

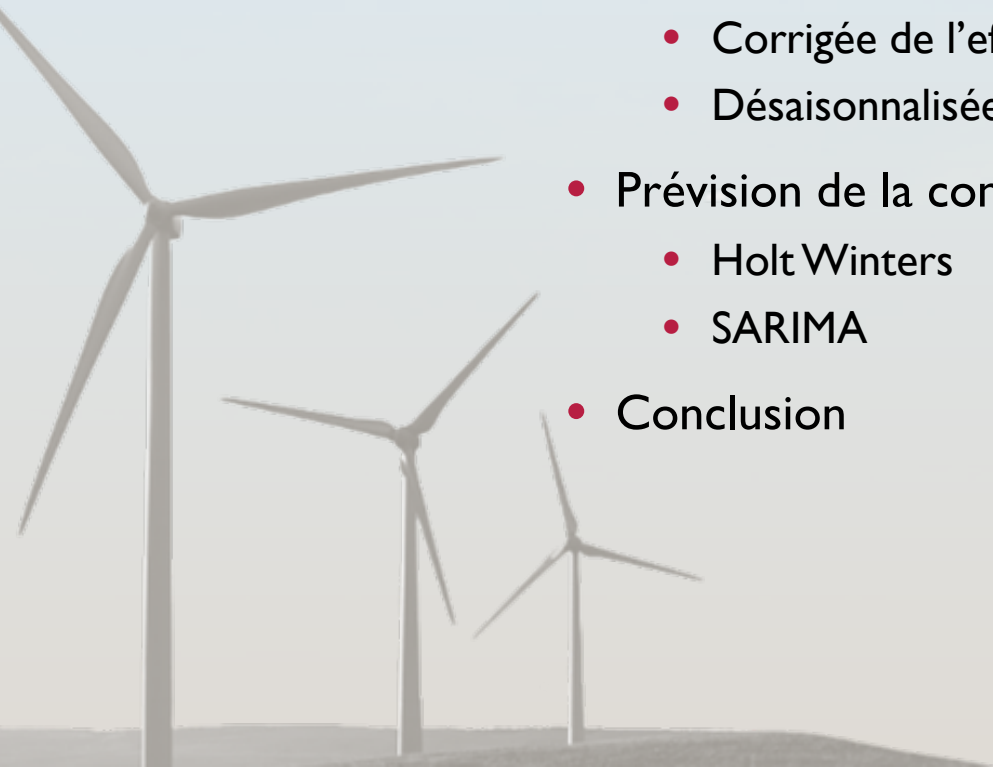
PRÉDICTION DE DEMANDE D'ÉLECTRICITÉ

CLADIERE NATHAN, PROJET 9, FORMATION DATA ANALYST OC



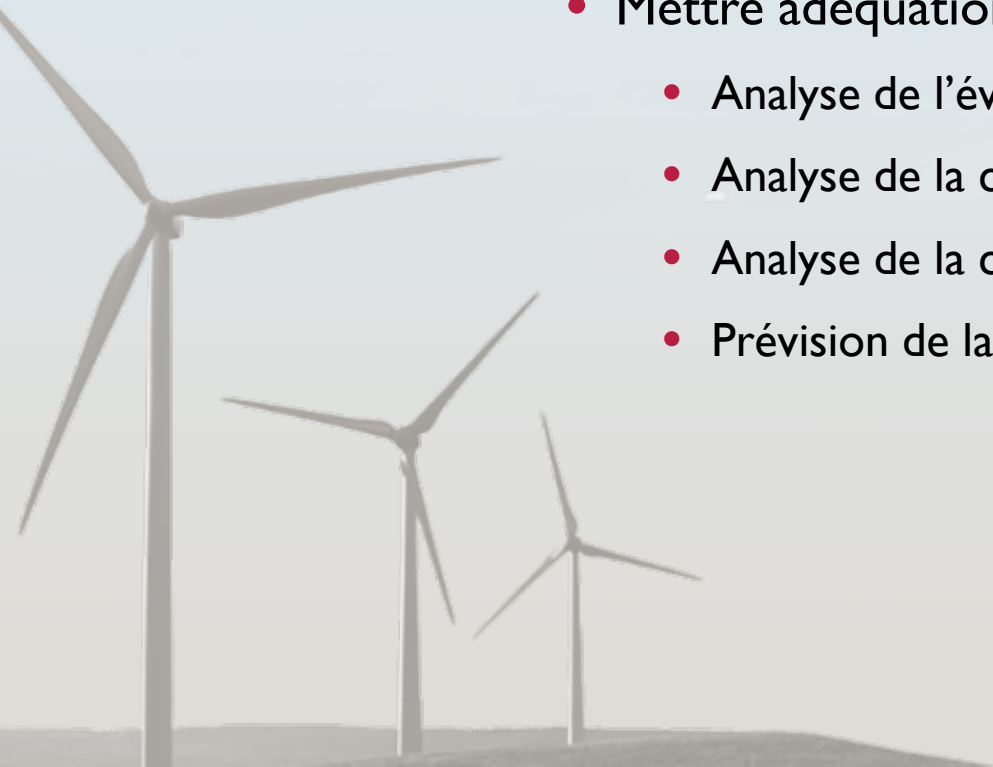
SOMMAIRE

- Contexte
- Présentation des données utilisées
- Analyse de la consommation
 - Corrigée de l'effet de température
 - Désaisonnalisée
- Prévion de la consommation
 - Holt Winters
 - SARIMA
- Conclusion



CONTEXTE

- Energies renouvelables intermittentes
- Demande en électricité varie au cours du temps
- Mettre adéquation offre et demande
 - Analyse de l'évolution de la de demande
 - Analyse de la consommation sans effet mensuel de température
 - Analyse de la consommation désaisonnalisé
 - Prévision de la consommation



PRÉSENTATION DES DONNÉES UTILISÉES



SOURCES ET PÉRIODES

DONNÉES DE CONSOMMATION

- Source : RTE (Réseau de transport d'électricité)
- Données des régions
- Groupées pour France
- Période: 2014 -2019

DONNÉES DE DJU

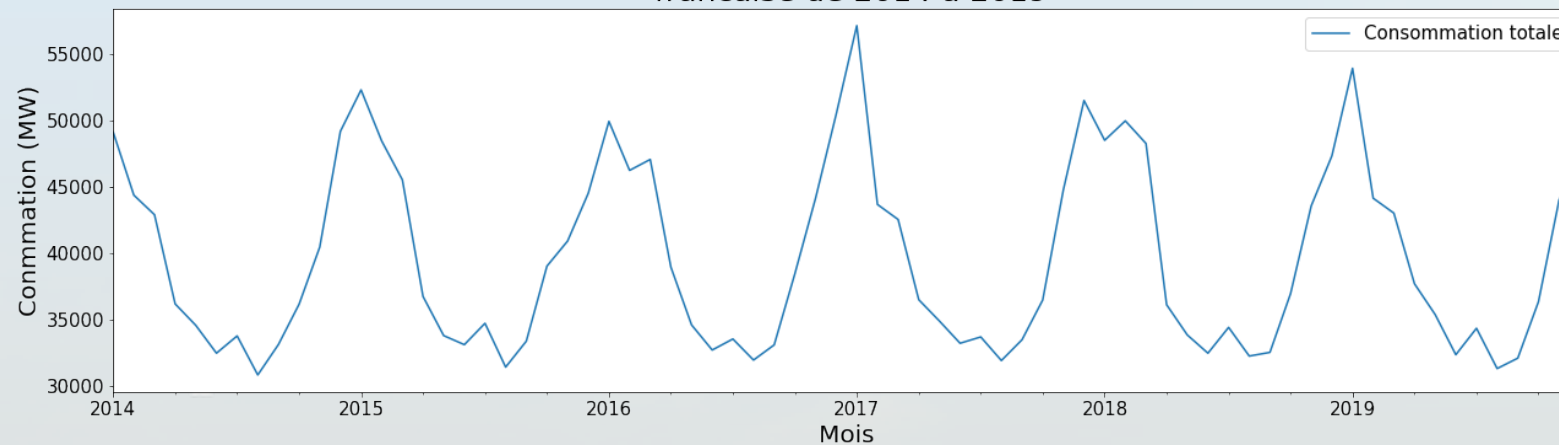
- Source : CEGIBAT (Météo France /GRDF)
- Données des régions
- Groupées pour France
- Période: 2014 -2019

