

# Dossier Techniques de Prévision et Conjoncture

## MASTER 1 ECAP

<p>Prévision et analyse de la consommation de gaz en France : une étude saisonnnière sur les ménages et l'industrie (2008–2020)</p>
---

Dossier réalisé par :

**Achille MARTERET**

**Timothée CADET**

Encadré par :

Monsieur Olivier DARNE

**Année universitaire 2024-2025**

## SOMMAIRE

<b>Introduction.....</b>	<b>3</b>
<b>I) Analyse exploratoire .....</b>	<b>4</b>
<b>II) Pr�vision de la s�rie saisonni�re corrig�e des points atypiques sur une ann�e 13</b>	
<b>III) Repr�sentation graphique de l'�volution des pr�visions des diff�rents mod�les .....</b>	<b>18</b>
<b>IV) Qualit� de pr�vision .....</b>	<b>20</b>
<b>V) Test de pr�cision .....</b>	<b>22</b>
<b>VI) Pr�vision sur une ann�e avec un pas de 1 mois avec le meilleur mod�le ...</b>	<b>23</b>
<b>Conclusion .....</b>	<b>25</b>
<b>Annexes .....</b>	<b>26</b>
<b>Bibliographies.....</b>	<b>40</b>
<b>Table des mati�res.....</b>	<b>41</b>

# Introduction

La consommation de gaz naturel en France s'inscrit au cœur des enjeux énergétiques, économiques, environnementaux et sociaux actuels. Le gaz est une ressource fossile longtemps centrale dans le mix énergétique français, le gaz naturel est utilisé principalement pour le chauffage, la production d'électricité, et les usages industriels. Malgré une tendance à la baisse ces dernières années, notamment en raison des politiques de transition énergétique, le gaz reste une composante essentielle du paysage énergétique national.<sup>1</sup>

Ce dossier s'intéresse à l'analyse de la consommation mensuelle de gaz en France entre janvier 2008 et janvier 2020, une période de plus de dix ans permettant d'étudier les dynamiques saisonnières, les tendances et les événements atypiques qui ont pu impacter cette série.

Cette période est également marquée par des événements significatifs comme la crise économique de 2008-2009, la montée en puissance des politiques de réduction des émissions de gaz à effet de serre, ainsi qu'une sensibilisation croissante aux enjeux énergétique. Dans un contexte européen de diversification des approvisionnements et de diminution de la dépendance énergétique, analyser l'évolution de la consommation de gaz en France permet de mieux comprendre les évolutions en cours.

L'analyse de ce dossier se déroulera en 3 étapes distincts. Dans un premier temps, une analyse exploratoire sera menée afin d'observer les principales dynamiques de la série temporelle de consommation de gaz. Cette phase permettra de visualiser les variations dans le temps, de détecter les pics de consommation, ainsi que d'identifier d'éventuels points atypiques et d'en proposer des explications. Nous procéderons également à une analyse de la saisonnalité afin d'évaluer la présence de cycles saisonniers. Dans un deuxième temps, différentes méthodes de prévision seront mises en œuvre afin de comparer leurs performances. L'objectif sera d'identifier le modèle offrant la meilleure qualité de prévision. Enfin, dans un troisième temps, nous utiliserons ce modèle optimal pour réaliser une prévision sur un horizon d'un an, avec une fréquence mensuelle.

---

<sup>1</sup> [Don't Make The same Mistake As Me! Almost-Returned\\_2 \(FR\)](#)

## **I) Analyse exploratoire**

### **a) Présentation et caractérisation de la série étudiée**

La base de données utilisée dans ce dossier provient d'Eurostat et est référencée sous le nom « nrg\_cb\_gasm », qui regroupe des données mensuelles sur l'approvisionnement, la transformation et la consommation de gaz naturel au sein des pays de l'Union européenne.<sup>2</sup>

La série temporelle analysée ici correspond à la consommation mensuelle de gaz naturel en France, telle que définie par la classification internationale des produits énergétiques (SIEC). Les données sont exprimées en térajoules (TJ) selon le pouvoir calorifique brut (Gross Calorific Value - GCV), qui représente la totalité de l'énergie libérée par la combustion.<sup>3</sup>

À titre d'exemple, la consommation française de gaz naturel s'élevait à 247 996 TJ en janvier 2008, ce qui équivaldrait à l'énergie nécessaire pour alimenter environ 59 millions de foyers pendant un mois. Or, la France ne comptait à cette époque qu'environ 30 millions de foyers. Cette différence s'explique par le fait que la base de données d'Eurostat "nrg\_cb\_gasm" ne se limite pas aux seuls ménages : selon la documentation officielle d'Eurostat, la consommation finale de gaz comprend aussi les usages industriels, tertiaires (bureaux, commerces, services), ainsi que, dans une moindre mesure, les transports et la production d'électricité. Cela illustre l'ampleur de la demande énergétique durant les mois d'hiver, période traditionnellement marquée par une hausse des besoins en chauffage.

Cette série constitue une base pertinente pour étudier les dynamiques de la demande énergétique en France, compte tenu de sa mesure mensuelle, de sa période d'observation étendue (2008-2020) et de son importance dans le contexte des enjeux énergétiques contemporains.<sup>4</sup>

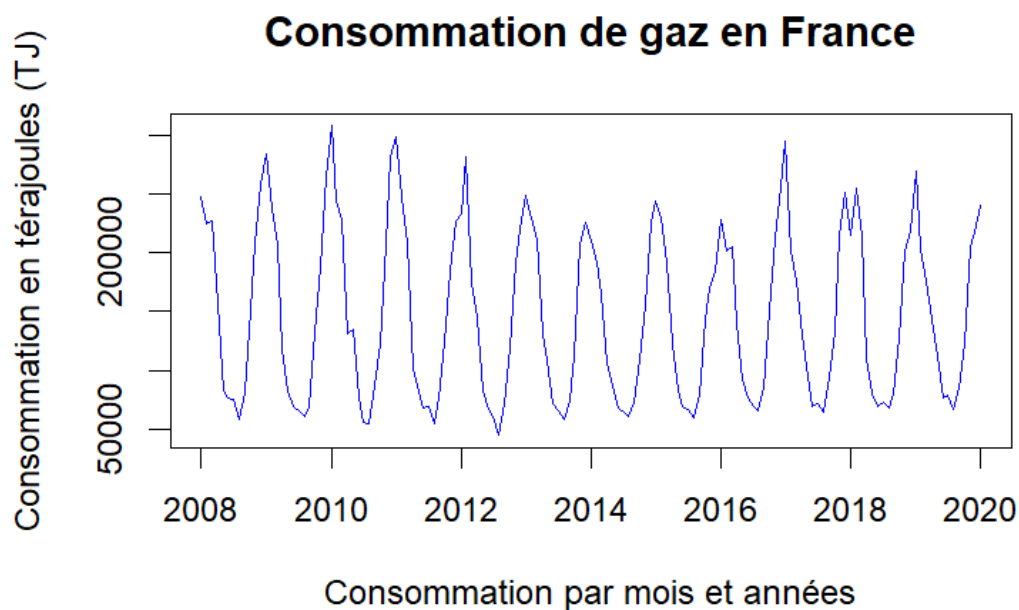
---

<sup>2</sup> [\[nrg\\_cb\\_gasm\] Supply, transformation and consumption of gas - monthly data](#)

<sup>3</sup> [Energy statistics - quantities \(nrg\\_quant\)](#)

<sup>4</sup> [Base de données - Eurostat](#)

Figure 1 : Série temporelle de la consommation de gaz en France en térajoules



La figure 1 met en évidence une forte saisonnalité dans la consommation mensuelle de gaz en France entre 2008 et 2020. Chaque année, on observe des pics de consommation réguliers durant les mois d'hiver, principalement liés à l'augmentation de la demande de chauffage, tandis que les creux se situent en été, période où les besoins énergétiques sont plus faibles.

On remarque également que certaines années présentent des variations plus marquées que d'autres, ce qui pourrait s'expliquer par des facteurs conjoncturels, tels que des conditions météorologiques inhabituelles (hivers particulièrement froids ou doux), des évolutions dans les politiques énergétiques ou encore des fluctuations économiques ayant un impact sur la consommation industrielle.

Enfin, une tendance générale à la baisse semble se dessiner à partir de 2012, ce qui pourrait refléter une évolution du mix énergétique français. Cette période marque en effet une montée en puissance progressive des énergies renouvelables, notamment le solaire et l'éolien, dans le cadre de la transition énergétique.

## b) Détection des points atypiques

L'analyse des points atypiques réalisée avec le package tso met en évidence trois valeurs extrêmes dans la série mensuelle de consommation de gaz en France entre 2008 et 2020 (voir annexe 1).

On identifie deux points atypiques de type AO (Additive Outlier), correspondant à des chocs ponctuels, ainsi qu'un point de type TC (Temporary Change), représentant une variation transitoire qui s'est estompée progressivement. Ces trois points atypiques sont résumés dans le tableau ci-dessous :

Tableau 1 : Détection des points atypiques

	Type	n°Observation	Time	Coefhat	tstat
1	AO	50	2012:02	83585	5.462
2	TC	73	2014:01	-50402	-3.663
3	AO	121	2018:01	-63238	-4.402

Le premier point atypique apparaît en février 2012. Il s'agit d'un Additive Outlier positif, signalant une hausse exceptionnelle de la consommation estimée à plus de 83 000 unités. Ce pic est très probablement lié à la vague de froid historique qui a touché la France ce mois-là, avec des températures largement inférieures aux normales saisonnières pendant plusieurs semaines sur l'ensemble du territoire. Cette situation a entraîné une forte augmentation de la demande en gaz, notamment pour le chauffage.<sup>56</sup>

Le deuxième point atypique concerne le mois de janvier 2014. Il s'agit ici d'un Temporary Change, c'est-à-dire un changement temporaire de niveau avec un effet initial estimé à une baisse de 50 000 unités, qui s'atténue au fil des mois suivants. Ce phénomène peut être attribué à un hiver exceptionnellement doux, comme l'indiquent les données climatiques de Météo France : l'hiver 2013-2014 figure parmi les plus doux depuis 1900. La douceur des températures a logiquement réduit les besoins de chauffage, entraînant une baisse temporaire de la consommation de gaz.<sup>78</sup>

Le troisième point atypique est observé en janvier 2018, sous la forme d'un Additive Outlier négatif, avec une chute ponctuelle de la consommation de plus de 62 000 unités. Plusieurs explications sont envisageables pour cet épisode : il pourrait s'agir de conditions météorologiques particulièrement clémentes, d'événements économiques comme des grèves dans le secteur de l'énergie ou des perturbations dans

---

<sup>5</sup> [Froid : la consommation de gaz atteint un nouveau sommet historique - ladepeche.fr](#)

<sup>6</sup> [Hausse de la consommation des ménages en février : retour à la confiance ou conséquence de la vague de froid ? - FioulReduc](#)

<sup>7</sup> [L'hiver 2013-2014 parmi les plus doux depuis 1900](#)

<sup>8</sup> [Gaz : la France a beaucoup moins consommé en 2014](#)

l'approvisionnement, ou encore de facteurs géopolitiques ou techniques, tels que des opérations de maintenance majeures ou des restrictions temporaires d'importation.<sup>9</sup>

Enfin, les statistiques t associées à ces trois points montrent qu'ils sont tous significatifs au seuil de 1 %, confirmant ainsi leur caractère réellement atypique du point de vue statistique et justifiant une interprétation économique approfondie.

