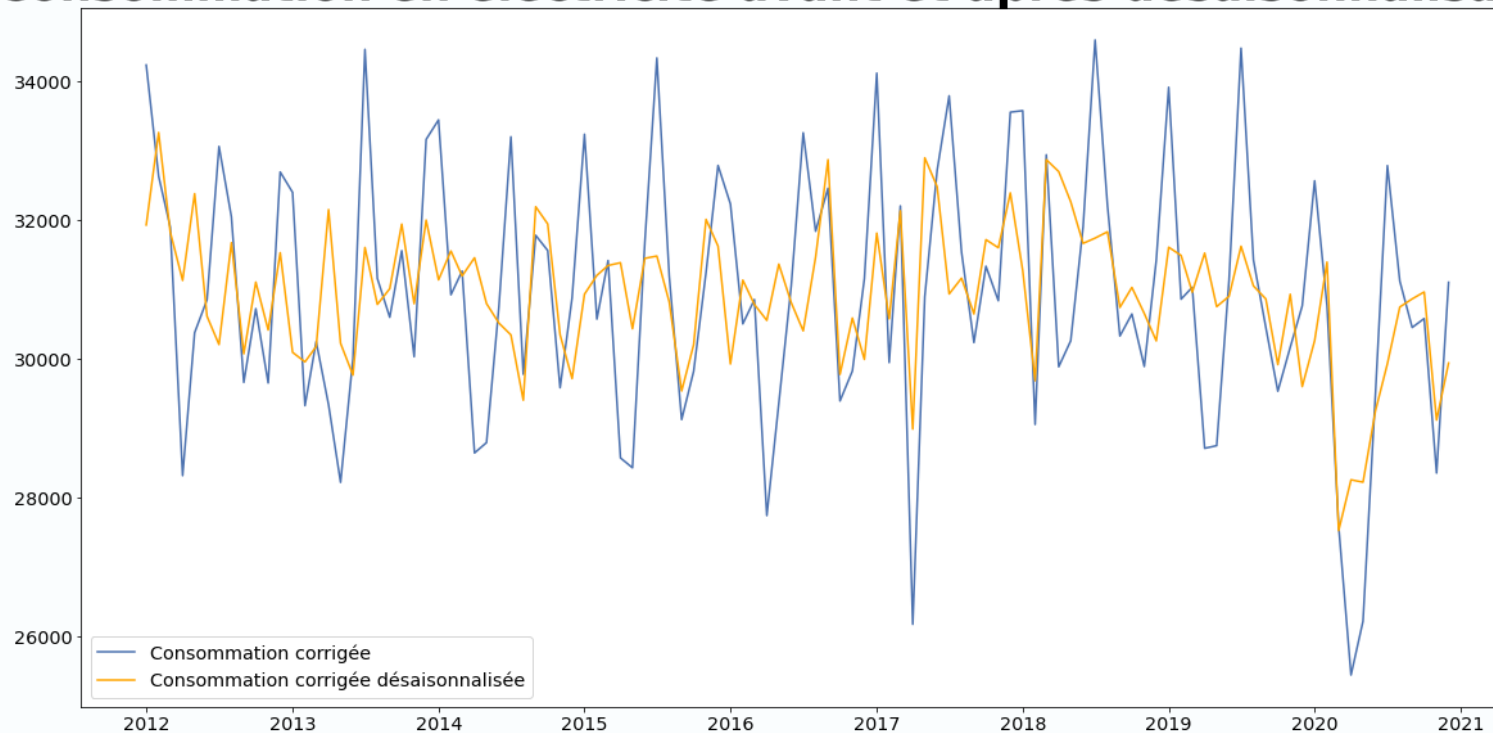


DÉSAISONNALISATION DE LA CONSOMMATION

Différence avec la saisonnalité pour obtenir une série temporelle hors impact saisonnier :

La série temporelle désaisonnalisée reste stable suivant un modèle additif.

Consommation en électricité avant et après désaisonnalisation



The background features a stylized, light-colored illustration of a sustainable landscape. It includes solar panels on the left, wind turbines on the right, and a small house in the foreground. The sky is light blue with a sun and clouds. The entire scene is framed by decorative circuit-like lines with circular nodes in the corners.