Utilisation des données région:
DJU n'est pas identique dans toutes les régions



CONSOMMATION : ANALYSE GRAPHIQUE

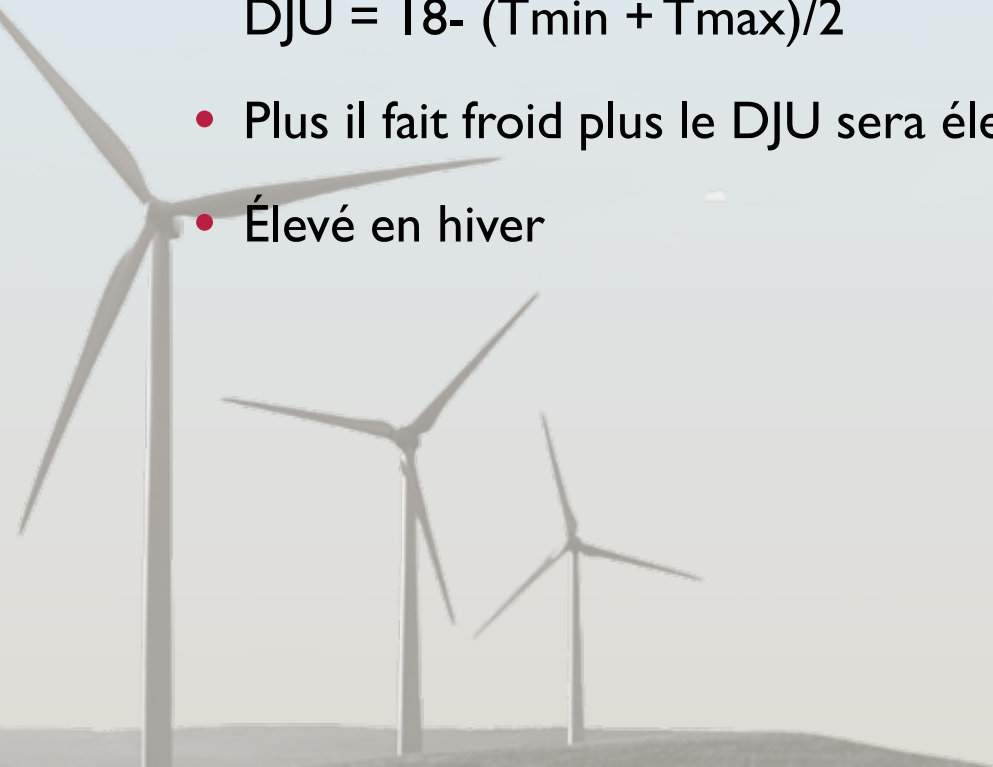
Evolution de la consommation
française de 2014 à 2019



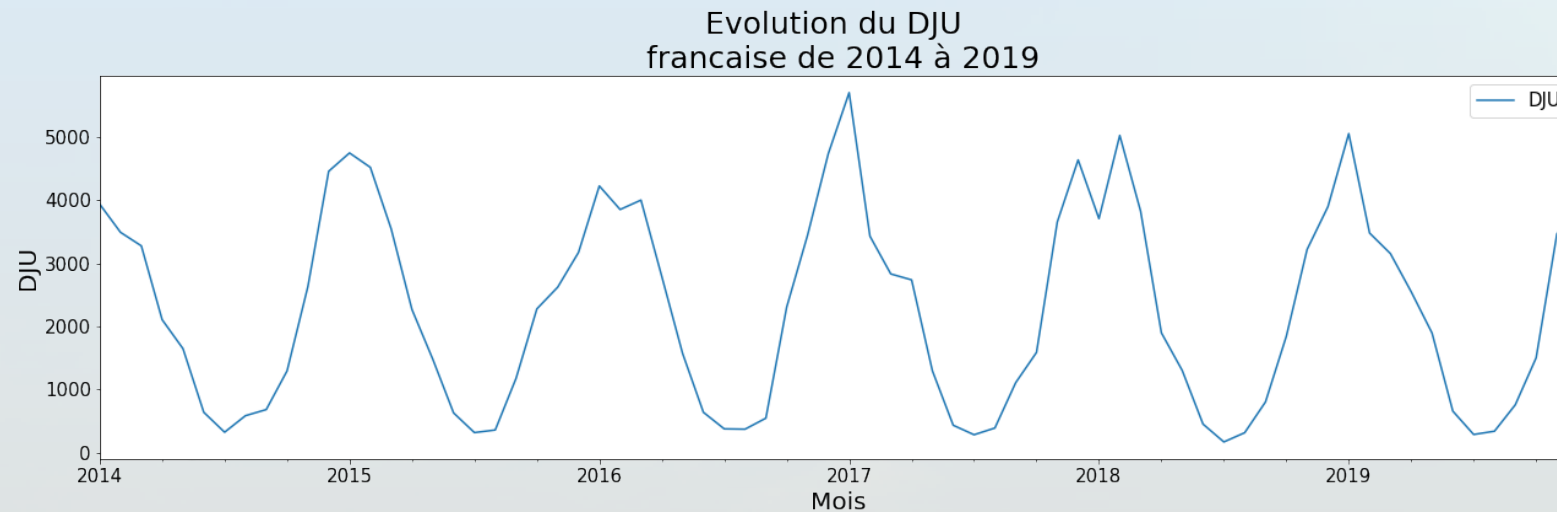
- Pic haut de consommation en hiver (moyenne Janvier: 51 800 MW)
- Pic bas de consommation en été (moyenne Aout: 31 600 MW)
- Petit pic de consommation Juillet
- Tendance stable
- Saisonnalité annuelle

DJU : QU'EST-CE ?

- Degré Jour Unifié : valeur représentant l'écart de la température d'une journée donnée et un seuil de température préétabli
- Formule :
$$DJU = 18 - (T_{min} + T_{max})/2$$
- Plus il fait froid plus le DJU sera élevé
- Élevé en hiver



ANALYSE GRAPHIQUE: DJU



- Pic haut du DJU en hiver (moyenne Janvier: 749)
- Pic bas de consommation en été (moyenne Juillet: 67)
- Pas de pic en Juillet
- Tendance stable
- Saisonnalité annuelle

PREMIÈRES ANALYSES: CONCLUSIONS

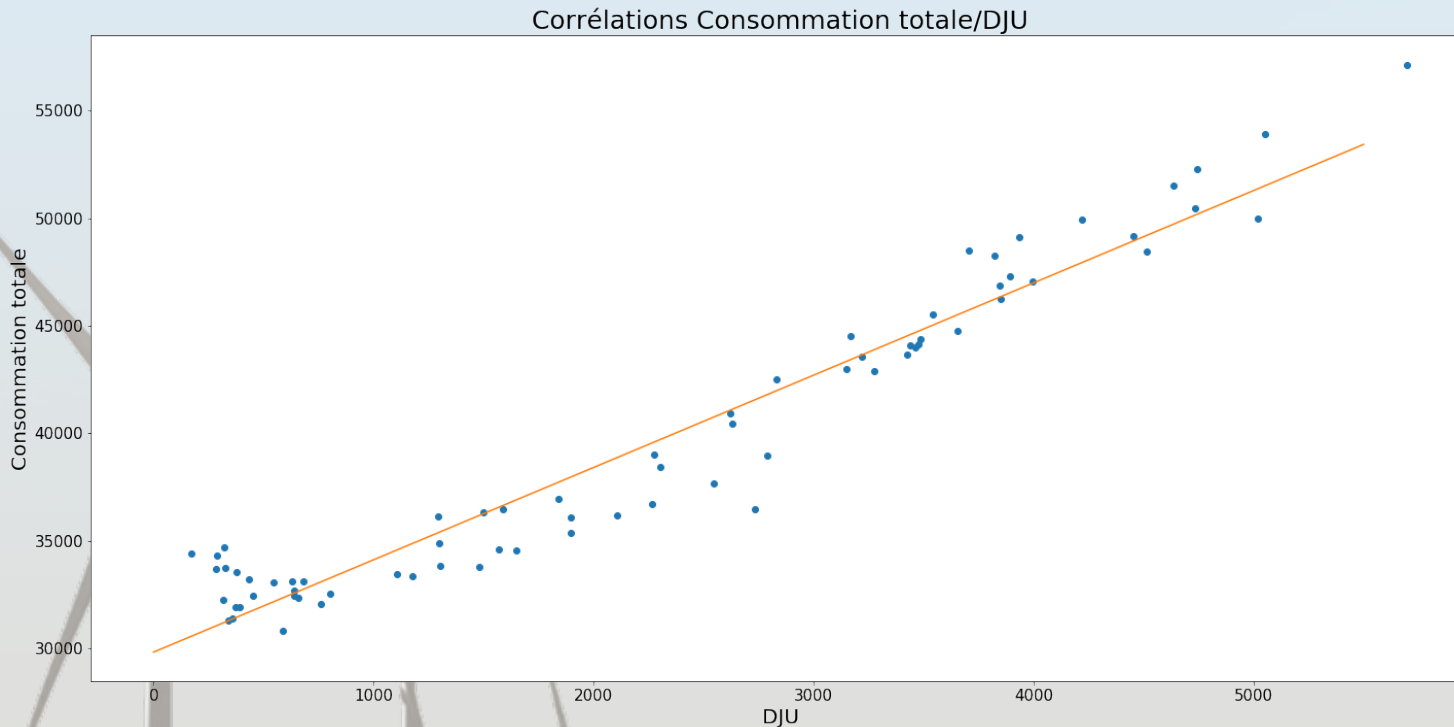
- Consommation et DJU suivent la même tendance et saisonnalité
- Y'a-t-il réellement une relation entre les deux ?
- Peut-on se servir du DJU pour voir la consommation d'électricité des français hors chauffage ?