### c) Statistiques descriptives sur la série corrigée

Tableau 2 : Statistiques descriptives de la série corrigée

	Min	1Q	Médiane	Moyenne	3Q	Max	Sd	Skewness	Kurtosis
Série	44 833	76 423	130 988	147 022	212 366	307 937	371.2	0,35	-1,29

La moyenne de la série corrigée (147 022) est supérieure à la médiane (130 988), ce qui suggère une asymétrie positive de la distribution. Autrement dit, la série présente une légère tendance à être étirée vers la droite, avec quelques mois de consommation particulièrement élevée qui tirent la moyenne vers le haut. Cette interprétation est cohérente avec la valeur positive de la skewness (0,35), qui confirme cette asymétrie en signalant une queue droite plus longue dans la distribution des données.

La série présente également une grande dispersion, avec une consommation mensuelle minimale de 44 833 TJ et un maximum atteignant 307 937 TJ. Cette large étendue traduit des variations importantes entre les mois les plus creux (en été) et les mois de forte demande (en hiver). En complément, la moyenne des racines carrées, qui donne une idée de la variabilité de la série transformée, est estimée à 371.2, ce qui illustre un écart type moyen relativement élevé.

Concernant la variabilité intra-annuelle, l'analyse des racines carrées de la série révèle des pics récurrents en décembre et janvier, correspondant à des mois où la demande en chauffage est particulièrement élevée. On note cependant une variabilité d'une année à l'autre : par exemple, la consommation en janvier 2010 (valeur transformée : 554,9) est nettement supérieure à celle de janvier 2012 (483,1), ce qui illustre l'influence possible des conditions climatiques ou d'autres facteurs contextuels sur la demande en gaz.

Enfin, la kurtosis négative (-1,29) indique que la distribution est plus aplatie que la normale, avec des valeurs moins concentrées autour de la moyenne et des extrêmes moins fréquents qu'une distribution normale.

---

<sup>9</sup> [Bilan météo de janvier 2018](#)

Figure 2 : Histogramme et densité de la série temporelle corrigé

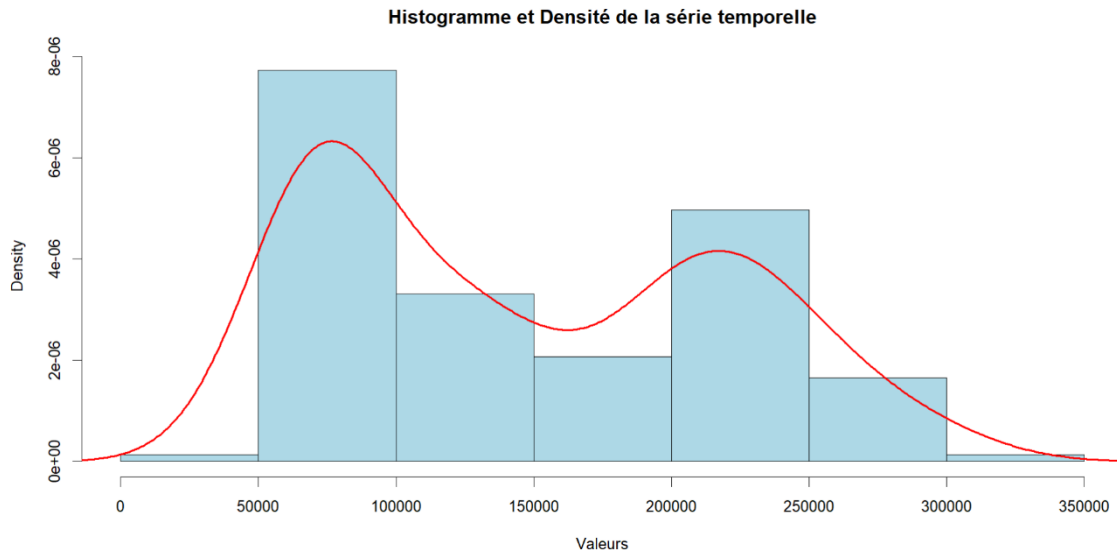
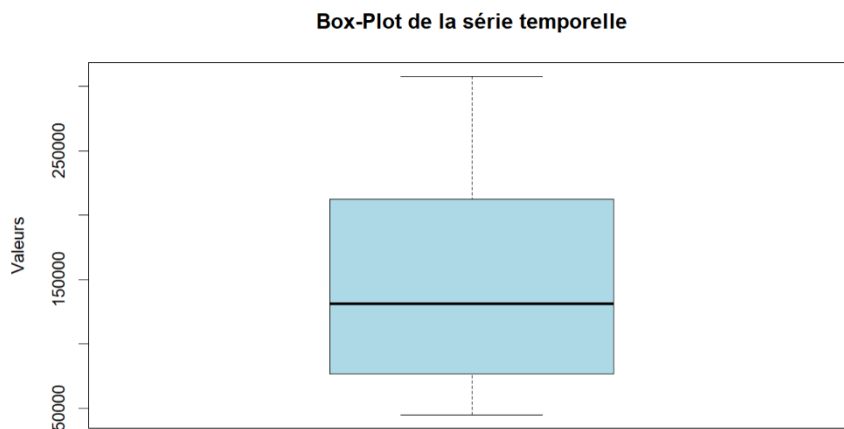


Figure 3 : Boîte à moustache de la série corrigée



L'histogramme ci-dessus, accompagné de sa courbe de densité (en rouge), permet de visualiser la distribution de la série temporelle corrigée. On observe clairement une asymétrie de la distribution, notamment avec une concentration de valeurs dans les intervalles bas (autour de 50 000 à 100 000) et une deuxième accumulation autour de 200 000 à 250 000. La forme de la courbe ne suit pas l'allure d'une distribution normale, elle présente deux bosses ce qui suggère une distribution bimodale.

Cette observation visuelle est renforcée par les résultats du test de Kolmogorov-Smirnov, dont la p-value est inférieure à 0,05. Cela implique le rejet de l'hypothèse de normalité au seuil de 5 %, confirmant que la série corrigée ne suit pas une loi normale. Ce résultat est



cohérent avec les valeurs de skewness positive et de kurtosis négative, déjà évoquées précédemment (voir annexe 2).

Concernant le boxplot, après correction de la série initiale en supprimant les points atypiques repérés, aucune nouvelle anomalie n'a été détectée dans la série ajustée. Cela montre que les irrégularités présentes au départ ont bien été traitées et que la série ne présente plus de valeurs atypiques.

#### **d) Détection de la saisonnalité**

Pour identifier la présence d'une saisonnalité dans la série de consommation de gaz, nous avons mobilisé quatre méthodes complémentaires : le test de Webel-Ollech (WO), les Seasonal Dummies (SD), l'analyse des fonctions d'autocorrélation (ACF/PACF), ainsi que le périodogramme.

Le test de Webel-Ollech permet de détecter la présence d'une structure saisonnière dans une série temporelle. Dans notre cas, la statistique du test est égale à 1, avec une p-value extrêmement faible ( $8.26e-11$ ). Cette valeur permet de rejeter l'hypothèse nulle d'absence de saisonnalité, suggérant ainsi une forte composante saisonnière dans la série.

De manière similaire, le test basé sur les Seasonal Dummies, qui consiste à introduire des variables fictives représentant les mois dans un modèle, il indique également une forte saisonnalité. En effet, la statistique du test est de 169.46, avec une p-value nulle, ce qui confirme de manière significative l'existence d'effets saisonniers.

En complément, le périodogramme de la série différenciée permet une visualisation des fréquences dominantes. On observe un pic très prononcé autour de la fréquence 0.08, ce qui correspond à une périodicité de 12 mois ( $1/0.08 = 12.5$ ). Cette fréquence est caractéristique d'une saisonnalité annuelle et vient appuyer les résultats des tests précédents.

Enfin, l'analyse des fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) permet de compléter l'étude de la saisonnalité. Sur le graphique de l'ACF, on observe des pics significatifs de 12 périodes, ce qui est typique d'une saisonnalité annuelle lorsque la fréquence de la série est mensuelle. Ces pics alternent entre positifs et négatifs, ce qui peut indiquer la présence de fluctuations saisonnières régulières dans le temps. La décroissance lente de l'ACF est également un signe caractéristique d'une composante saisonnière non négligeable. Le graphique de la PACF montre un pic élevé au lag 1, suivi de plusieurs valeurs significatives décroissantes. Cela suggère qu'un modèle AR (autorégressif) pourrait être approprié, et qu'il existe une dépendance à court terme dans la série, en plus de la structure saisonnière.

L'ensemble de ces éléments convergent vers la conclusion que la consommation de gaz présente une saisonnalité annuelle. Cela justifie le recours à des modèles intégrant une composante saisonnière pour l'analyse et la prévision de cette série temporelle.