SOMMAIRE

I. Le jeu des données et son traitement

II. Correction des données de consommation mensuelles de l'effet température

III. Désaisonnalisation de la consommation

IV. Prédiction de la consommation

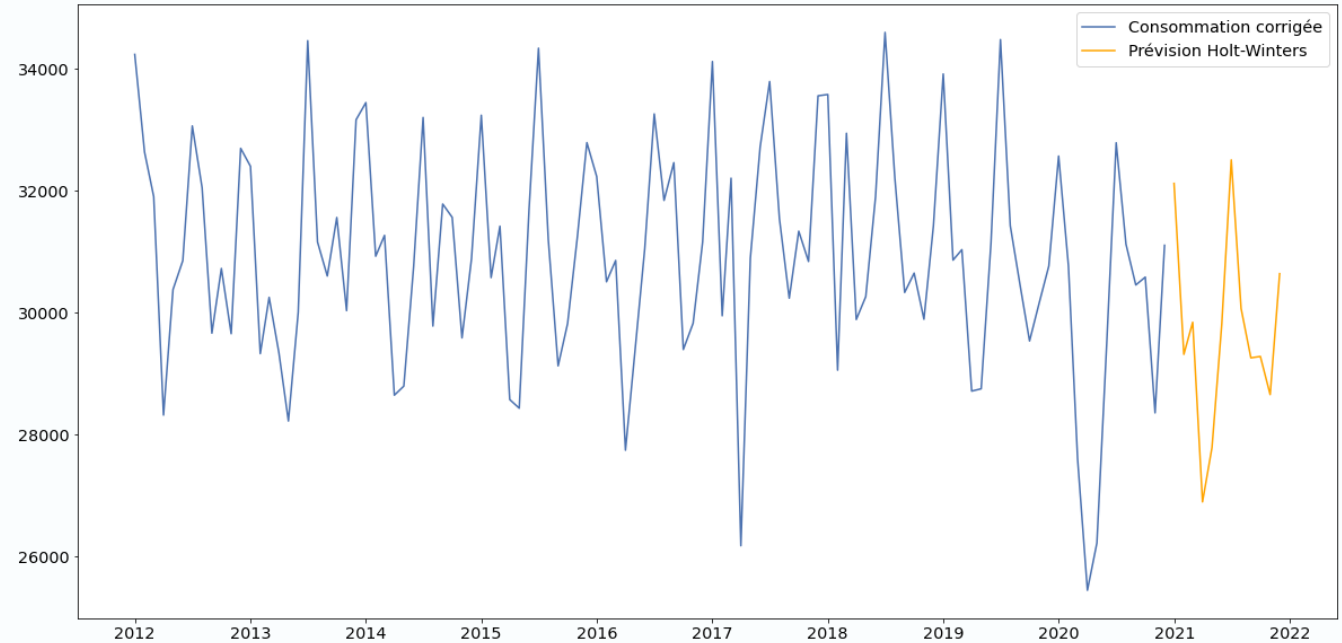
V. Conclusion

PRÉVISION DE LA CONSOMMATION

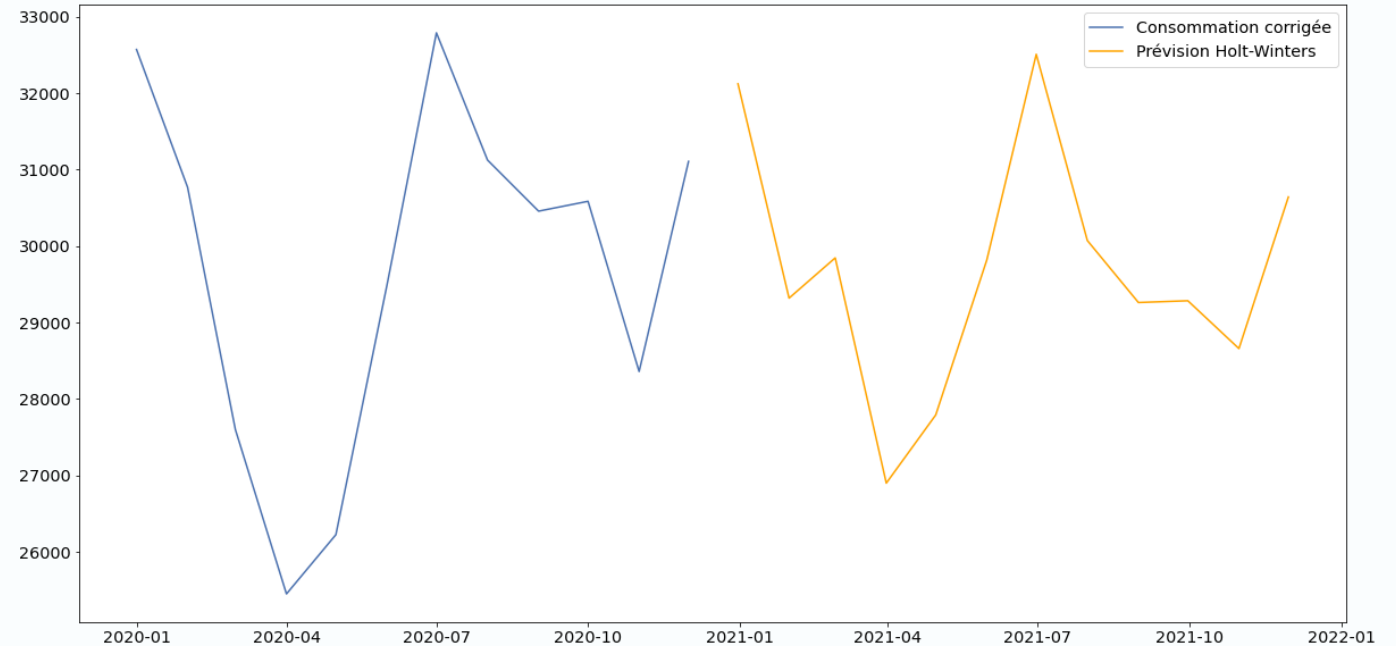
La méthode de Holt-Winters (lissage exponentiel):

- Le modèle est une extension du lissage exponentiel de Holt qui capture la saisonnalité.
- Cette méthode génère des valeurs lissées de façon exponentielle pour le niveau, la tendance et l'ajustement saisonnier de la prévision.

Consommation d'électricité en France - Prévision 2021



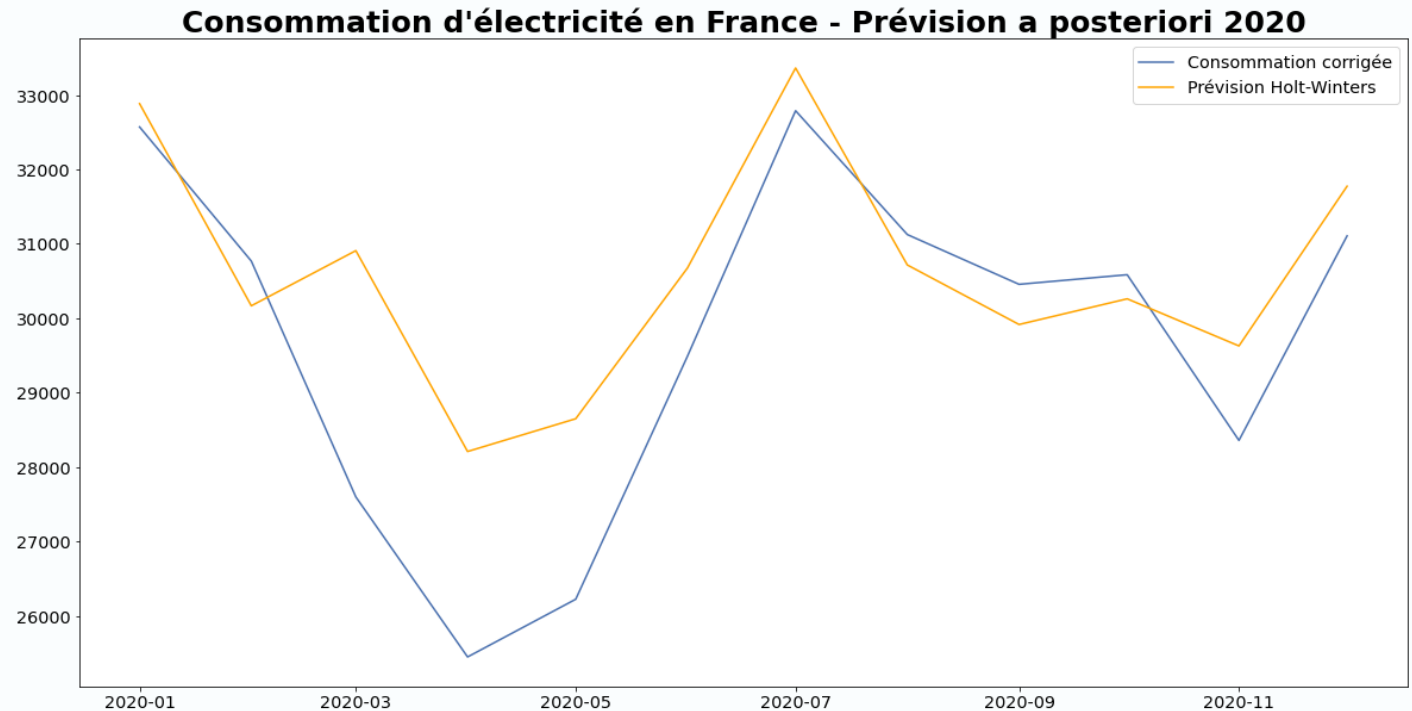
Consommation d'électricité en France 2021 - Prévision 2021