ANALYSE DE LA CONSOMMATION

- Données corrigées de la température
- Données dessaisonnalisées

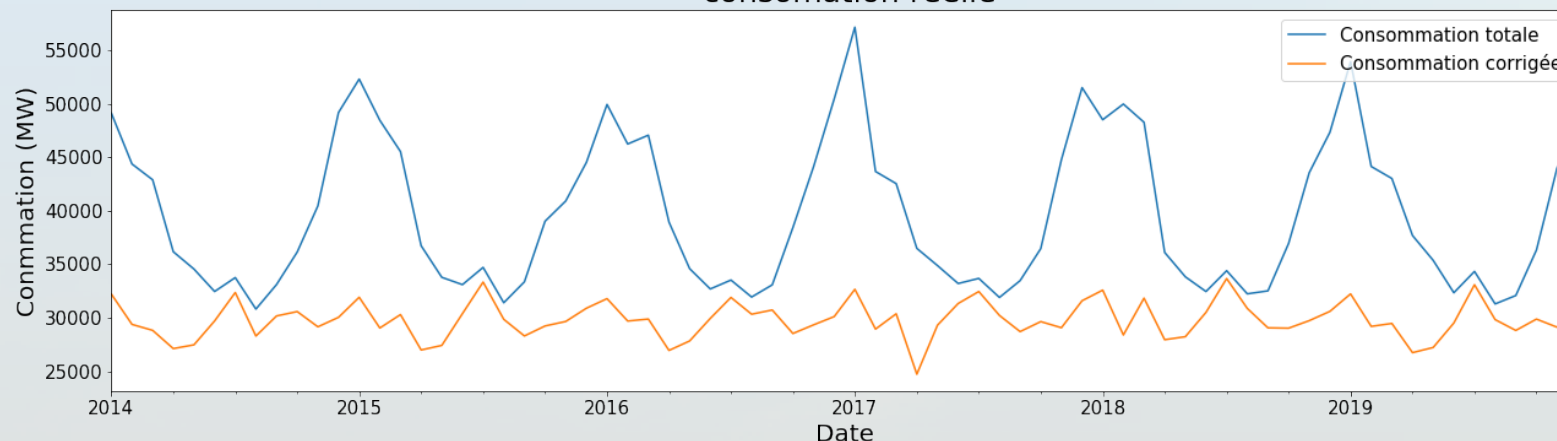
DONNÉES CORRIGÉES DE LA TEMPÉRATURE: RÉGRESSION LINÉAIRE



- Corrélation apparaît
 - Constante 30000 MW
 - Coefficient 4,29
 - Normalité des résidus
- Correction de la température avec cette régression :
$$\text{Corrigée} = \text{consommation} - (\text{DJU} * \text{coeff})$$

DONNÉES CORRIGÉES DE LA TEMPÉRATURE: ANALYSE DE LA CORRECTION

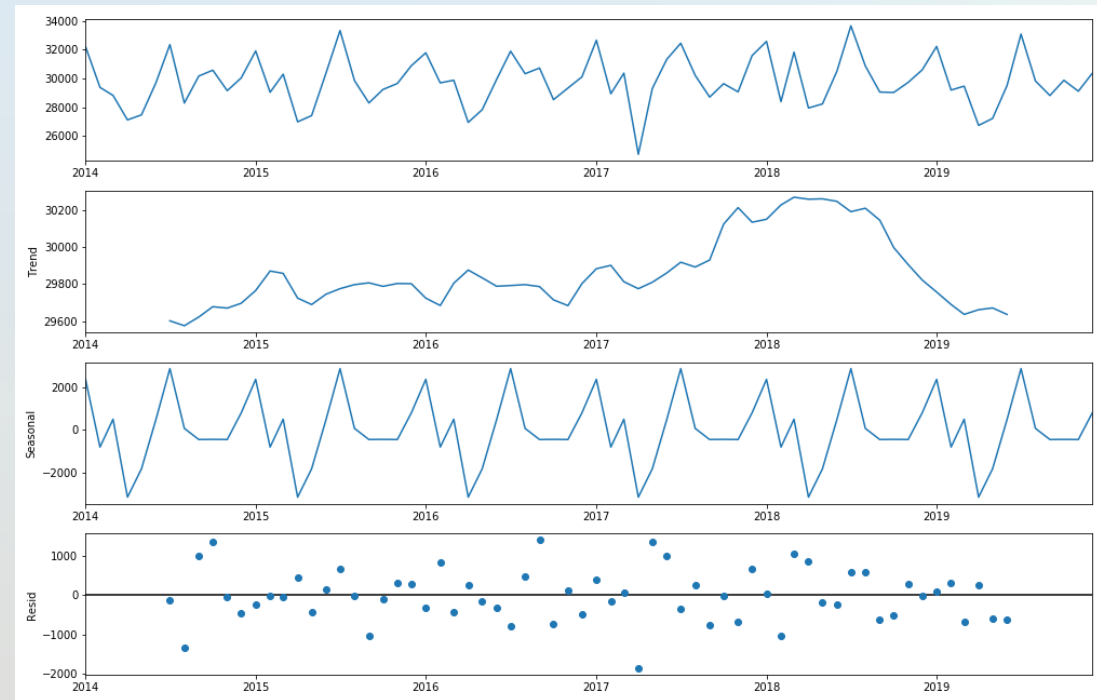
Comparaison données corrigées de températures
consommation réelle



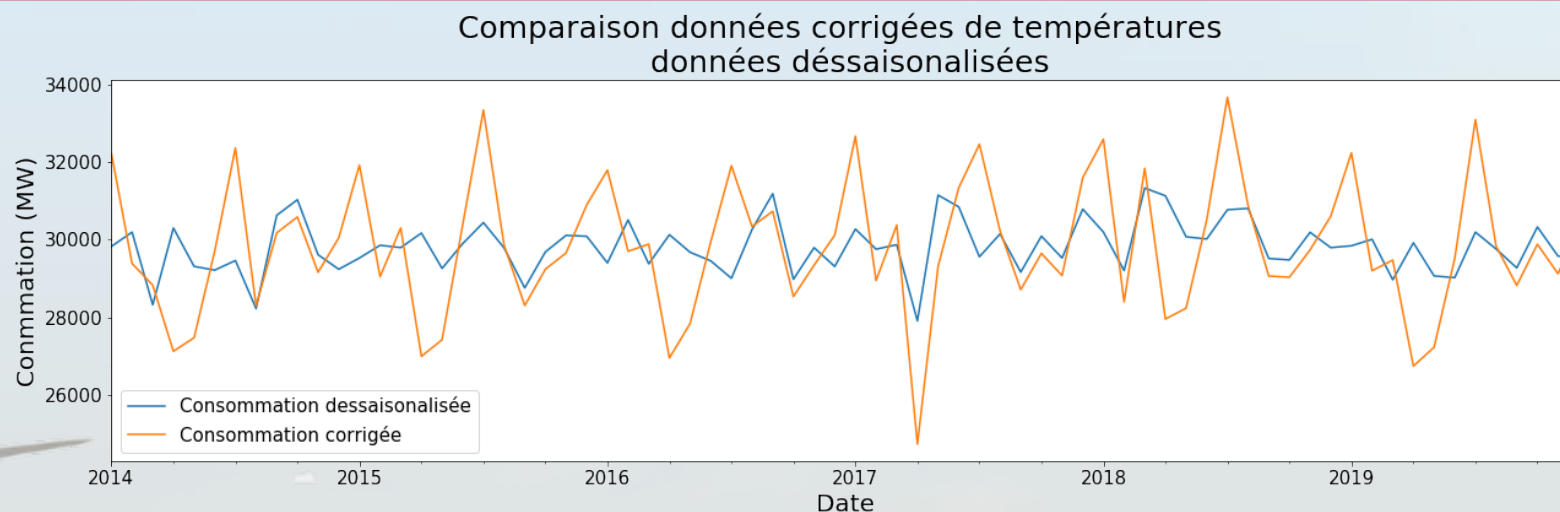
- Plus de variation due à la température
 - Variation consommation réelle: 17%
 - Variation consommation corrigée: 5%
- Saisonnalité annuelle à supprimer pour une modélisation