Tableau 3 : Tests de Webel-Ollech et Seasonal Dummies (voir annexe 3 et 4)

Test	Statistique du test	p-value
Webel-Ollech Test (WO)	1	8.2647e-11
Seasonal Dummies (SD)	169.46	0

Figure 4 : Périodogramme sur la série ajustée

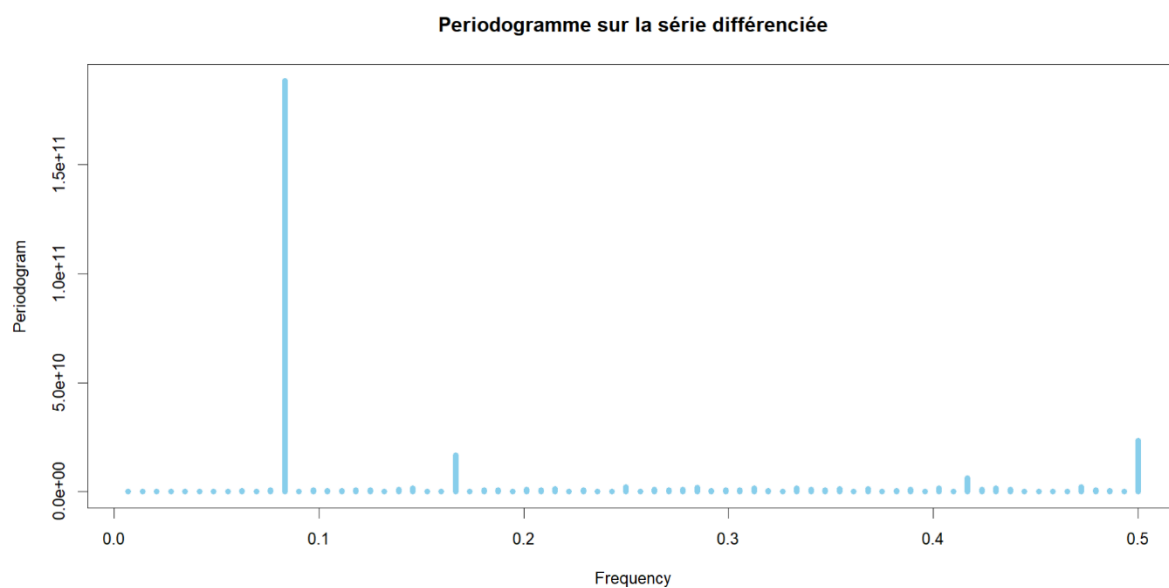
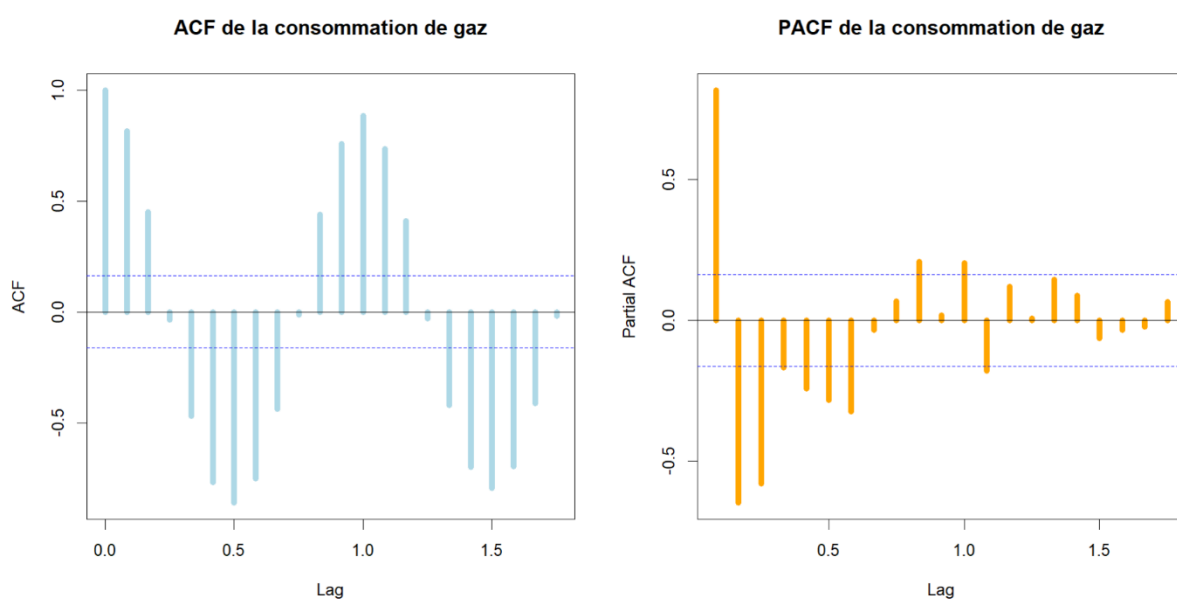


Figure 5 : ACF et PACF de la consommation de gaz



## e) Désaisonnalisation et décomposition

Nous avons utilisé trois approches pour déterminer le type de schéma additif ou multiplicatif de la série, deux approches visuelles permettent d'identifier le schéma de manière implicite et une méthode statistique va permettre de confirmer ces résultats.

La première méthode est la méthode du profil, qui consiste à superposer les saisons représentées par des courbes de profil sur un même graphique. Le graphique montre que les courbes semblent parallèles ce qui indique que le schéma de notre distribution semble additif. Par ailleurs, le graphique de comparaison de la saisonnalité par année, met bien en évidence la baisse de la consommation de gaz en été par rapport aux périodes d'hivers. Malgré la correction de la série un point semble également être atypique pour le mois de mai 2010 où la consommation de gaz était plus forte que les autres années (voir annexe 5).

La deuxième approche est la méthode de la bande, qui consiste à réaliser un graphique représentant la série chronologique, et tracer une droite passant respectivement par les valeurs minimums et maximums de chaque saison. Si ces deux droites sont parallèles, nous sommes en présence d'un modèle additif. Dans le cas contraire, c'est un modèle multiplicatif. Dans notre cas les droites semblent relativement parallèles ce qui confirme le schéma additif de notre distribution (voir annexe 6).

Ensuite la dernière approche est la méthode X13-ARIMA- SEATS qui propose un test pour la spécification log-level de la série. C'est la fonction `s_transform` qui permet de donner les résultats de cette méthode. Pour notre série on obtient `adjust = None` ce qui confirme que la série ne nécessite pas de transformation en logarithme. Également, `aicdiff = -2` donne la différence d'AICc entre les 2 spécifications =  $AICc(level) - AICc(log)$ , ce qui montre que le modèle sans transformation logarithmique a un AICc inférieur au modèle avec la transformation, car la valeur est négative.

Tableau 4 : Test log-level

tfunction	adjust	aicdiff
Auto	None	-2

Les trois approches différentes valident donc l'utilisation de la spécification d'un schéma additif pour notre série.

Ensuite, nous allons examiner le modèle X13-ARIMA-SEATS pour déterminer s'il existe des effets déterministes et observer les différents paramètres ARIMA.

La méthode X13-SARIMA-SEATS s'appuie également sur un modèle ARIMA mais pousse plus loin l'idée de filtrage statistique. En effet, elle identifie la structure statistique de la

série pour estimer séparément les différentes composantes. Le modèle estime donc les 3 composantes suivante : Tendances-cycle - l'évolution de long terme, Saisonnalité - les effets réguliers liés aux saisons ou aux périodes fixes et Irrégulier - les variations aléatoires, imprévues (bruit) (voir annexe 7).

Le modèle a appliqué une transformation log, ce qui indique que le modèle a choisi un schéma multiplicatif. Cela n'est pas en contradiction avec les résultats du test précédent, cependant la transformation en logarithme effectué par le modèle est probablement dans l'objectif de stabiliser la variance.

Le modèle ARIMA obtenu est un ARIMA(1, 0, 0, 0, 1, 1), ce qui suggère la présence d'une partie non-saisonnaire avec AR(1), d'une partie saisonnière avec SMA(1) et une saisonnalité d'ordre 12, d'où ARIMA(0,1,1)[12] où SMA(1) capture la composante saisonnière lissée. Enfin, il y a aucune différenciation non-saisonnaire, la série est considérée comme stationnaire en niveau (après log-transformation), sauf pour la partie saisonnière (une différenciation saisonnière D=1).

De plus, aucun effet déterministe n'est pris en compte, ce qui signifie qu'il n'y a pas de correction pour les effets de calendrier tels que les jours ouvrables, Pâques, les années bissextiles, ni d'effet moyen. L'absence d'effets déterministes semble logique, car la consommation de gaz dépend davantage de facteurs externes comme la température ou des chocs extérieurs.

Comme vu par la méthode du profil, il y a l'apparition d'une nouvelle valeur atypique. AO (Additive Outlier) : anomalie ponctuelle en mai 2010 et elle est hautement significative car la p-value est très faible.

Concernant les paramètres ARIMA :  $\Phi(1) = -0.39198$ ,  $B\Theta(1) = -0.85136$ . Ils indiquent une dépendance négative à court terme (effet de retour à la moyenne d'un mois sur l'autre) et une forte composante saisonnière négative à 12 mois, suggérant que les chocs survenus une année influencent fortement, mais de manière opposée, la consommation un an plus tard.

Enfin pour la qualité du modèle, l'erreur standard est faible (0.096), ce qui est bon signe. L'AIC et le BIC sont des indicateurs à comparer entre modèles. Ici, le BIC corrigé est négatif, ce qui montre un bon compromis entre ajustement et complexité.

## **II) Prédiction de la série saisonnière corrigée des points atypiques sur une année**

Dans cette section, nous nous intéressons à la prédiction de la série temporelle saisonnière après avoir corrigé les points atypiques. L'objectif est d'identifier les modèles les plus performants pour prédire les valeurs futures sur un horizon d'un an, en tenant compte de la structure saisonnière et de la dynamique propre à la série.

### **a) Estimation et prévisions des modèles**

#### **(i) Les méthodes Naïve, X13-SARIMA-SEATS, STL et STS**

##### **- Méthode Naïve**

La méthode naïve consiste à utiliser la dernière valeur observée de la série comme prévision pour toutes les périodes futures. Elle ne prend pas en compte les tendances ni les effets saisonniers, mais sert souvent de modèle de référence pour évaluer la performance des autres méthodes. Sa simplicité permet de comparer cette méthode par rapport aux autres, bien que ses résultats soient rarement les plus précis lorsque la série présente une dynamique complexe. Elle est une marche aléatoire et va représenter exactement la même prévision que les valeurs observées de l'année précédentes 2019. (Voir annexe 8).