PRÉVISION DE LA CONSOMMATION

Analyse a posteriori sur 12 pour le modèle Holt Winters :

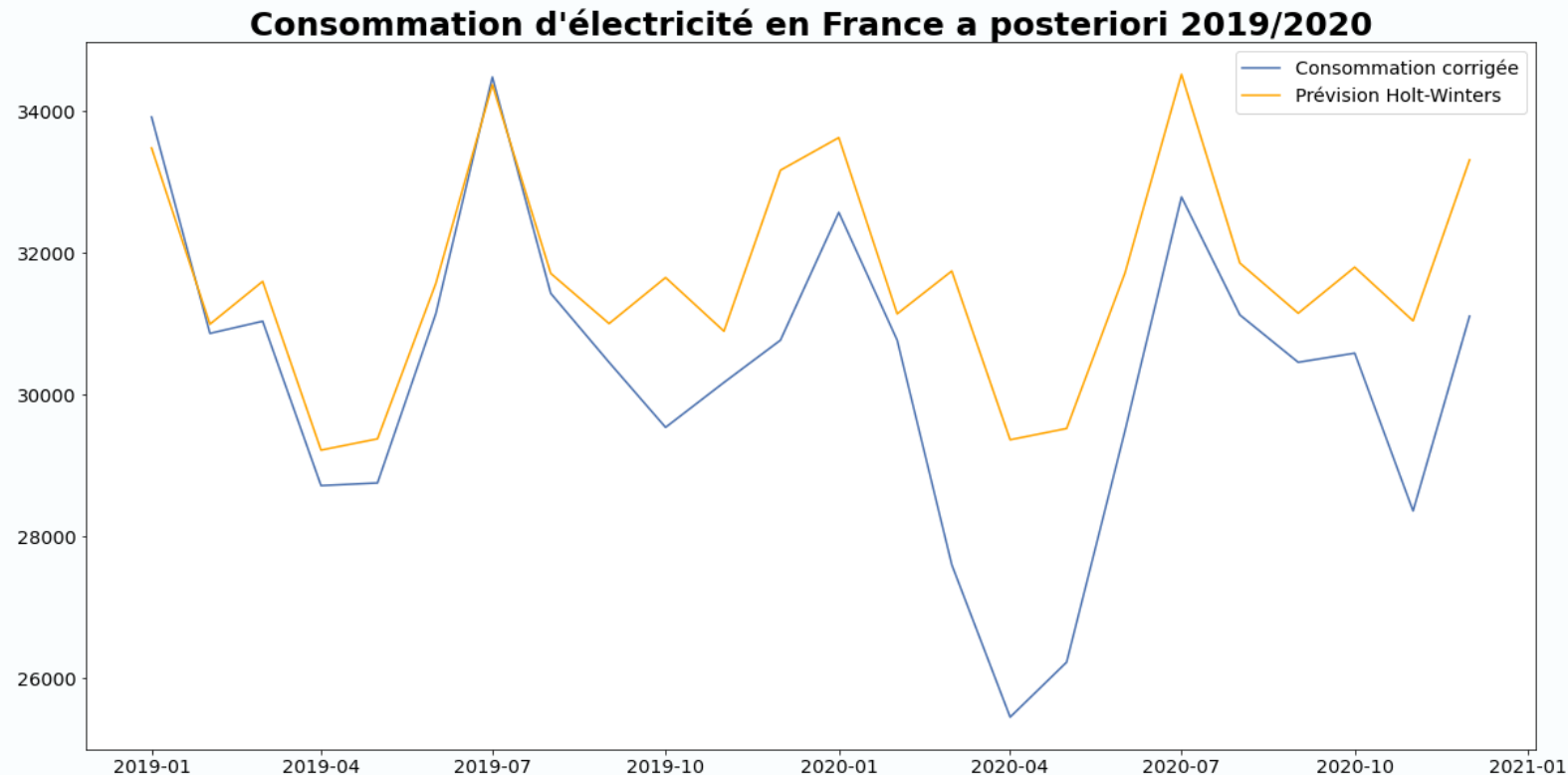
- Pour tester la performance du modèle, nous avons réalisé une prédiction de la dernière année connue (2020)
- Nous pourrions comparer les valeurs prédites par le modèle avec les valeurs réels grâce au calcul de la Mean Absolute Percentage Error (MAPE).
- Etant donné qu'une MAPE inférieure à 10% dénote un modèle plus précis, notre calcul de la MAPE à 4,29% pour cette projection montre que celle-ci est représentative de la réalité.