DONNÉES DESSAISONNALISÉES: DÉCOMPOSITION AVEC MOYENNES MOBILES

- Tendence légère hausse en 2018
- Moyenne des résidus = 0 (après dessaisonalisation)
- Saisonnalité annuelle
- Besoin d'une dessaisonalisation



DONNÉES DESSAISONNALISÉES: ANALYSE GRAPHIQUE



- Plus de variation due à la température
 - Variation données corrigées de températures: 5%
 - Variation données déssaisonnalisées: 2%

DONNÉES DESSAISONNALISÉES: VIABILITÉ DE LA SÉRIE POUR MODÉLISATION

- Série doit avoir un bruit Blanc:
 - Test de Ljung-box
- Pas d'autocorrélation des données dans le temps
 - Utilisation de l'ACF
- Distribution suivant une lois normale
 - Test de shapiro



DONNÉES DESSAISONNALISÉES: VIABILITÉ DE LA SÉRIE POUR MODÉLISATION

Série corrigée de la
température

```
6 : 0.04764412917136542
12 : 0.023822074269388562
18 : 0.01588138284625905
24 : 0.011911037134694285
30 : 0.009528829707755428
36 : 0.007940691423129524
```



Ljung-box

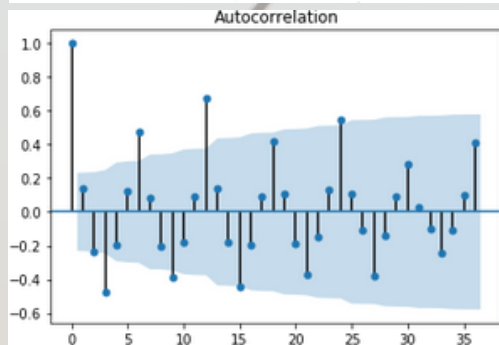
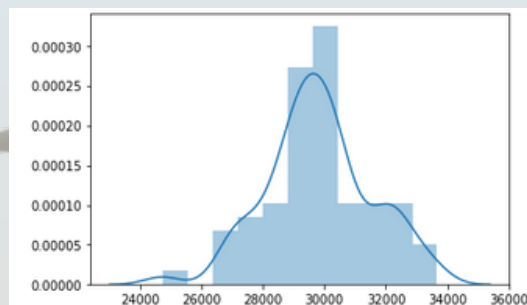


Série dessaisonnalisée

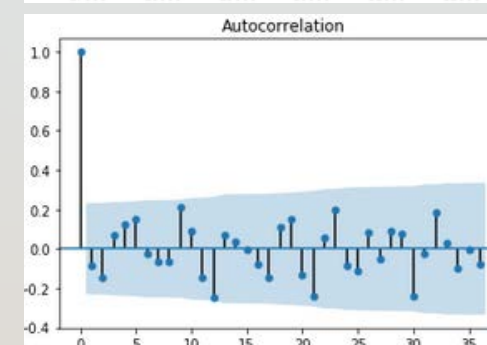
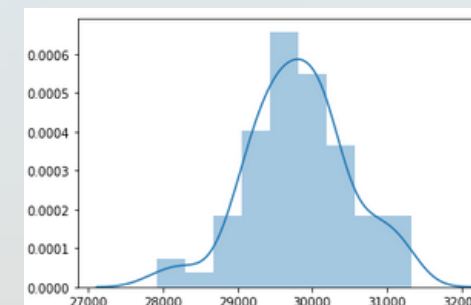
```
6 : 0.43384514324069
12 : 0.4123797487839376
18 : 0.3422576429360732
24 : 0.2765534886544823
30 : 0.22873228456138583
36 : 0.19278357089523887
```



Normalité



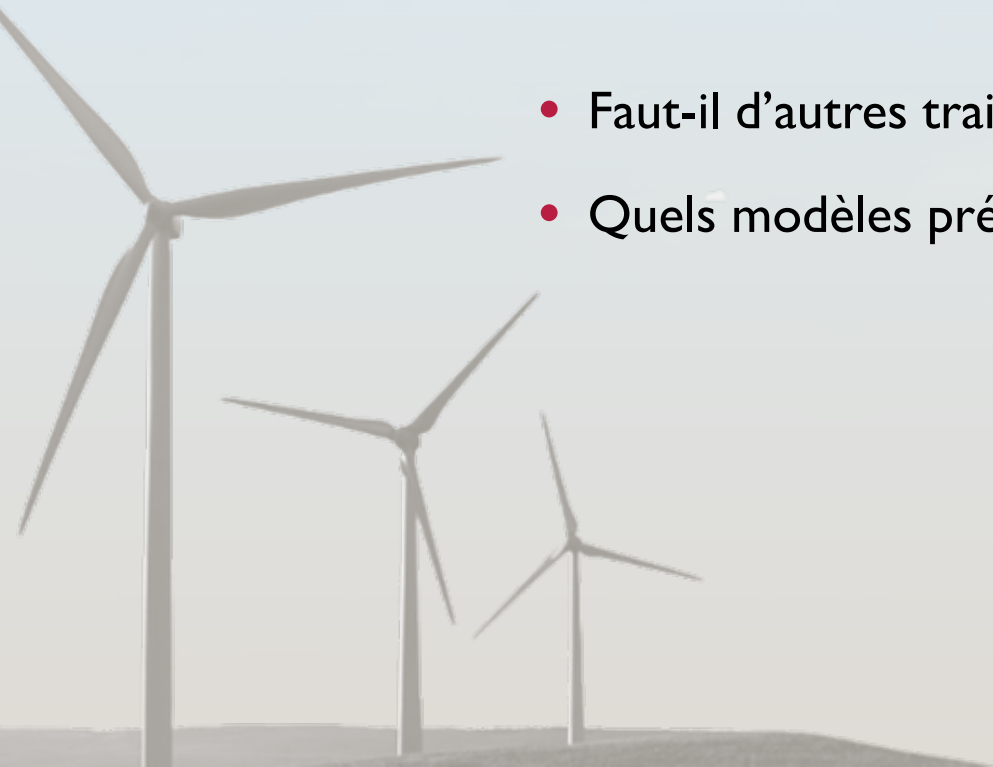
ACF



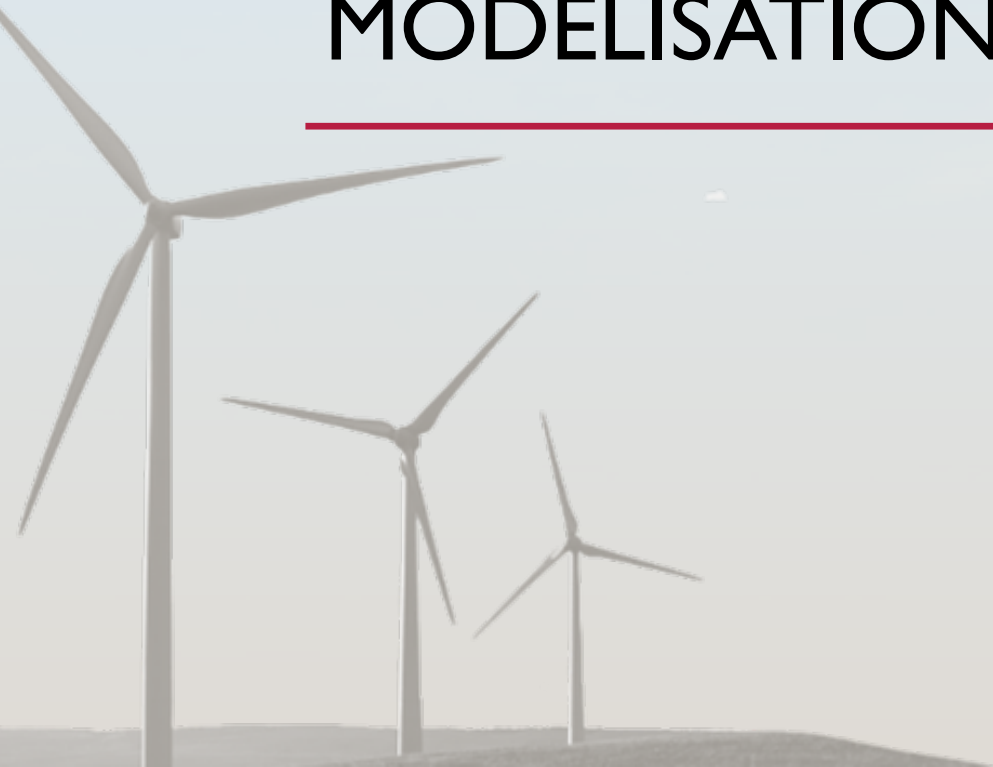
ANALYSE DE LA CONSOMMATION

CONCLUSION

- Afin de prédire la consommation pour notre production:
 - Enlever les facteurs de température
 - Dessaisonnaliser
- Faut-il d'autres traitement ?
- Quels modèles prédit le mieux la consommation ?