##### **- X13-SARIMA-SEATS**

L'application de la méthode X13-ARIMA-SEATS à la série des prévisions révèle l'absence d'effets déterministes. Une transformation logarithmique a été appliquée à la série afin de stabiliser la variance. Le modèle sélectionné intègre une composante autorégressive non saisonnière d'ordre 1 (AR(1)) ainsi qu'une différenciation saisonnière. Le coefficient autorégressif  $\Phi(1) = -0.39556$  est significatif ( $p < 0.001$ ), indiquant une dépendance négative avec la période précédente. Par ailleurs, le terme de moyenne mobile saisonnière  $B\Theta(1) = -0.99955$  est hautement significatif ( $p < 2e-16$ ), traduisant une forte composante saisonnière. (Voir annexes 9 et 10).

##### **- STL (Seasonal-Trend decomposition Loess)**

STL est une méthode de décomposition souple permettant de séparer une série temporelle en trois composantes : tendance, saisonnalité et bruit résiduel. Elle repose sur des régressions locales (loess) pour ajuster ces composantes, ce qui lui permet de s'adapter à des évolutions non linéaires ou à une saisonnalité changeante dans le temps.

Elle est particulièrement efficace pour détecter les points atypiques et suivre les évolutions structurelles d'une série. Une fois la série décomposée, un modèle de type ETS peut être appliqué à la composante résiduelle désaisonnalisée. Dans notre cas, le modèle retenu est ETS(M,N,N), c'est-à-dire un modèle avec erreur multiplicative, sans tendance ni saisonnalité (car déjà capturé par STL). Le graphique de prévision obtenu montre que le modèle restitue fidèlement la forte saisonnalité annuelle de la consommation de gaz, avec des pics hivernaux clairement identifiables et des creux en été. Ce type de modèle est particulièrement adapté aux séries saisonnières sans tendance marquée à long terme (voir annexes 11 et 12).

- STS (Structural Time Series)

L'approche STS consiste à modéliser une série temporelle en la décomposant en composantes inobservables : tendance, saisonnalité et irrégulier. Chaque composante est spécifiée par un modèle simple (souvent de type ARIMA) et l'ensemble est formulé sous forme d'un modèle en espace d'état. Dans notre cas, les résultats montrent une variance élevée du niveau et de la pente (environ 83 millions et environ 939 millions). La variance de la saisonnalité est relativement faible (environ 3,6 millions). Enfin, la variance du bruit résiduel est nulle, ce qui est inhabituelle : cela suggère que la série est presque entièrement expliquée par les composantes structurelles (niveau, pente, saisonnalité), sans aléa résiduel. Cependant, le graphique de prévision révèle une limite importante du modèle : les valeurs prévues s'envolent rapidement, avec une incertitude croissante. Cette dérive est liée à l'absence d'une composante saisonnière bien intégrée, pourtant essentielle dans le cas de la consommation de gaz, qui suit un cycle annuel régulier. Cela suggère que le modèle STS utilisé ici s'appuie surtout sur la tendance récente pour faire la prévision, ce qui entraîne une dérive importante et irréaliste à moyen terme (voir annexe 13).

## **(ii) Les méthodes de lissage exponentiel : Holt-Winters, ETS, TBATS, ADAM ETS et ADAM ETS SARIMA**

- Holt-Winters

La méthode de Holt-Winters est une extension du lissage exponentiel simple, permettant de modéliser à la fois la tendance et la saisonnalité. Elle existe sous deux formes : additive (pour les séries où les variations saisonnières sont constantes) et multiplicative (si elles augmentent avec le niveau de la série). Très intuitive, cette méthode est adaptée aux séries présentant des schémas réguliers et bien définis, notamment lorsqu'on observe des cycles saisonniers nets. Dans le cas de la consommation de gaz, la méthode de Holt-Winters s'adapte correctement. Le graphique de prévision montre que le modèle capture

efficacement le cycle saisonnier, en suivant les hausses hivernales et les baisses estivales, tout en projetant ces schémas sur les périodes futures (voir annexe 14).

- ETS (Error, Trend, Seasonality)

Le modèle ETS est une famille de modèles de séries temporelles fondée sur le lissage exponentiel. Il formalise de façon systématique les composantes de la série : l'erreur (E), la tendance (T), et la saisonnalité (S), chacune pouvant être additive ou multiplicative. Le modèle sélectionne automatiquement la meilleure combinaison (par ex. ETS(A,A,A) ou ETS(M,A,N)). Il est particulièrement efficace pour des prévisions rapides, fiables et interprétables, notamment pour des séries économiques ou de consommation. Ici, le modèle retenu est un ETS(M,N,M), ce qui signifie que l'erreur et la saisonnalité sont multiplicatives, tandis que la série ne présente pas de tendance structurée. Les valeurs estimées montrent une saisonnalité marquée (avec les coefficients saisonniers variant significativement d'un mois à l'autre), et une erreur multiplicative modérée ( $\sigma = 0.1035$ ), indiquant une bonne capacité du modèle à ajuster les données. De plus, le modèle donne davantage de poids aux valeurs passées ( $\alpha = 0.2423$ ) et réagit lentement aux changements récents. Les prévisions sont cohérentes et respectent bien la dynamique saisonnière observée dans l'historique de la série. Le graphique des prévisions reflète cette régularité : les fluctuations saisonnières sont bien reproduites, et les intervalles de confiance sont réalistes. Cela fait du modèle ETS un bon compromis entre simplicité, performance et interprétabilité (voir annexes 15 et 16).

- TBATS (Trigonometric, Box-Cox, ARMA errors, Trend and Seasonal)

Le modèle TBATS est conçu pour les séries temporelles présentant une saisonnalité multiple ou complexe (par exemple : hebdomadaire + annuelle). Il intègre plusieurs outils : transformation Box-Cox (pour stabiliser la variance), composantes trigonométriques (pour modéliser la saisonnalité), erreurs ARMA, et tendance. TBATS est particulièrement utile lorsque les données montrent des cycles non standards, par exemple avec des pics irréguliers ou des périodes de saisonnalité qui changent dans le temps. Dans notre cas, il détecte une saisonnalité annuelle bien marquée et applique une transformation de Box-Cox ( $\lambda \approx 0$ ), ce qui suggère une stabilisation efficace de la variance. Les paramètres estimés ( $\sigma \approx 0.10$ , valeurs faibles de  $\gamma$ ) indiquent une bonne qualité d'ajustement, avec une tendance globalement stable et une saisonnalité dominante. Aussi, un  $\alpha$  proche de 0 ( $\alpha = 0.07851826$ ) signale une forte dépendance aux valeurs passées. Les prévisions sur 12 mois sont cohérentes avec le comportement observé de la série (voir annexes 17 et 18).

#### - ADAM ETS

Le modèle ADAM (Automatic Dynamic Adaptive Model) est une extension moderne et puissante des modèles ETS classiques. Il conserve la logique des modèles ETS, en décomposant la série en Erreur (E), Tendance (T) et Saisonnalité (S), tout en y ajoutant plusieurs améliorations : une structure d'erreur plus flexible (erreur additive, multiplicative, ou distributions non gaussiennes comme Laplace), la possibilité d'intégrer des variables explicatives (exogènes) et un cadre bayésien pour une estimation dynamique et adaptative des paramètres.

On obtient le modèle ADAM ETS(MNM) + ARIMA(1,0,0). Contrairement aux modèles traditionnels, ADAM permet une structure d'erreur flexible (ici multiplicative), l'intégration de composantes ARIMA (avec un  $AR(1) = 0.3693$ ), et l'utilisation de distributions non gaussiennes comme la Laplace, offrant une meilleure tolérance aux valeurs extrêmes. Le cadre bayésien considère les paramètres comme des variables aléatoires, ce qui permet au modèle de s'ajuster dynamiquement à de nouvelles données. Les coefficients de lissage très faibles ( $\alpha = 0.0000$ ,  $\gamma = 0.0034$ ) indiquent une série stable dans ses composantes internes. Ce modèle a produit des prévisions réalistes et saisonnières, telles que 271 441 en janvier 2020 et 71 588 en juin 2019, illustrant clairement les pics de consommation hivernale et les creux estivaux (voir annexes 19, 20 et 21).

#### - ADAM ETS SARIMA

Le modèle ADAM ETS(MNN) + ARIMA(2,0,0) combine la structure des modèles de lissage exponentiel (ETS) avec une dynamique autorégressive (ARIMA), offrant ainsi une double modélisation des composantes structurelles (niveau, saisonnalité) et des dépendances temporelles (effets mémoire). Ici, le modèle utilise une erreur multiplicative, aucune tendance ni saisonnalité dans le bloc ETS, et deux composantes AR (avec  $AR(1) = 1.4574$  et  $AR(2) = -0.7445$ ), capturant une mémoire forte dans la série. La distribution choisie automatiquement est l'inverse gaussienne, souvent utilisée pour des séries strictement positives et asymétriques, comme c'est le cas ici avec la consommation de gaz. Le coefficient de lissage très faible ( $\alpha = 0.0001$ ) suggère que le niveau de la série reste stable dans le temps, tandis que les variations sont surtout captées par la composante AR. Ce modèle a généré des prévisions réalistes, reflétant des dynamiques saisonnières tout en tenant compte de l'inertie et des ajustements à court terme. Grâce à cette combinaison ETS + ARIMA, ce modèle est bien adapté aux séries présentant à la fois des effets saisonniers marqués et des corrélations internes complexes (voir annexes 22, 23 et 24).



### (iii) les modèles SARIMA(p,d,q)(P,D,Q)[12] et SSARIMA.

#### - SARIMA(p,d,q)(P,D,Q)[12]

Le modèle SARIMA (Seasonal AutoRegressive Integrated Moving Average) est une extension du modèle ARIMA qui permet de modéliser des séries temporelles présentant à la fois des composantes non saisonnières (ARIMA) et des effets saisonniers récurrents. Dans notre cas, le modèle sélectionné automatiquement est ARIMA(1,0,0)(2,1,0)[12], ce qui signifie qu'il inclut une composante autorégressive non saisonnière d'ordre 1 ( $AR(1) = 0.4638$ ), deux composantes autorégressives saisonnières ( $SAR(1) = -0.5400$  et  $SAR(2) = -0.4611$ ), et une différenciation saisonnière d'ordre 1 pour rendre la série stationnaire. Ce modèle capture efficacement les dépendances à court terme ainsi que les effets saisonniers à un et deux ans d'intervalle. Les performances sont globalement bonnes avec un RMSE de 15 785 et une autocorrélation résiduelle faible ( $ACF1 = 0.044$ ), ce qui indique que les résidus ne présentent pas de structure temporelle notable. Ainsi, ce modèle offre une représentation robuste de la dynamique saisonnière de la consommation de gaz naturel (voir annexes 25 et 26).