PRÉVISION DE LA CONSOMMATION

Analyse a posteriori sur 24 pour le modèle Holt Winters :

- Le modèle résiste plutôt bien aux impacts saisonniers sur du long terme avec une MAPE à 4,77%.
- Malgré cela, les excédents des pics et creux ne sont pas totalement prévisibles par le modèle, les prévisions SARIMA seront-elles plus représentatives ?

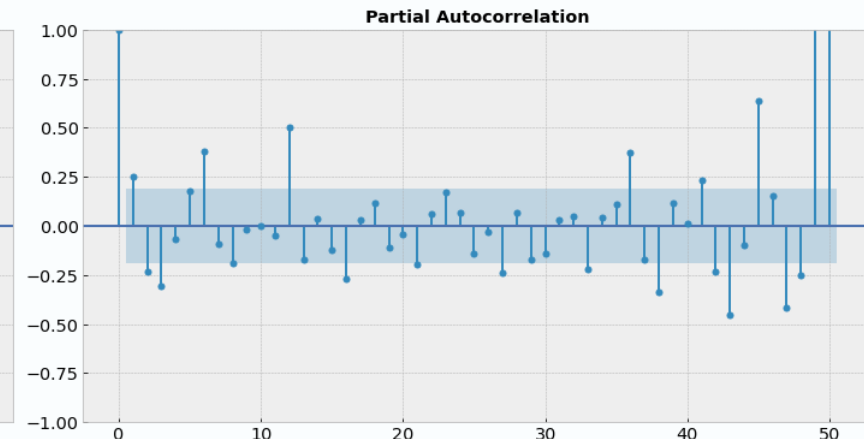
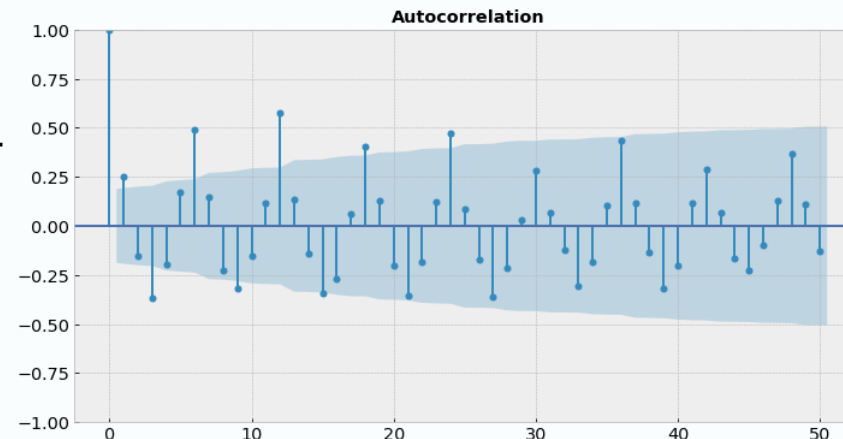


PRÉVISION DE LA CONSOMMATION

Le modèle ARIMA saisonnier de format SARIMA :

Stationnarité et différenciation de la série :

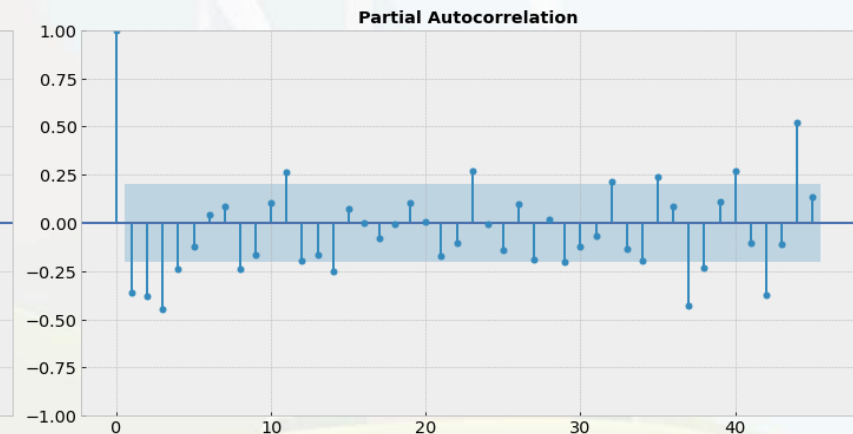
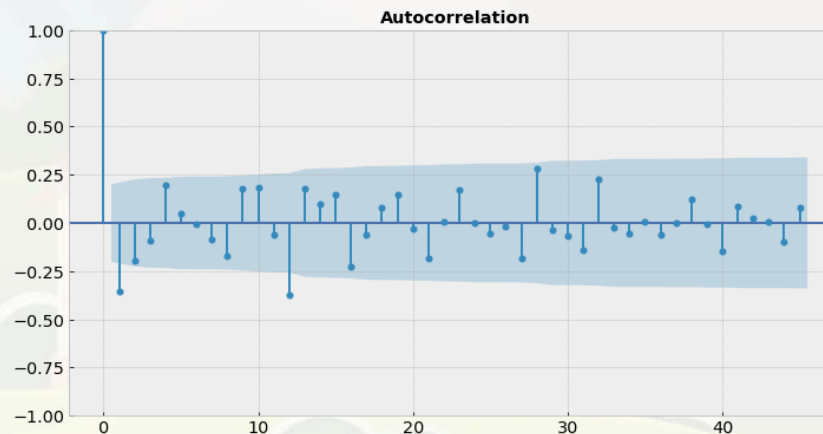
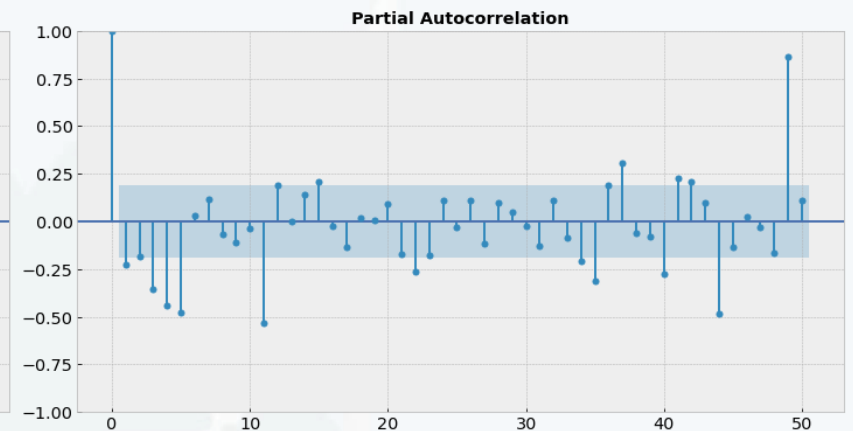
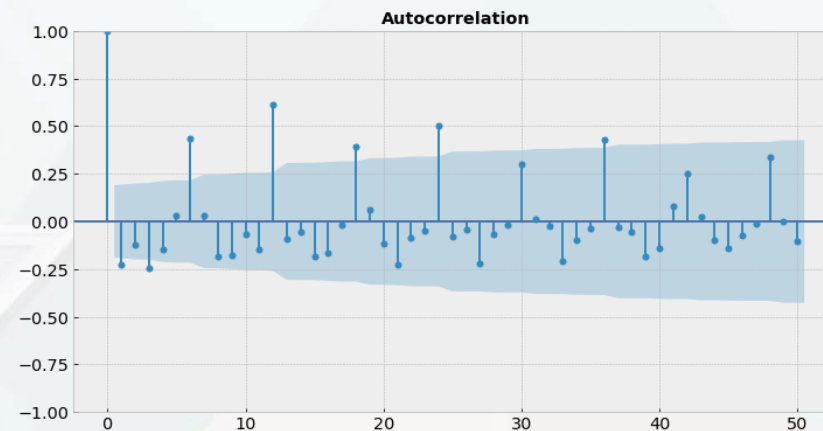
- Nous sommes en présence d'une série temporelle marquée par une forte saisonnalité, classique sur les températures climatiques.
- Il n'est pas possible de modéliser directement par un processus stationnaire, la méthode retenue sera la différenciation.
- Cette méthode génère des valeurs lissées de façon exponentielle pour le niveau, la tendance et l'ajustement saisonnier de la prévision.
- La sortie ACF présente une décroissance lente vers 0, ce qui traduit un problème de non-stationnarité. L'hypothèse nulle à la non-stationnarité de la série ne peut pas être rejetée. Nous devons effectuer donc une première différenciation.