MODÉLISATION: HOLT WINTERS



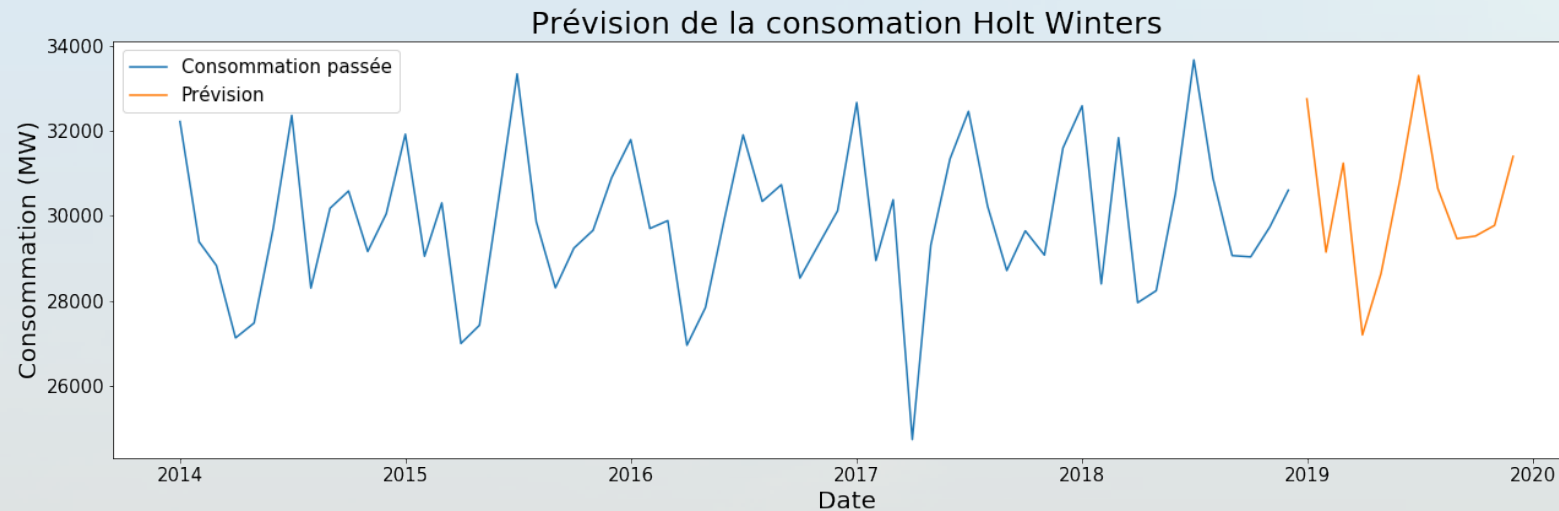
HOLT WINTERS: PARAMÉTRAGES

```
hw = ExponentialSmoothing(np.asarray(dataQ1FinalPrev["Consomation corrigée"]),  
                           seasonal_periods=12,  
                           trend='add',  
                           seasonal='add').fit()
```

- Données utilisées : données corrigées de température
- Dessaisonnalisation : seasonal_period = 12
- Tendence et saisonnalité : additif



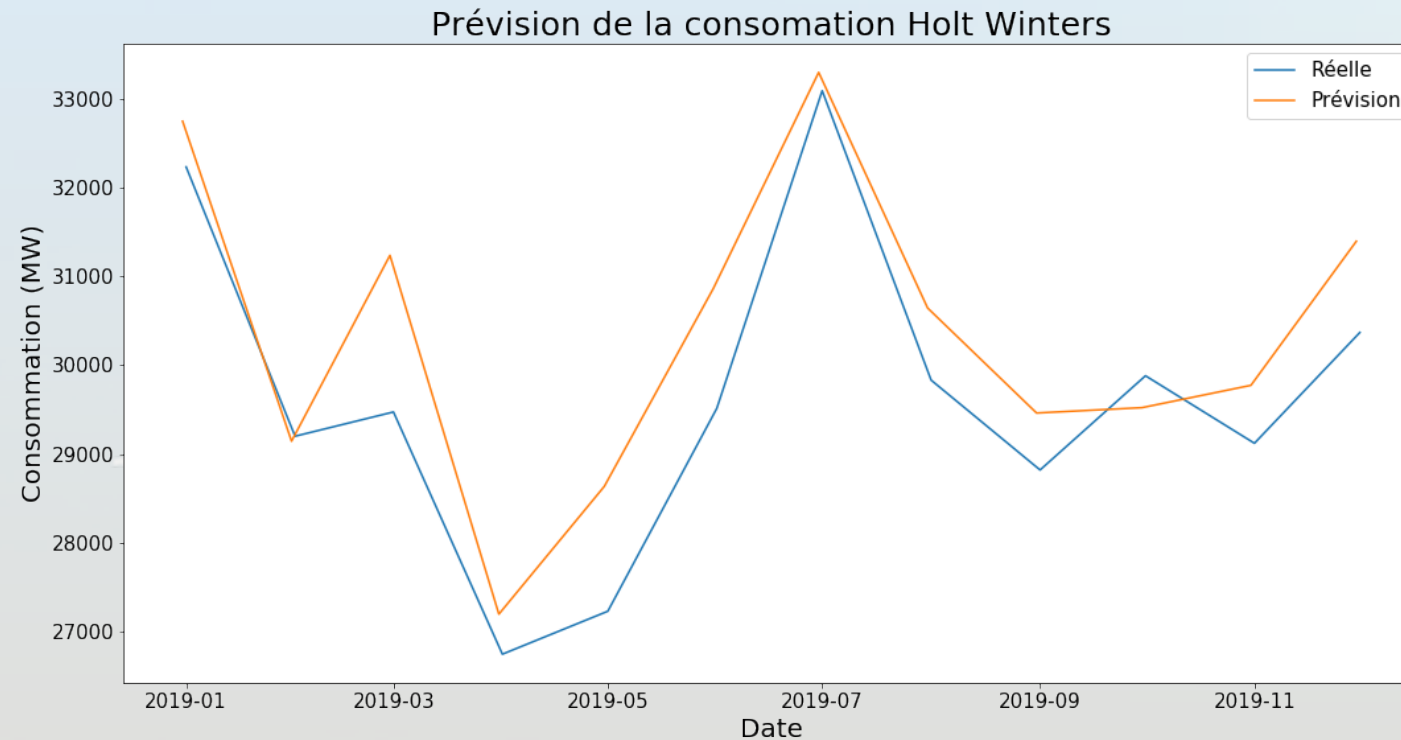
HOLT WINTERS: ANALYSE GRAPHIQUE



- Paramètres

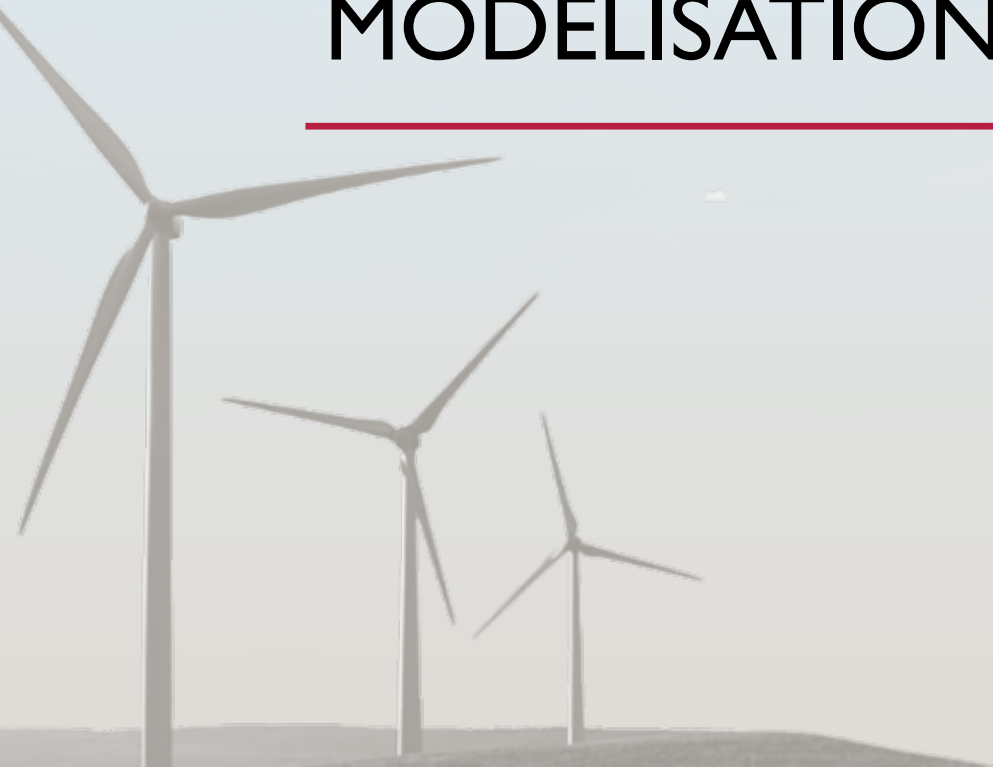
- Alpha : Ensemble des valeurs (0,052)
- Beta : Ensemble des tendances (0,052)
- Gamma : saisonnalité récente (0,47)