#### - SSARIMA

Le modèle SSARIMA (State Space AutoRegressive Integrated Moving Average) est une généralisation du modèle SARIMA, formulée dans un cadre espace d'états. Il permet une estimation plus souple et souvent plus efficace, en combinant différentes fréquences de saisonnalité, il est performant pour de petites tailles d'échantillon et plus flexible. Dans notre cas, le modèle retenu est SARIMA(2,0,3)[1](3,0,0)[12], ce qui signifie qu'il y a une partie non saisonnière avec  $AR(2)$  et  $MA(3)$  (lag 1), capturant les dynamiques locales. Une composante saisonnière avec  $AR(3)$  (lag 12), qui exerce une forte influence. Et, aucun terme différenciateur, la série étant considérée comme stationnaire en niveau et en saison. Les coefficients estimés montrent notamment une forte composante de moyenne mobile à court terme ( $MA(1) = 0.7063$ ) ainsi qu'une saisonnalité marquée ( $AR(3)[lag 12] = 0.4780$ ). Les prévisions fournies sont cohérentes avec la logique saisonnière de la série (voir annexes 27 et 28).

## b) Présentation sous forme de tableau des paramètres des modèles

Tableau 5 : Paramètres des modèles

Modèles	X13 ARIMA SEATS	STL + ETS(M,N,N)	ETS (M,N,M)	TBATS (0, {0,0}, -, {<12,4>})	ADAM ETS(MNM)+ ARIMA(1,0,0)	ADAM ETS (MNN)+ ARIMA(2,0,0)	ARIMA (1,0,0) (2,1,0)[12]	SARIMA (2,0,3)[1] (3,0,0)[12]
alpha		0.2713	0.2423	0.07851826	0.0000	1e-04		
beta	-0.99955							
gamma			1e-04	0.007687197	0.0034			
phi	-0.39556							
lambda				0				
ar1					0.3693	1.4574	0.4638	-0.1391
ar2						-0.7445		0.1297
ar3								0.0000
sar1							-0.5400	0.3759
sar2							-0.4611	0.1133
sar3								0.4780
ma1								0.7063
ma2								0.1982
ma3								0.0722
sma1								
sma2								
Distribution	Normale	Normale	Normale	Normale	Laplace	Inverse Gaussian	Normale	Normale
AIC	2644	3208.813	3192.504	3193.354	2905.907	3091.625	2709.94	2952.119
AICc	2644	3209.000	3196.607		2911.228	3092.521	2710.29	2953.583

Le modèle X13-ARIMA-SEATS est le meilleur selon les critères AIC et AICc, car il présente les valeurs les plus faibles pour ces deux indicateurs. Un AIC et un AICc plus bas indiquent que ce modèle offre un meilleur équilibre entre la qualité de l'ajustement et la complexité du modèle. Cela signifie qu'il est plus efficace pour capturer les dynamiques des données tout en évitant le surajustement.

## III) Représentation graphique de l'évolution des prévisions des différents modèles

Le graphique (voir figure 6) compare l'évolution des prévisions obtenues à partir de plusieurs modèles statistiques avec la série observée, sur une période de 12 mois. L'objectif est de visualiser la capacité de chaque modèle à reproduire la dynamique réelle des données (courbe des données observées présente en noir).

Le modèle Naïf, représenté en bleu, suit les grandes tendances mais avec un certain décalage comme on peut le constater, notamment lors du rebond. Ce modèle repose sur

l'hypothèse que la meilleure prévision pour une période est la dernière valeur observée, ce qui le rend simple mais peu performant lorsqu'il y a des variations brutales.

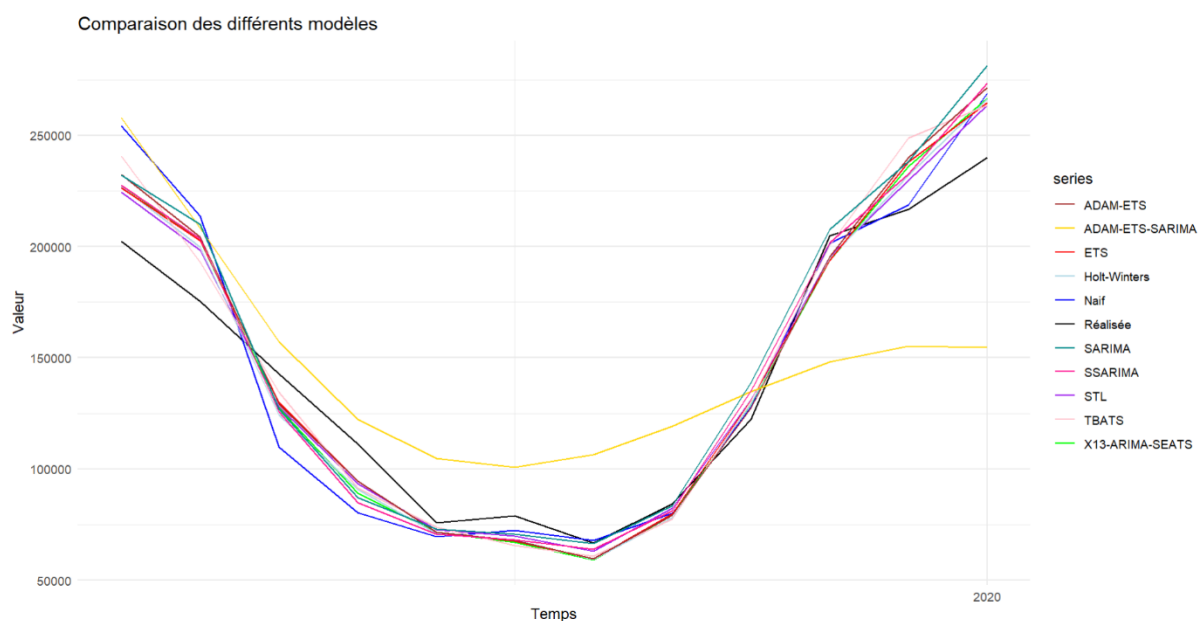
Le modèle qui ne suit pas du tout la réalité c'est le modèle STS (voir annexe 29), il affiche une performance nettement inférieure aux autres modèles. Alors que la plupart des séries suivent la courbe réalisée avec une certaine précision, StructTS s'en éloigne de manière croissante au fil du temps.

Le modèle ADAM-ETS-SARIMA (jaune) se distingue par une performance relativement faible. Il surestime de manière systématique les valeurs, en particulier lors de la phase de baisse. Ce modèle est assez lisse et ne prend pas bien en compte les variations réelles ce qui suggère un ajustement mal spécifié.

En revanche, les modèles ETS (rouge foncé), Holt-Winters (bleu ciel), SARIMA (cyan), SSARIMA (rose), STL (violet), TBATS (rose clair), et X13-ARIMA-SEATS (vert) suivent très fidèlement la courbe réalisée. Ils captent à la fois la forte baisse centrale pour la période d'été et le rebond final pour l'hiver avec une hausse de la consommation de gaz, avec des écarts très faibles par rapport à la réalité. Ces modèles confirment leur efficacité pour représenter des séries avec des effets saisonniers.

À la suite de la visualisation graphique des différents modèles, nous constatons qu'un grand nombre d'entre eux parviennent à reproduire assez fidèlement la série réalisée. Cependant, cette première comparaison visuelle ne permet pas à elle seule de désigner le modèle le plus performant. C'est pourquoi nous allons désormais approfondir l'analyse afin d'identifier celui offrant la meilleure qualité de prévision et la plus grande précision.

**Figure 6 : Comparaison des différents modèles de prévisions (sans le modèle STS)**



#### IV) Qualité de prévision

Le tableau présente deux types d'évaluation pour les différents modèles : le MSE (Mean Squared Error) et le R2OOS ( $R^2$  Out-Of-Sample).

Le MSE mesure directement l'écart quadratique moyen entre les valeurs prédites par chaque modèle et les valeurs réelles observées. Plus cette valeur est basse, plus le modèle est précis dans ses prévisions. Dans notre cas, le modèle STL se distingue avec un MSE de 213678906, ce qui est relativement faible comparé à la plupart des autres modèles.

Le R2OOS, quant à lui, compare spécifiquement la performance de chaque modèle par rapport au modèle naïf (qui sert de référence). Une valeur positive indique une amélioration par rapport au modèle naïf, tandis qu'une valeur négative signifie que le modèle performe moins bien que cette référence simple. Le modèle STL affiche le meilleur R2OOS avec 0,6430646, ce qui signifie qu'il améliore les prévisions de 64,3% par rapport au modèle naïf.

Les modèles StructTS et ADAM-ETS-SARIMA présentent des R2OOS fortement négatifs (-179,0441163 et -2,2127756), indiquant qu'ils sont moins performants que le simple modèle naïf. Le modèle StructTS est particulièrement problématique avec un MSE extrêmement élevé (10778317866).

Parmi les modèles offrant une amélioration par rapport au modèle naïf, on peut établir un classement décroissant selon le R2OOS : STL (0,64) > Holt-Winter (0,55) > ETS (0,55) > X13-ARIMA-SEATS (0,50) > SSARIMA (0,43) > ADAM-ETS (0,43) > TBATS (0,42) > SARIMA (0,24).

En conclusion, le modèle STL apparaît comme le choix optimal pour cette série temporelle particulière, combinant l'erreur de prédiction la plus faible (bas MSE) et l'amélioration la plus élevée par rapport à l'approche naïve (R2OOS élevé).