PRÉVISION DE LA CONSOMMATION

Le modèle ARIMA saisonnier de format SARIMA :

- La sortie ACF de la série ainsi différenciée présente encore une décroissance lente vers 0 pour les multiples de 12. Il faut analyser la différenciation d'ordre 12.
- Pour la différenciation, les résultats sont plus satisfaisants et nous pouvons nous appuyer sur les autocorrélogrammes simple et partiels estimés.



PRÉVISION DE LA CONSOMMATION

Le modèle ARIMA saisonnier de format SARIMA

Identification a priori de modèles potentiels :

- Après la stationnarisation effectuée, nous pouvons nous consacrer aux choix potentiels des polynômes AR et MA.
- Ici, nous avons identifié les autocorrélations simples et partielles significatives pour caler ensuite des polynômes qui reflètent ces liens temporels, afin d'émettre une hypothèse sur la partie autorégressive (simple et saisonnière).
- Nous avons testé plusieurs modèles potentiels en regardant l'autocorrélogramme simple et partiel du résidu.
- Après plusieurs essais, nous avons finalement gardé le modèle avec les paramètres suivants :

```

=====
SARIMAX Results
=====
Dep. Variable:          y      No. Observations:      108
Model:                SARIMAX(1, 0, 1)x(1, 1, 1, 12)  Log Likelihood      -830.752
Date:                  Sat, 19 Feb 2022              AIC              1671.505
Time:                  09:55:53                     BIC              1684.327
Sample:                0                            HQIC             1676.688
                                - 108
Covariance Type:      opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.L1          -0.3769        0.651      -0.579      0.562      -1.652        0.899
ma.L1           0.4471        0.650       0.688      0.491      -0.826        1.720
ar.S.L12        0.0069        0.591       0.012      0.991      -1.151        1.165
ma.S.L12       -0.0457        0.591      -0.077      0.938      -1.205        1.114
sigma2         1.834e+06      2.82e+05      6.499      0.000      1.28e+06      2.39e+06
=====
Ljung-Box (L1) (Q):                2.14   Jarque-Bera (JB):                0.48
Prob(Q):                           0.14   Prob(JB):                  0.79
Heteroskedasticity (H):              1.29   Skew:                      0.09
Prob(H) (two-sided):                0.47   Kurtosis:                  2.71
=====