HOLT WINTERS: ANALYSE VIABILITÉ DU MODÈLE



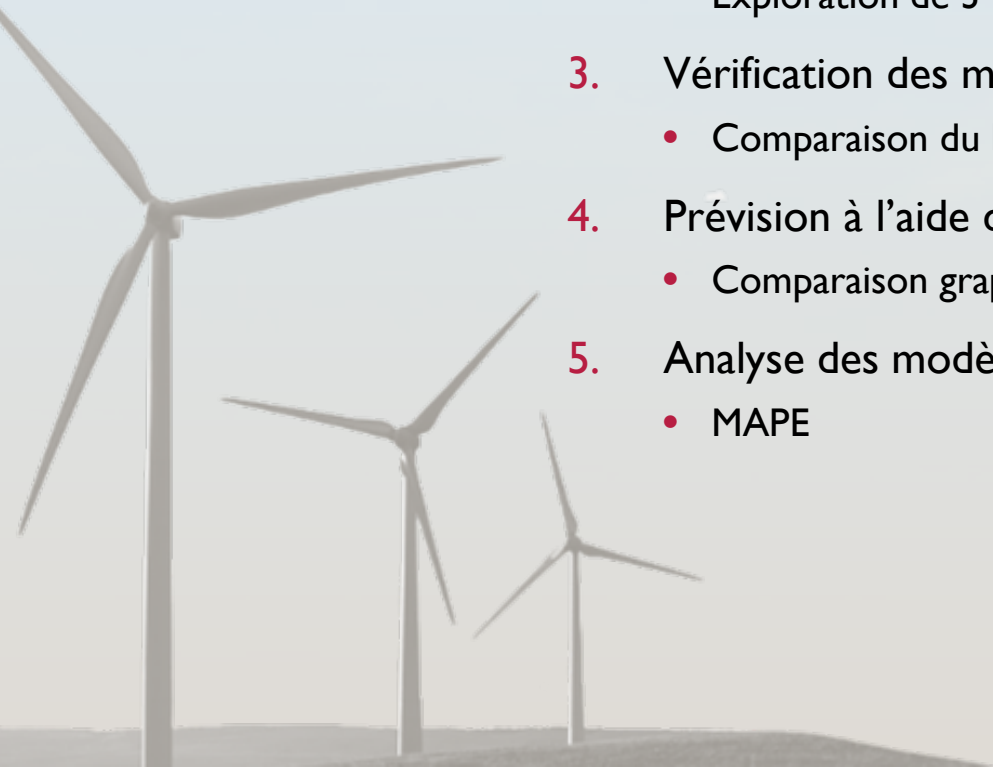
• MAPE: 2,63%

MODÉLISATION: SARIMA



SARIMA: MÉTHODOLOGIE

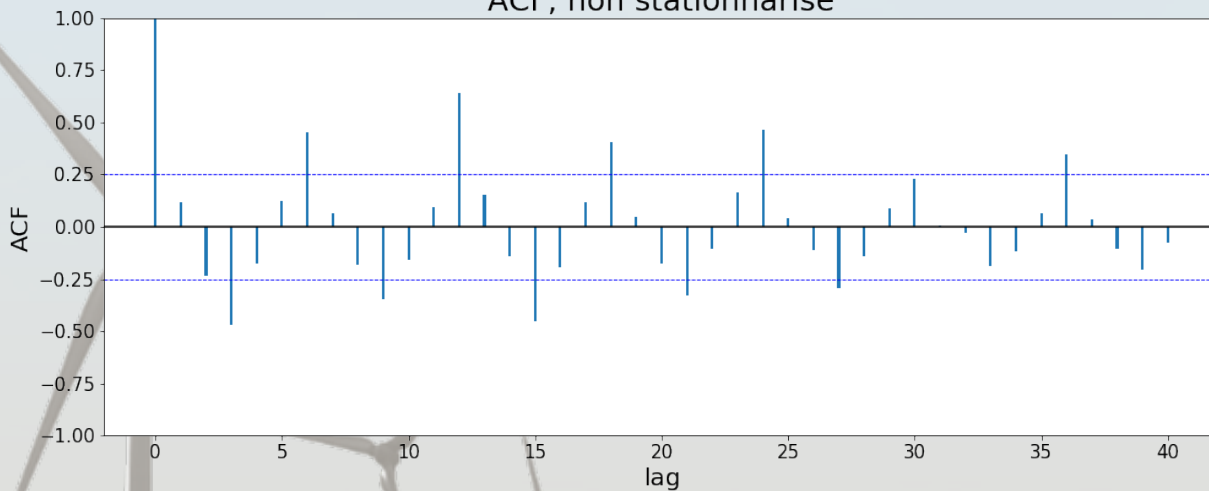
1. Stationnarisation
 - Etude des autocorrélations et test de Dickey-Fuller
2. Identifications des modèles potentiels
 - Exploration de 3 modèles
3. Vérification des modèles
 - Comparaison du bruit blanc des résidus des modèles
4. Prévion à l'aide des modèles
 - Comparaison graphique
5. Analyse des modèles
 - MAPE



SARIMA: STATIONNARISATION

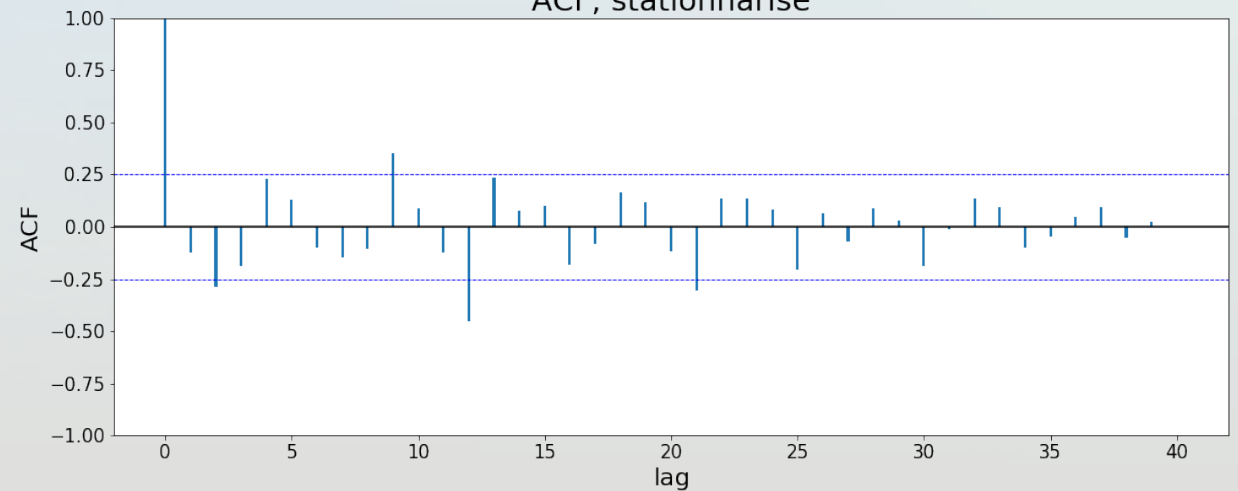
- Saisonnalité à tout les lags 12 : différenciation 12