Tableau 6 : MSE et R2OOS des différents modèles de prévisions

Modèle	MSE	R2OOS
Naïf	598648714	0.0000000
STL	213678906	0.6430646
StructTS	107783178664	-179.0441163
X13-ARIMA-SEATS	298759649	0.5009433
Holt-Winter	272238536	0.5452449
ETS	270336652	0.5484219
TBATS	347640217	0.4192918
ADAM-ETS	338847829	0.4339789
ADAM-ETS-SARIMA	1923324000	-2.2127756
SARIMA	452146991	0.2447207
SSARIMA	341525501	0.4295060

La figure 7 ci-dessous présente l'évolution de l'erreur quadratique cumulée pour les différents modèles de prévision sur 12 mois. On observe que :

Le modèle ADAM\_ETS\_SARIMA (ligne jaune) montre une performance assez mauvaise. Son erreur quadratique cumulée augmente très rapidement, surtout à partir de la période 8. Cette forte évolution de l'erreur quadratique confirme le R2OOS négatif observé précédemment.

Le modèle Naïf (ligne bleue) présente également une performance médiocre, avec une erreur cumulée plus élevée comparé aux autres modèles. Cette ligne sert de référence pour évaluer l'amélioration apportée par les autres modèles.

Le modèle qui reste le moins bon est comme on pouvait s'y attendre au vu des résultats précédents, c'est le modèle STS. En effet, l'évolution de son erreur de prédiction quadratique cumulée (CSPE) démontre une détérioration de sa précision prévisionnelle dès les premières périodes (voir annexe 30).

Le modèle SARIMA (cyan) affiche une performance moyenne, avec une erreur cumulée qui reste contrôlée jusqu'à la période 10, puis augmente significativement vers la fin.

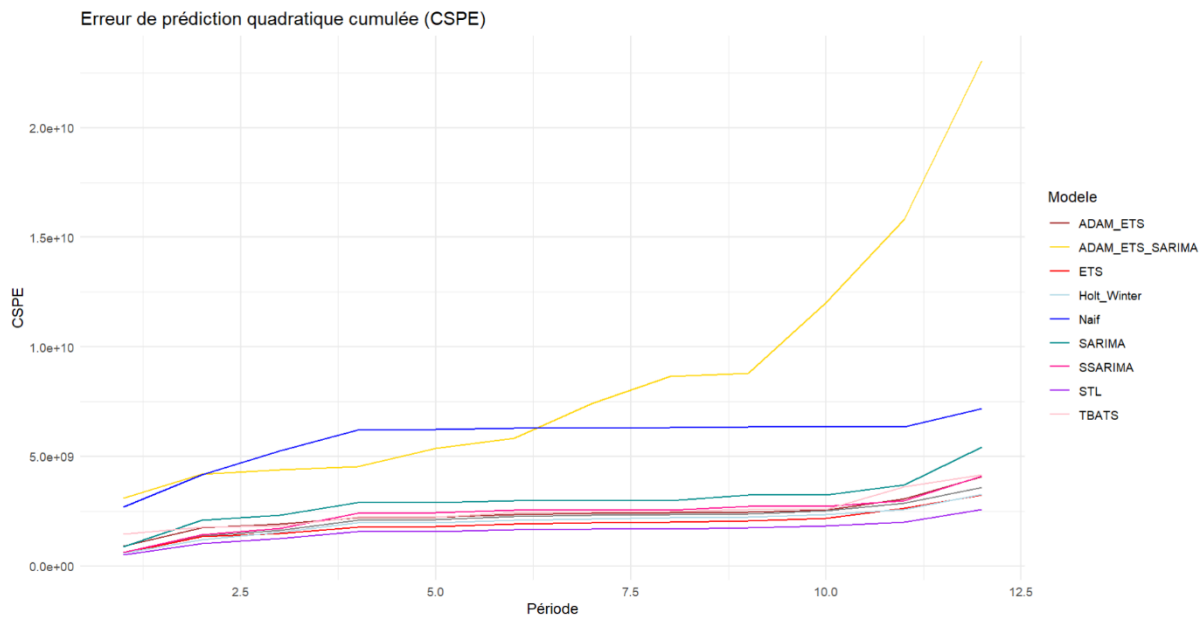
Les modèles les plus performants sont ceux dont les lignes restent dans la partie inférieure du graphique, notamment :

- STL (ligne violette) qui maintient l'erreur cumulée la plus basse sur l'ensemble des périodes
- ETS et Holt-Winter qui suivent des trajectoires similaires avec des erreurs cumulées relativement contenues

Cette visualisation confirme l'analyse des MSE et du R2OOS du tableau précédent : le modèle STL se révèle être le plus performant avec l'erreur cumulée la plus faible, tandis que ADAM\_ETS\_SARIMA et STS présentent les pires résultats avec une erreur qui explose sur les dernières périodes.

On observe également que la plupart des modèles connaissent une augmentation de l'erreur vers la fin de la période d'observation (après la période 10), ce qui pourrait indiquer une difficulté générale à prévoir avec précision les valeurs futures de la série temporelle.

**Figure 7 : Erreur de prédiction quadratique cumulée (CSPE) sans le modèle STS**



## V) Test de précision

Le tableau présente les résultats du test de Diebold-Mariano (DM) qui évalue statistiquement si les différences de précision entre les modèles de prévision sont significatives. Ce test compare la qualité prédictive d'un modèle par rapport au modèle naïf, avec la p-value comme indicateur clé.

Pour le test DM, une p-value inférieure à 0,05 indique que le modèle testé est significativement meilleur que le modèle naïf. À l'inverse, une p-value élevée suggère que la différence n'est pas statistiquement significative.

Le modèle STL présente une p-value de 0,039, ce qui est inférieur au seuil de 0,05. Cela confirme que ces prévisions sont meilleures par rapport au modèle naïf, c'est ce qui avait déjà été observée dans les calculs du MSE et R2OOS précédents. C'est le seul modèle qui satisfait ce critère.

Les modèles Holt-Winter (0,057), ETS (0,064) et X13-ARIMA-SEATS (0,072) présentent des p-values proches mais légèrement supérieures au seuil de 0,05. Ces résultats suggèrent que leurs performances sont meilleures que le modèle naïf mais cette amélioration n'est pas tout à fait significative au seuil de 95%.

Les modèles STS (0,999) et ADAM-ETS-SARIMA (0,971) affichent des p-values extrêmement élevées, ce qui confirme leur faible performance déjà observée dans les analyses précédentes.

Les autres modèles (TBATS, ADAM-ETS, SARIMA, SSARIMA) présentent des p-values intermédiaires (entre 0,090 et 0,234), indiquant que leur amélioration par rapport au modèle naïf n'est pas statistiquement significative.

En conclusion, le test de Diebold-Mariano renforce les conclusions précédentes : le modèle STL est statistiquement le plus performant pour cette série temporelle, alors que les modèles STS et ADAM-ETS-SARIMA sont les moins adaptés.

**Tableau 7 : Test de Diebold-Mariano des différents modèles par rapport au modèle naïf**

Modèle	p-value
STL	0.039
STS	0.999
X13-ARIMA-SEATS	0.072
Holt-Winter	0.057
ETS	0.064
TBATS	0.102
ADAM-ETS	0.090
ADAM-ETS-SARIMA	0.971
SARIMA	0.234
SSARIMA	0.102

## **VI) Prédiction sur une année avec un pas de 1 mois avec le meilleur modèle**

On utilise donc le modèle STL pour faire la prédiction sur une année avec un pas de 1 mois en fenêtre roulante. L'idée est de savoir si le modèle fait des bonnes prévisions dans le futur et pas seulement avec les données utilisées pour l'entraîner. La prédiction en fenêtre roulante est un indicateur de robustesse du modèle hors-échantillons.

Concrètement, on définit une fenêtre d'estimation fixe (correspondant à l'ensemble des données disponibles, hors période de prédiction), sur laquelle on ajuste le modèle et on génère une prédiction à un horizon d'un mois. Ensuite, on décale cette fenêtre d'un mois vers l'avant (en retirant le mois le plus ancien et en ajoutant le mois suivant), et on répète le processus pour chaque mois de l'année à prévoir, soit 12 itérations. Ce procédé permet d'obtenir des prévisions plus réalistes, reflétant une situation où les données arrivent au fur et à mesure du temps. Cette méthode permet d'éviter le sur-apprentissage et d'identifier quand le modèle commence à ne plus prédire correctement.

La figure 8 illustre la comparaison entre les prévisions issues du modèle STL en rolling window et les observations réelles. On observe une bonne adéquation, notamment sur la

deuxième moitié de l'horizon de prévision : les mois de septembre, octobre et novembre présentent des valeurs prévues très proches des valeurs observées (voir annexes 31 et 32).

En ce qui concerne les indicateurs de performance, le MSE (Mean Squared Error) de la prévision rolling est de 395 007 329, ce qui est plus élevé que celui obtenu avec une prévision STL classique. Cela signifie que la prévision en rolling window génère davantage d'erreur que la prévision réalisée directement à partir de l'historique complet. En revanche, le R<sup>2</sup>OS (ou R2OOS), qui mesure l'amélioration par rapport à une prévision naïve, montre une amélioration de 34 %, ce qui, bien que positif, reste légèrement inférieur à la performance du modèle STL sans rolling.

Pour conclure, la prévision en fenêtre roulante permet de tester le modèle STL dans des conditions réalistes et en temps réel. Bien qu'elle produise une erreur globale légèrement supérieure à la version classique, elle démontre une capacité intéressante à capter la dynamique future de la série, tout en évitant les biais liés à un entraînement sur l'ensemble des données. Elle reste donc un outil pertinent pour évaluer la fiabilité temporelle d'un modèle de prévision.