```

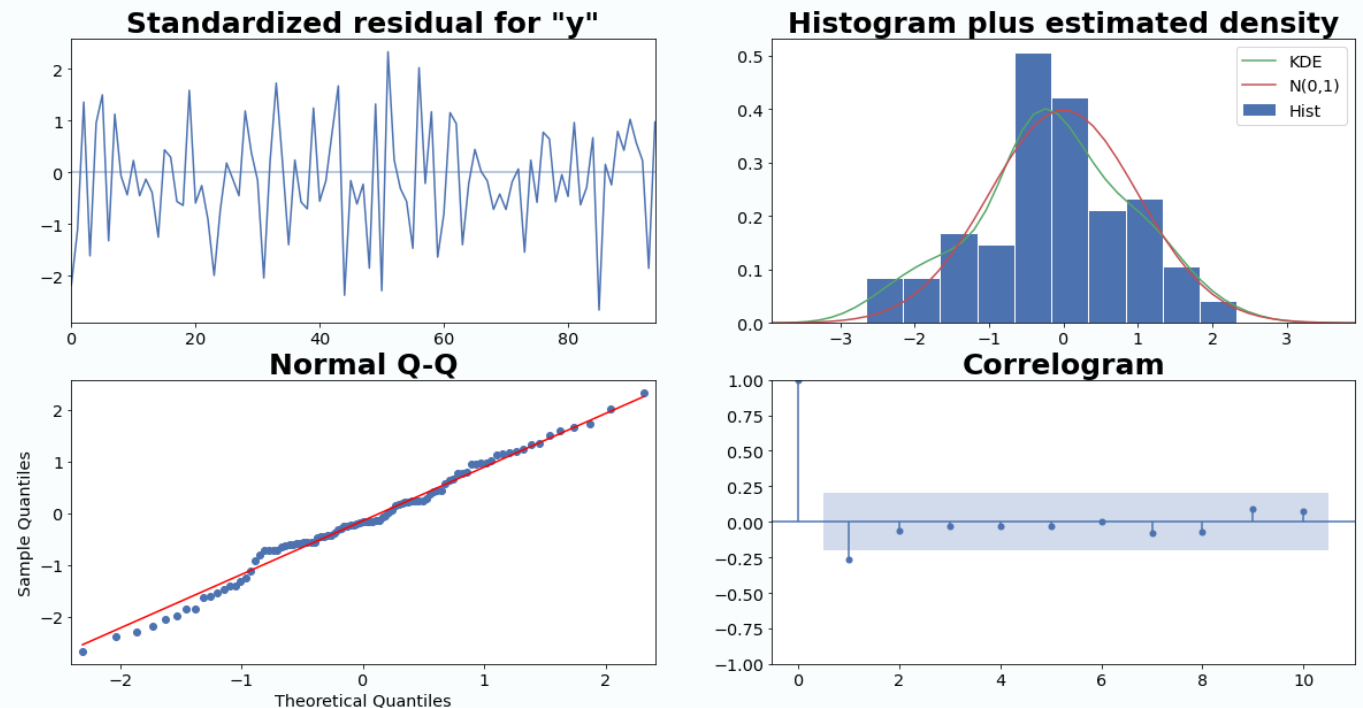
Sarima (1,0,1)(1,1,1,12)

PRÉVISION DE LA CONSOMMATION

Le modèle ARIMA saisonnier de format SARIMA :

Vérification de la normalité des résidus

- Dans le cadre de prévisions, il convient de vérifier la normalité des résidus pour tester l'adéquation de notre modèle. La normalité peut-être détectée de façon graphique et avec l'utilisation des tests statistiques.
- La représentation "Standardized residual" et "Correlogram" confirment qu'il n'y a pas de corrélation entre les résidus.
- Les résidus sont normalement distribués.
- La distribution ordonnée des résidus représentée par le du Q-Q plot est globalement satisfaisant, il y a quand même des petites divergences vers les queues de distribution.

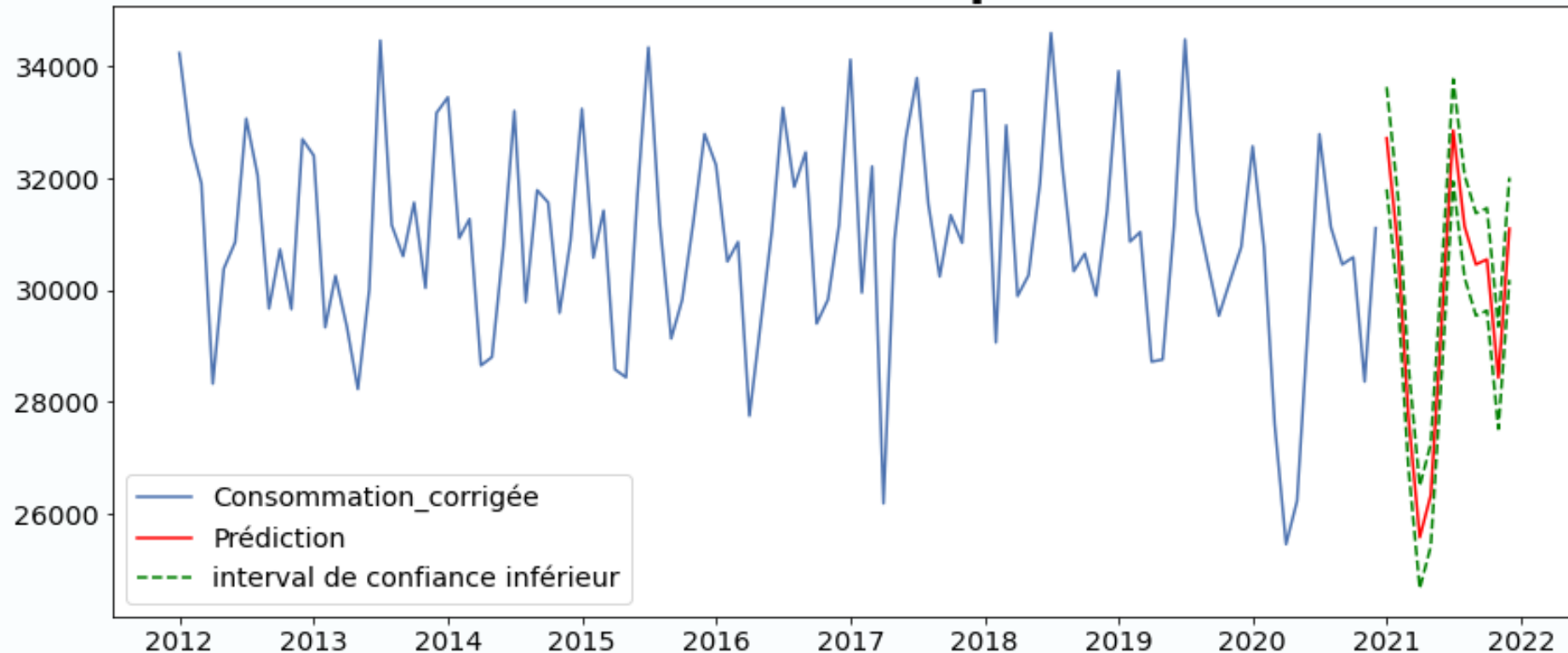


PRÉVISION DE LA CONSOMMATION

Le modèle ARIMA saisonnier de format SARIMA :

- Nous observons que le modèle englobe bien l'année 2021 sur un intervalle de confiance à 95%.