ACF, non stationnarisé



- P-value Dickey-Fuller = 0,72

ACF, stationnarisé



- P-value Dickey-Fuller = $2,31 \times 10^{-8}$

SARIMA: IDENTIFICATIONS DES MODÈLES POTENTIELS

MANUELLE 1

- Identification des paramètres p , d et q ; P , D et Q avec l'ACF et le PCAF

MANUELLE 2

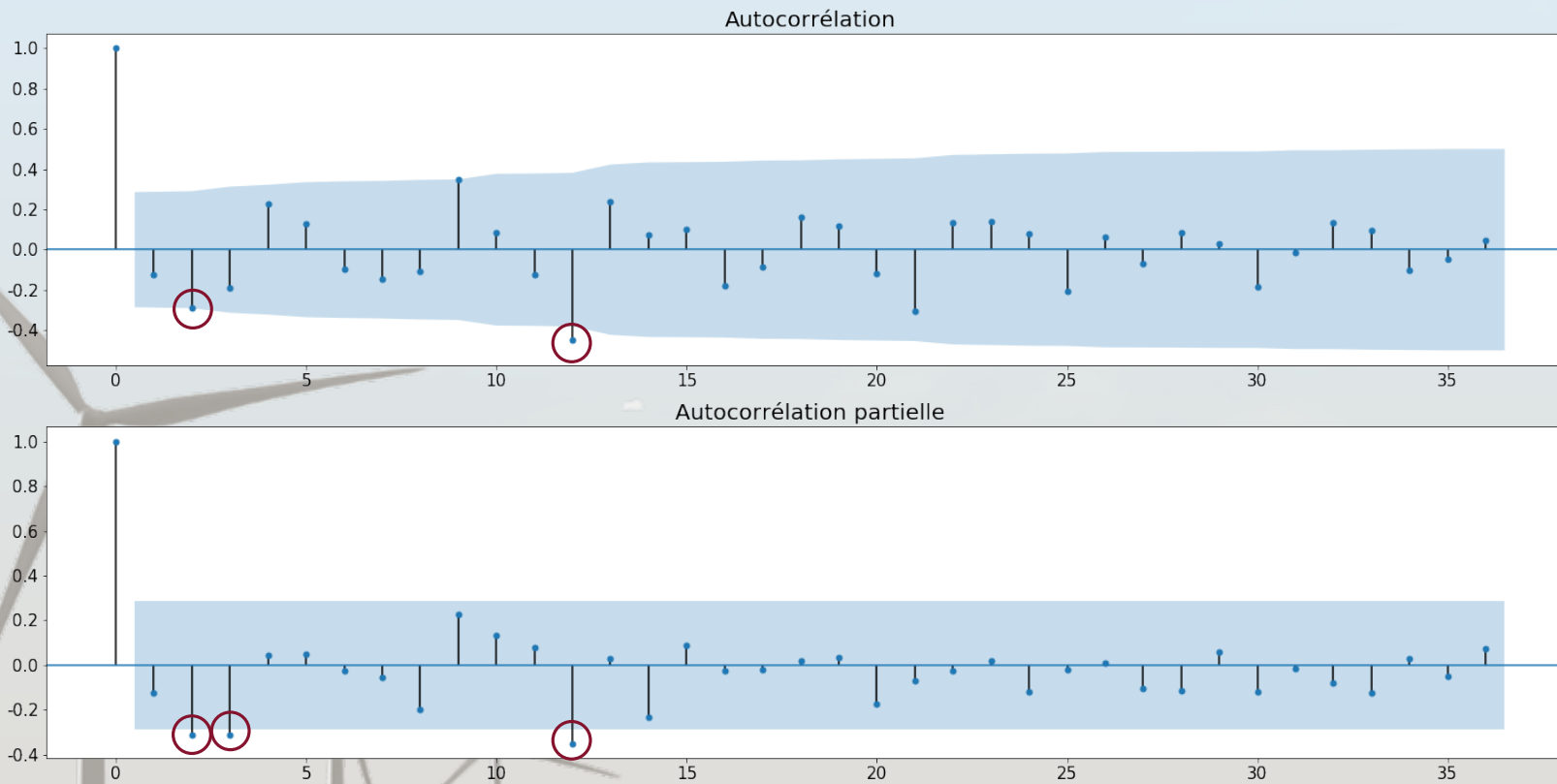
- Initialisation des paramètres $(1,1,1)(1,1,1)$, retrait si pas significatif

AUTOMATIQUE

- Minimisation des AIC et BIC



SARIMA: IDENTIFICATIONS DES MODÈLES POTENTIELS MANUELLE I



- q: nombre de lag pas significatif (ACF)
q:2
Q:0 (pas de saisonnalité)
- p: nombre de lag pas significatif (PACF)
p:3
P:0 (pas de saisonnalité)
- différenciation pour stationnariser la série
D:1
d: 0 plus de saisonnalité dans l'ACF
- (2,0,3)(0,1,0,12)

SARIMA: IDENTIFICATIONS DES MODÈLES POTENTIELS MANUELLE 2

SARIMAX Results						
Dep. Variable:	y	No. Observations:	60			
Model:	SARIMAX(1, 1, 1)x(1, 1, 1, 12)	Log Likelihood	-411.708			
Date:	Fri, 11 Dec 2020	AIC	833.416			
Time:	18:10:44	BIC	842.667			
Sample:	0	HQIC	836.897			
	- 60					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.3377	0.449	0.752	0.452	-0.543	1.219
ma.L1	-0.4817	0.443	-1.088	0.277	-1.349	0.386
ar.S.L12	0.6988	0.131	5.348	0.000	0.443	0.955
ma.S.L12	-1.0000	0.242	-4.140	0.000	-1.473	-0.527
sigma2	2.009e+06	1.2e-07	1.67e+13	0.000	2.01e+06	2.01e+06
Ljung-Box (Q):	69.54	Jarque-Bera (JB):	0.40			
Prob(Q):	0.00	Prob(JB):	0.82			
Heteroskedasticity (H):	1.42	Skew:	-0.08			
Prob(H) (two-sided):	0.49	Kurtosis:	2.58			

SARIMAX Results

Dep. Variable: y

No. Observations: 60

Model: SARIMAX(0, 1, 1)x(1, 1, 1, 12)

Log Likelihood -412.243

Date: Fri, 11 Dec 2020

AIC 832.486

Time: 18:11:14

BIC 839.887

Sample: 0

HQIC 835.271

- 60

Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.1397	0.052	-2.682	0.007	-0.242	-0.038
ar.S.L12	0.6950	0.129	5.375	0.000	0.442	0.948
ma.S.L12	-1.0000	0.245	-4.085	0.000	-1.480	-0.520
sigma2	2.072e+06	1.18e-07	1.75e+13	0.000	2.07e+06	2.07e+06

Ljung-Box (Q): 72.22

Jarque-Bera (JB): 0.53

Prob(Q): 0.00

Prob(JB): 0.77

Heteroskedasticity (H): 1.42

Skew: -0.10

Prob(H) (two-sided): 0.49

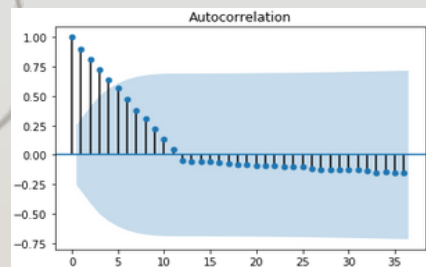
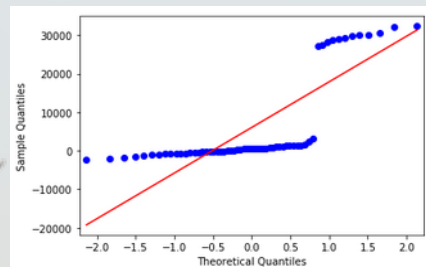
Kurtosis: 2.52

- Retrait du modèle AR non significatif

SARIMA: VÉRIFICATION DU MODÈLE BLANCHEUR DES RÉSIDUS

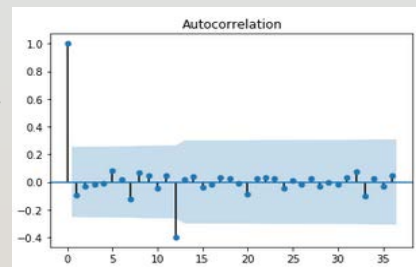
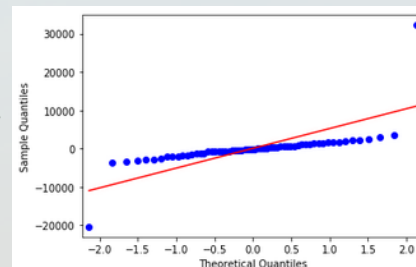
MANUELLE 1

6 : 2.0112978346934906e-13
12 : 1.0056489173467453e-13
18 : 6.704326115644968e-14
24 : 5.0282445867337264e-14
30 : 4.0225956693869814e-14
36 : 3.352163057822484e-14