Figure 8 : Comparaison prévisions STL (fenêtre roulante) avec les valeurs observées

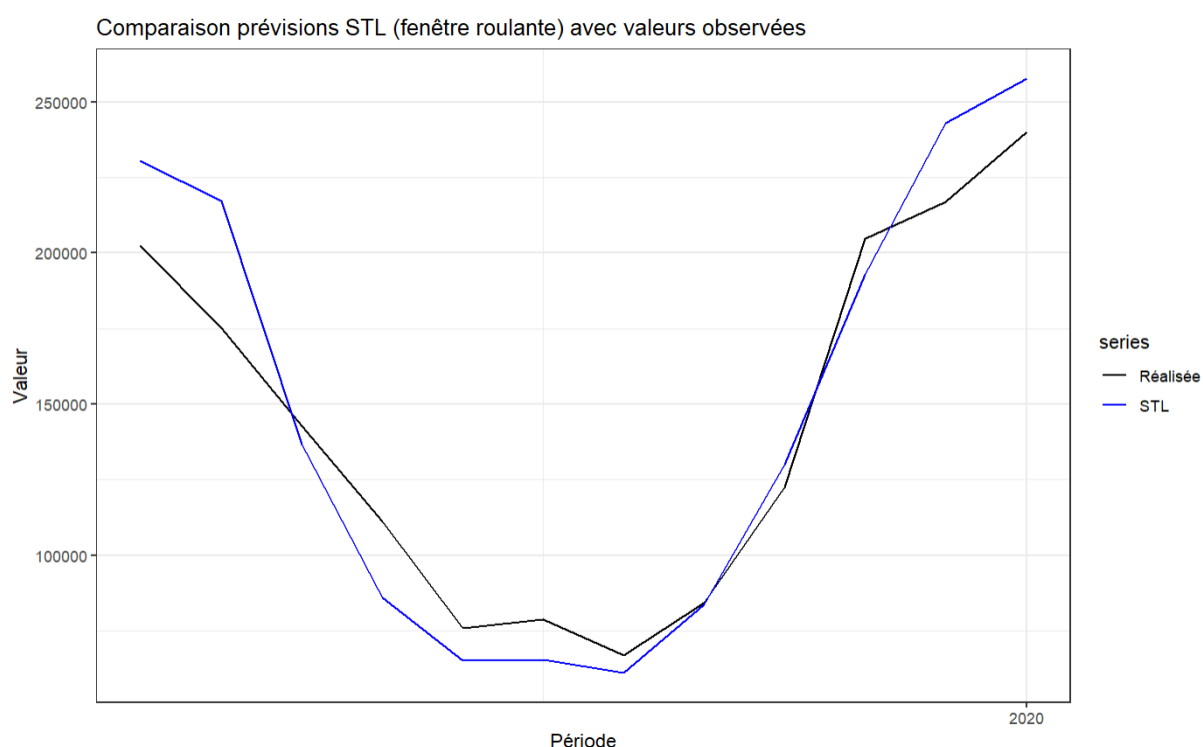


Tableau 8 : Qualité de prévision du modèle STL en fenêtre roulante

Modèle	MSE	R2OOS
STL en fenêtre roulante	395007329	0.34



## Conclusion

Cette étude a permis d'analyser en profondeur la consommation mensuelle de gaz naturel en France sur la période 2008-2020, ce qui nous a permis d'avoir un éclairage sur les dynamiques de cette ressource énergétique importante dans le paysage français.

L'analyse exploratoire a révélé une série temporelle caractérisée par une forte saisonnalité annuelle, avec des pics de consommation systématiques en période hivernale et des creux durant les mois d'été. Cette structure saisonnière a été confirmée par différentes méthodes statistiques. La série présente également une légère tendance à la baisse à partir de 2012, probablement liée aux politiques de transition énergétique et à l'évolution du mix énergétique français.

La comparaison de différentes méthodes de prévision a démontré que le modèle STL offre les meilleures performances. Ce modèle surpasse nettement les approches naïves et confirme sa pertinence pour des séries présentant une forte saisonnalité. À l'inverse, les modèles StructTS et ADAM-ETS-SARIMA se sont révélés moins adaptés à cette série temporelle.

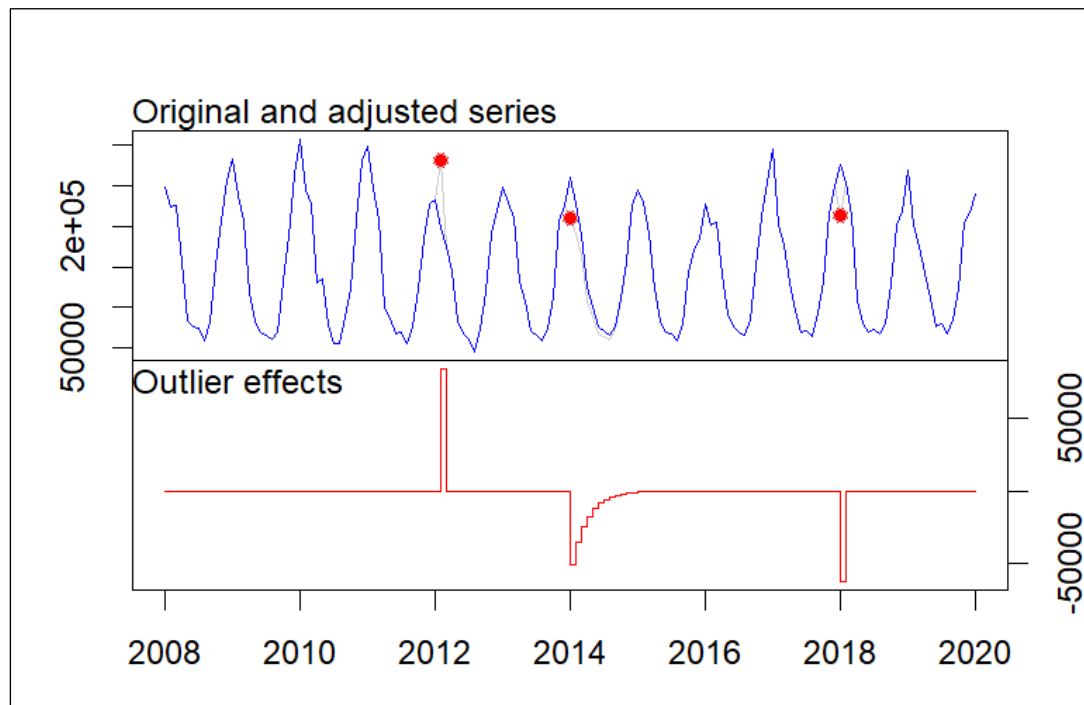
L'application d'une approche en fenêtre roulante au modèle STL a permis de tester sa robustesse dans des conditions proches d'une utilisation en temps réel. Bien qu'elle soit légèrement moins précise que la version STL classique.

Ces résultats soulignent l'importance d'une modélisation adaptée aux caractéristiques spécifiques de la série pour obtenir des prévisions fiables de la consommation de gaz. Ils peuvent servir de base à la planification énergétique, à l'optimisation des infrastructures de distribution, et à l'élaboration de politiques publiques visant à réduire la dépendance aux énergies fossiles.

Dans une perspective plus large, cette analyse s'inscrit dans un contexte de transition énergétique où la compréhension des dynamiques de consommation des énergies fossiles devient cruciale pour anticiper et accompagner les changements du système énergétique français dans sa transition vers une économie bas-carbone.

## Annexes

### ANNEXE 1 : Graphique de la détection des points atypiques avec tsoutliers



### ANNEXE 2 : Test de Kolmogorov-Smirnov

```
Asymptotic one-sample Kolmogorov-Smirnov test  
data: serie_std  
D = 0.15646, p-value = 0.001652  
alternative hypothesis: two-sided
```

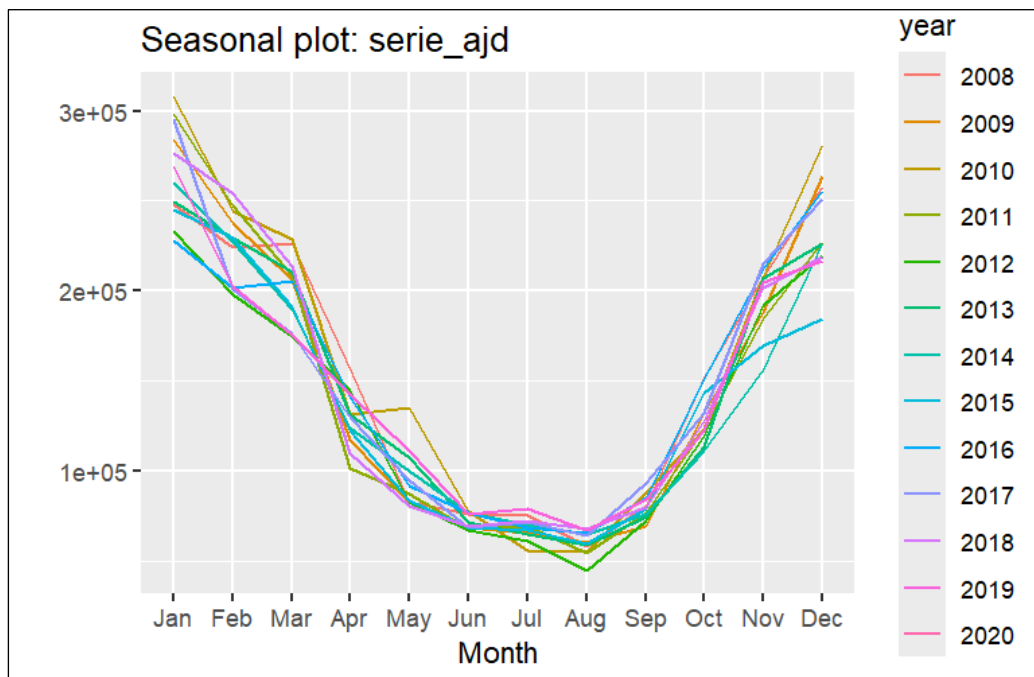
### ANNEXE 3: Test de Weibel-Ollech

```
Test used: WO  
Test statistic: 1  
P-value: 0 8.2647e-11 7.842338e-11
```