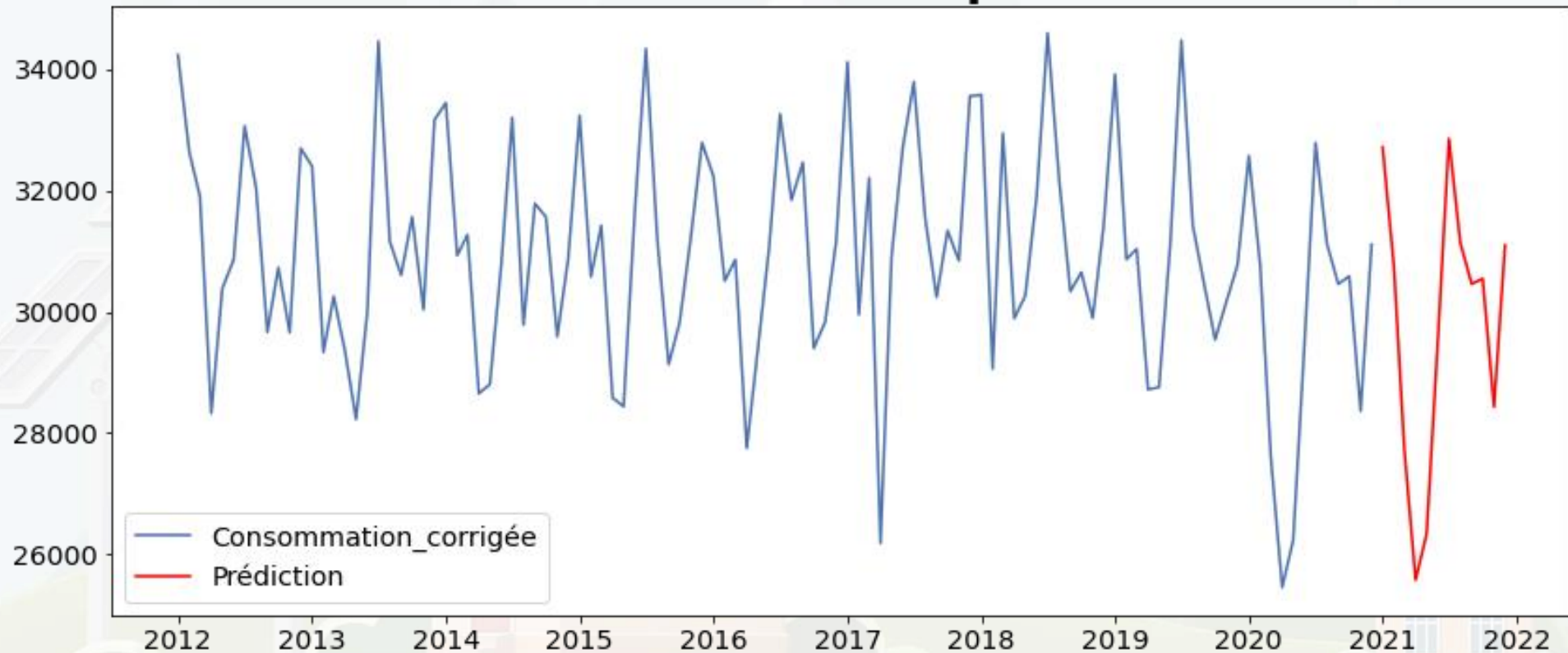
Prédiction de la consommation électrique à l'aide du model SARIMA



PRÉVISION DE LA CONSOMMATION

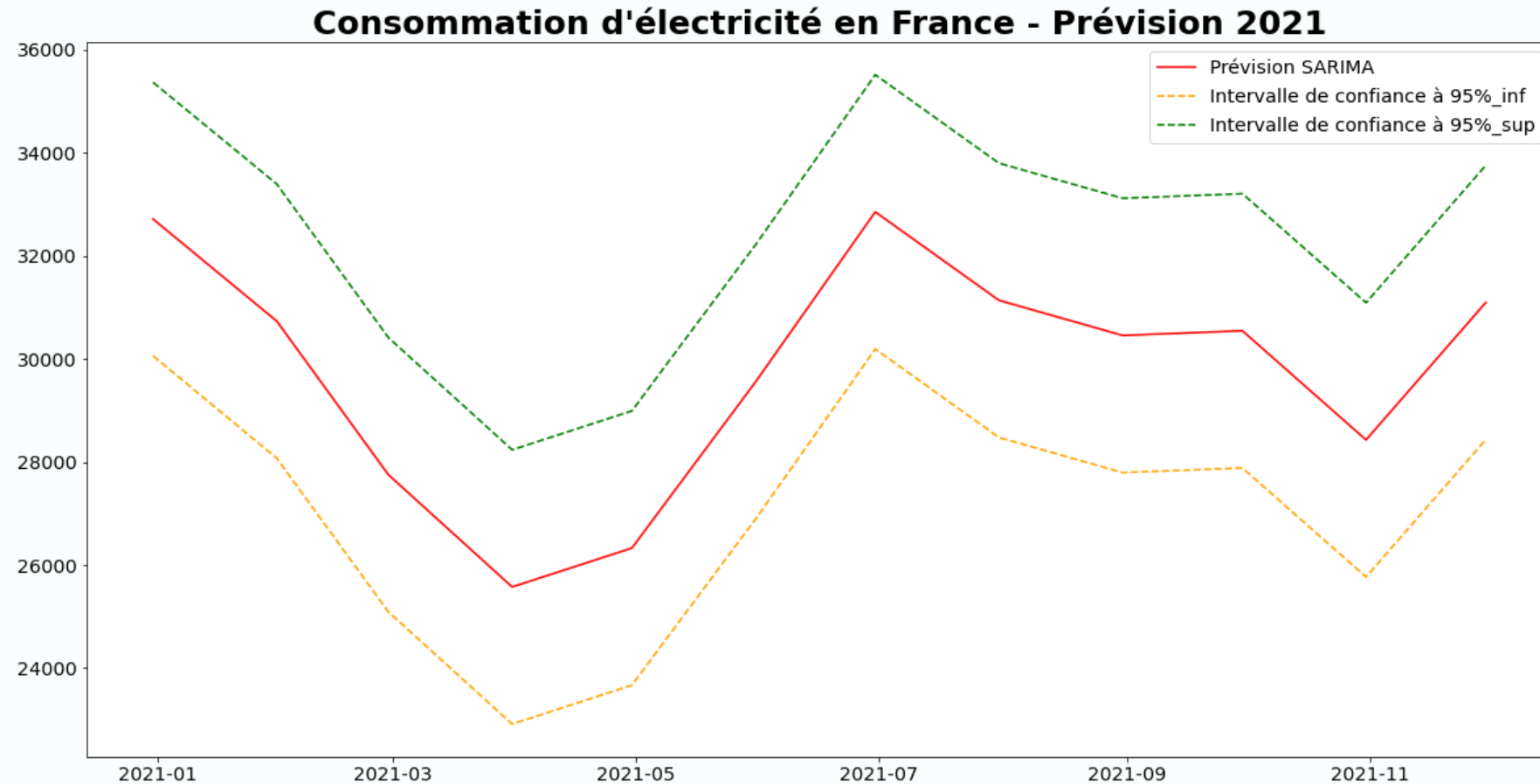
Le modèle ARIMA saisonnier de format SARIMA :