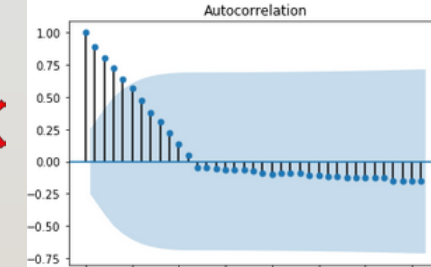
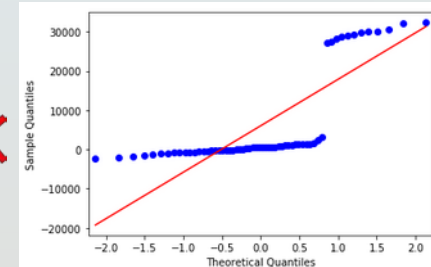
MANUELLE 2

6 : 0.9834945659586819
12 : 0.9255676217764474
18 : 0.7782687087104672
24 : 0.780777465234489
30 : 0.8137683054600248
36 : 0.8422705276321688



AUTOMATIQUE

6 : 2.4560178289469295e-13
12 : 1.2280089144734647e-13
18 : 8.186726096489765e-14
24 : 6.140044572367324e-14
30 : 4.912035657893859e-14
36 : 4.0933630482448825e-14



Ljung-box

✗

Normalité

✗

ACF

✗

✓

✓

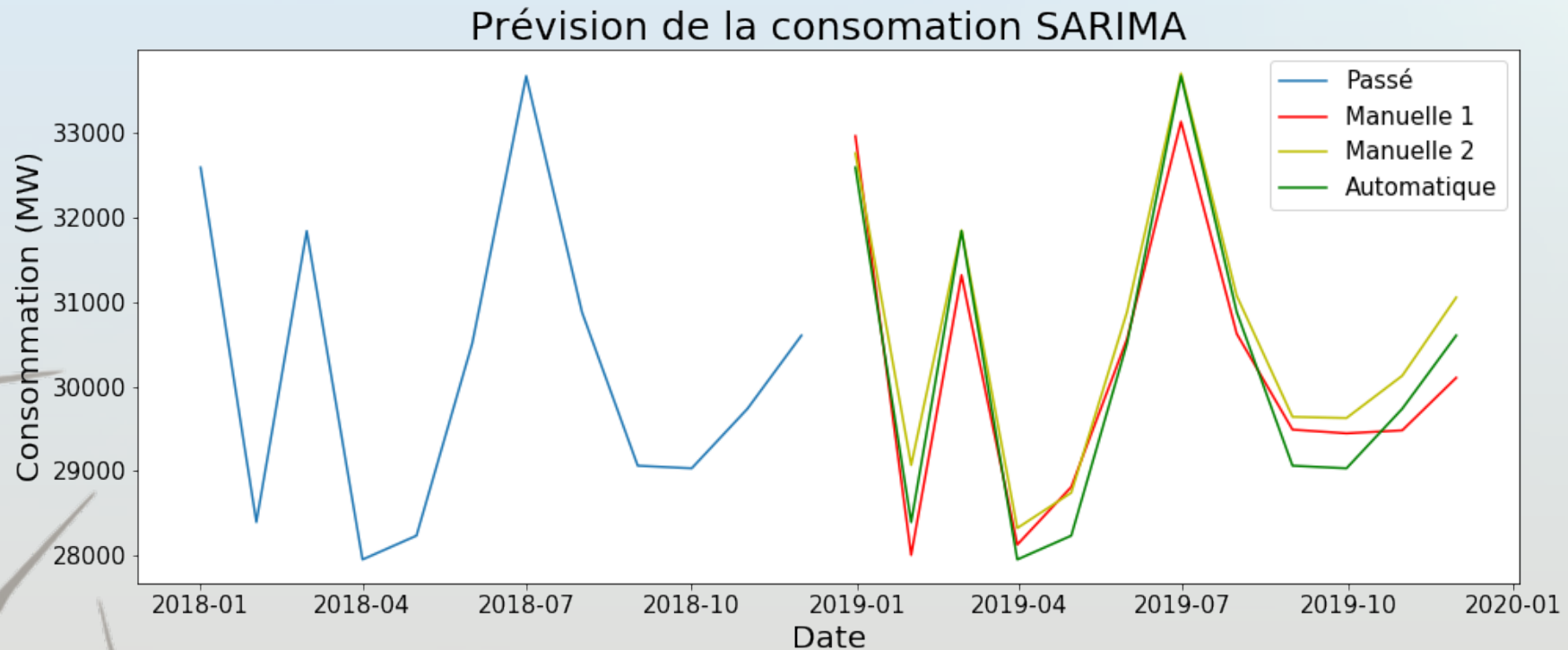
✓

✗

✗

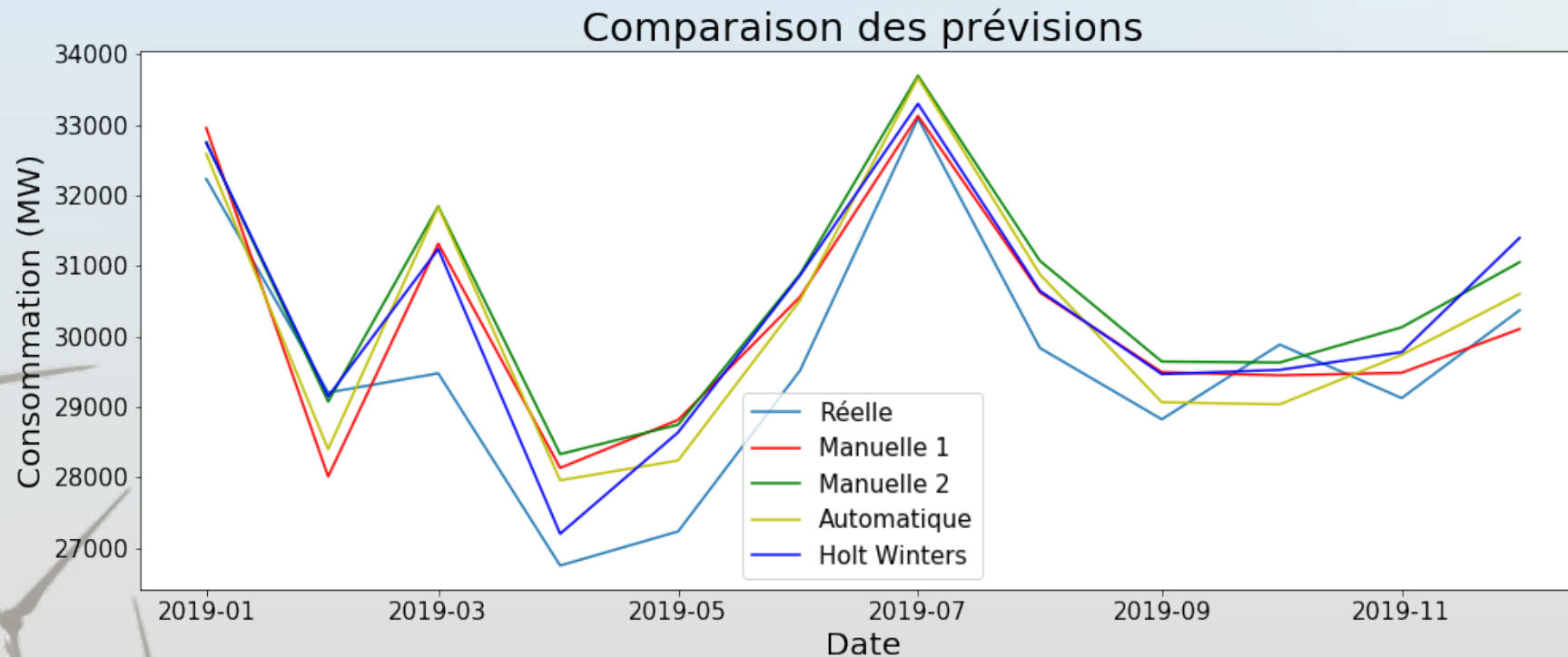
✗

SARIMA: PRÉVISION À L'AIDE DES MODÈLES



- Sensiblement identique
- Modèle automatique accentue les pics

SARIMA: ANALYSE DES MODÈLES



MAPE:

M1: 2,97%

M2: 3,46%

Auto: 2,93%

HW: 2,63%

MODÉLISATION: CONCLUSION

- Choix du modèle dépendra de nos critères de décisions
- Meilleur MAPE : Holt Winters



CONCLUSION



CONCLUSION GLOBALE

- Attention aux grosses variations été/hiver
- Différent modèles de prédictions possibles
- A mettre en adéquation avec notre capacité de production
- Etudier l'effet de la climatisation