Prédiction de la consommation électrique à l'aide du model SARIMA



PRÉVISION DE LA CONSOMMATION

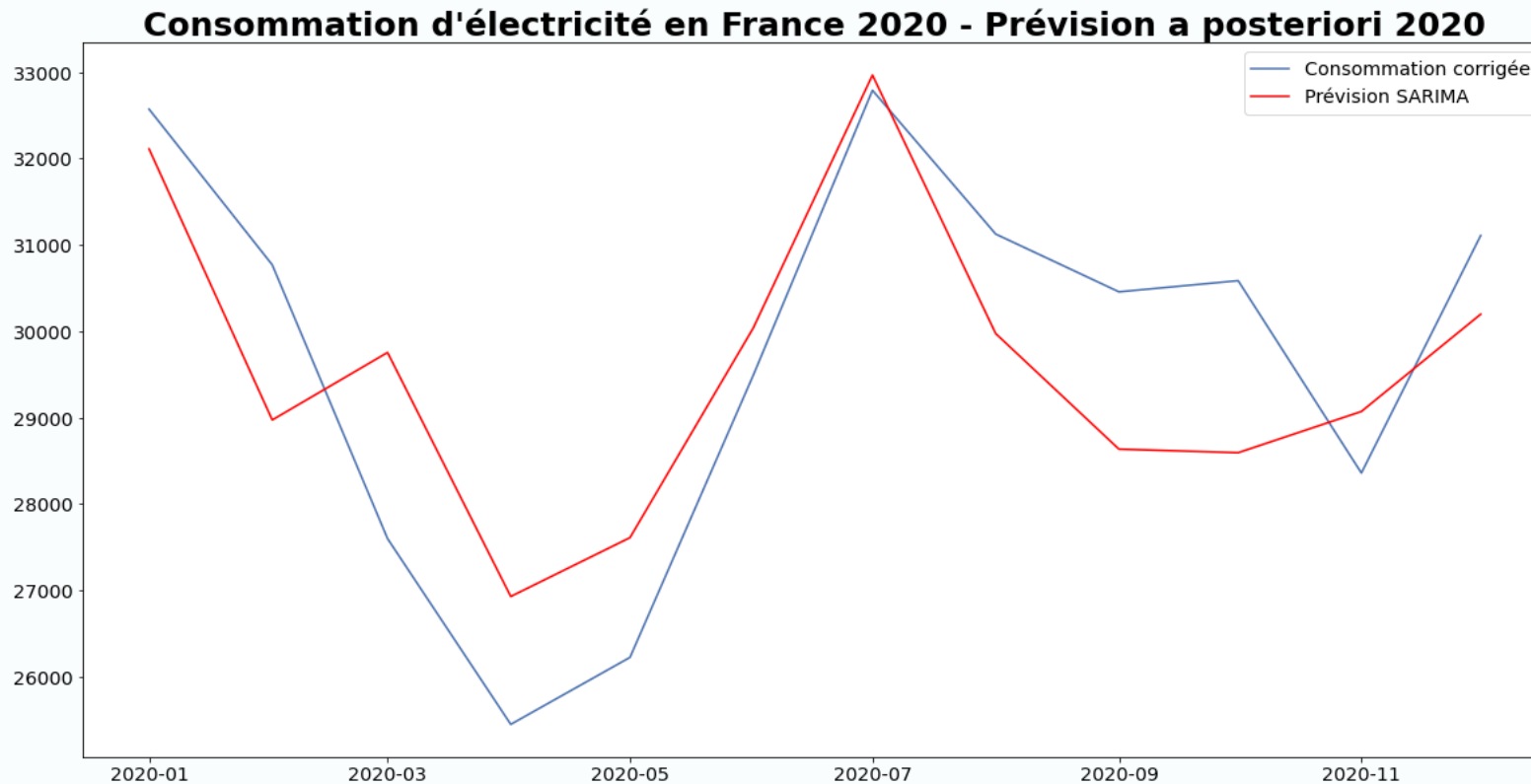
Le modèle ARIMA saisonnier de format SARIMA :



PRÉVISION DE LA CONSOMMATION

Le modèle ARIMA saisonnier de format SARIMA : prévision à posteriori sur 12 mois

- L'analyse à posteriori permet de valider avec appui du critère d'erreur (4,18%) que les écarts entre les prévisions et les réalisations sont acceptables.





SOMMAIRE

I. Le jeu des données et son traitement

II. Correction des données de consommation mensuelles de l'effet température

III. Désaisonnalisation de la consommation

IV. Prédiction de la consommation

V. Conclusion

CONCLUSION

- Le comparatif a posteriori sur 12 mois des projections SARIMA x Holt-Winters, nous montre que les deux modèles sont performantes.
- SARIMA, pour cette projection, affiche une MAPE légèrement moins élevé.
- La prévision à court terme reste plus précise que celles à long terme (24 mois et plus)
- Ces modèles n'intègrent pas de variables exogènes, dans le cotexte métier il est donc nécessaire d'incorporer d'autres variables climatiques et des signaux de prix pour une analyse plus complète.

2020	SARIMA	Holt-Winters
MAPE	4,18	4,29

Consommation d'électricité et Prévision a posteriori 2020 - SARIMA x Holt-Winters