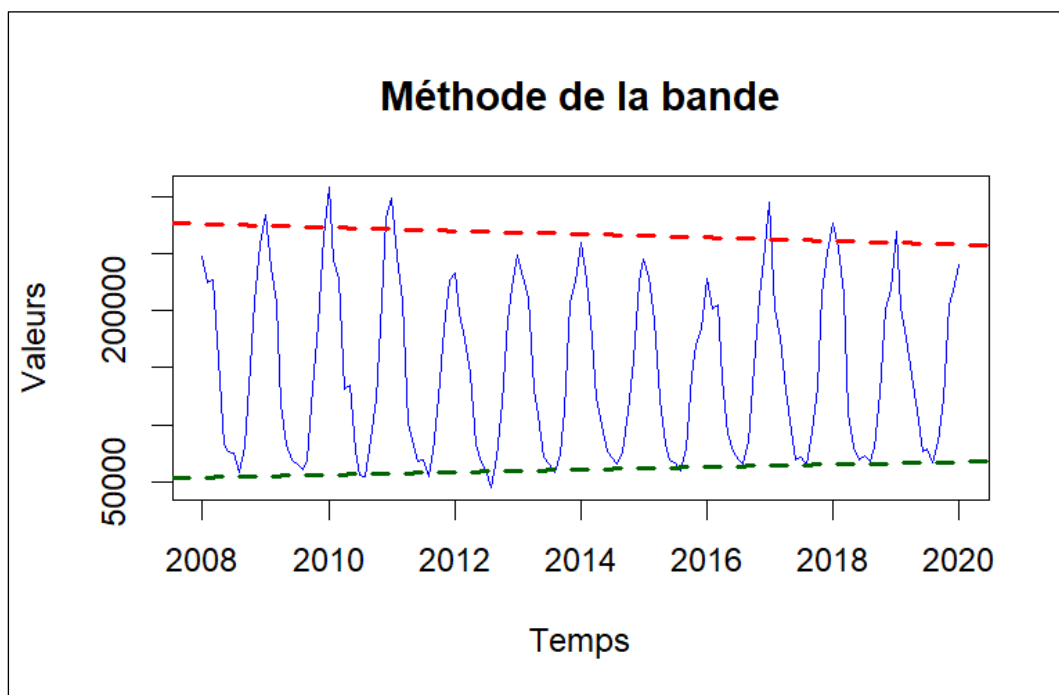
### ANNEXE 4: Test de Seasonal Dummies

```
Test used: SeasonalDummies  
Test statistic: 169.46  
P-value: 0
```

#### ANNEXE 5 : Graphique de la méthode du profil



#### ANNEXE 6 : Méthode de la bande



## ANNEXE 7 : Régression de la méthode X13-ARIMA-SEATS

```
y = regression model + arima (1, 0, 0, 0, 1, 1)

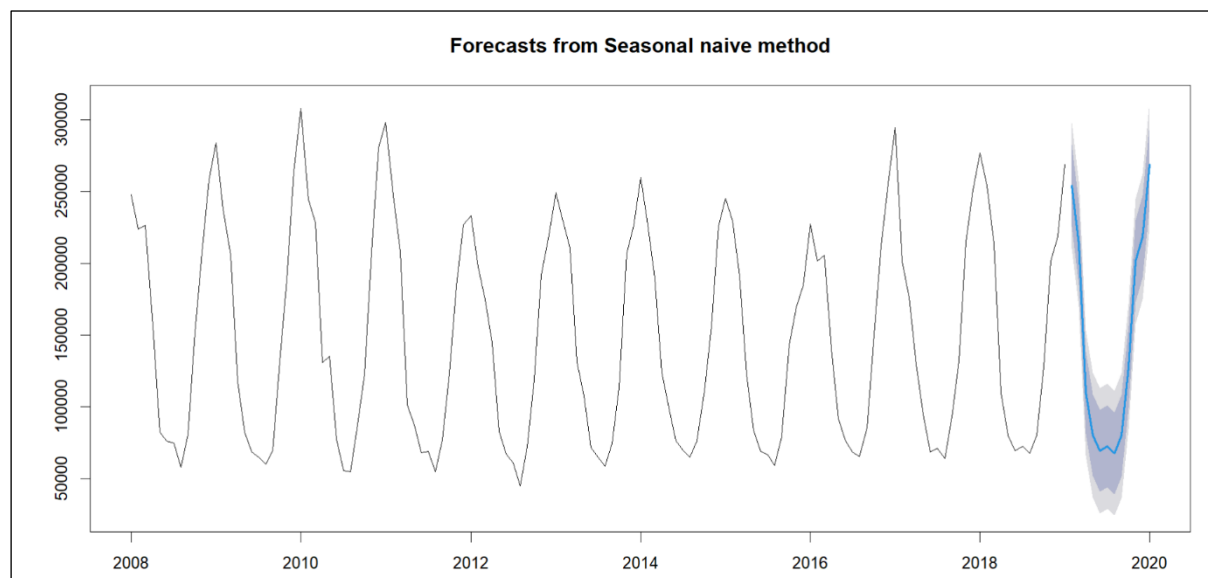
Model: RegARIMA - X13
Estimation span: from 1-2008 to 1-2020
Log-transformation: yes
Regression model: no mean, no trading days effect, no leap year effect, no Easter effect,
outliers(1)

Coefficients:
ARIMA:
      Estimate Std. Error  T-stat Pr(>|t|)
Phi(1)   -0.39198    0.08091  -4.844 3.48e-06 ***
BTheta(1) -0.85136    0.06035 -14.106 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

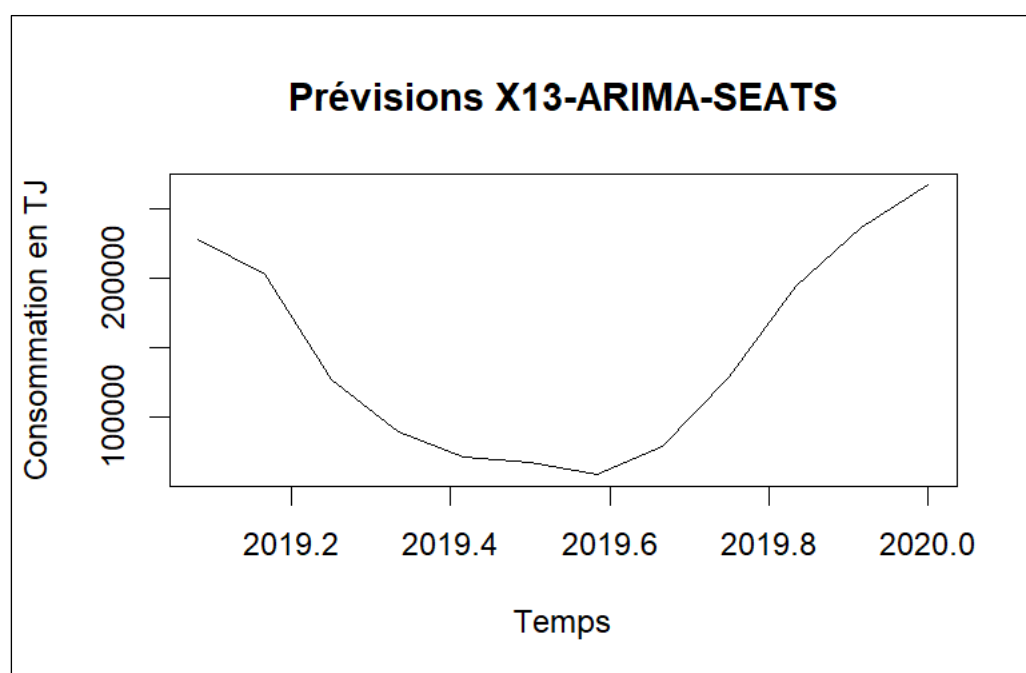
Regression model:
      Estimate Std. Error  T-stat Pr(>|t|)
AO (5-2010)  0.37663    0.08943   4.211 4.65e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09631 on 129 degrees of freedom
Log likelihood = 114.8, aic = 2907, aicc = 2907, bic(corrected for length) = -4.57
```

## ANNEXE 8 : Prévisions avec la méthode naïve



## ANNEXE 9 : Prédiction avec la méthode X13-ARIMA-SEATS à partir de 2019



## ANNEXE 10 : Régression de la méthode X13-ARIMA-SEATS à partir de 2019

```
y = regression model + arima (1, 0, 0, 0, 1, 1)
```

Model: RegARIMA - X13

Estimation span: from 1-2008 to 1-2019

Log-transformation: yes

Regression model: no mean, no trading days effect, no leap year effect, no Easter effect, outliers(1)

Coefficients:

ARIMA:

	Estimate	Std. Error	T-stat	Pr(> t )
Phi(1)	-0.39556	0.08458	-4.677	7.64e-06 ***
BTheta(1)	-0.99955	0.06991	-14.299	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Regression model:

	Estimate	Std. Error	T-stat	Pr(> t )
AO (5-2010)	0.37882	0.08918	4.248	4.26e-05 ***

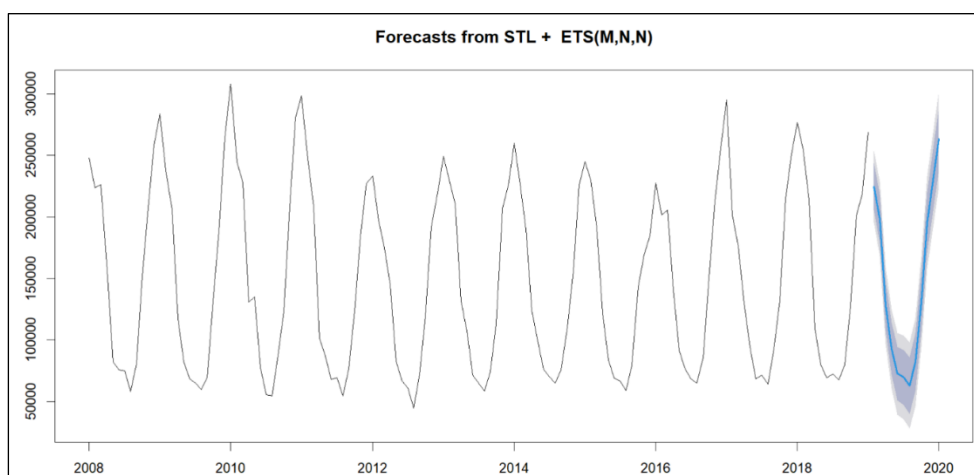
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09032 on 117 degrees of freedom

Log likelihood = 104.8, aic = 2644, aicc = 2644, bic(corrected for length) = -4.69

## ANNEXE 11 : Pr vision avec la m thode STL



## ANNEXE 12 : Estimation de la pr vision avec la m thode STL

```
Forecast method: STL + ETS(M,N,N)

Model Information:
ETS(M,N,N)

Call:
ets(y = x, model = etsmodel, allow.multiplicative.trend = allow.multiplicative.trend)

Smoothing parameters:
alpha = 0.2713

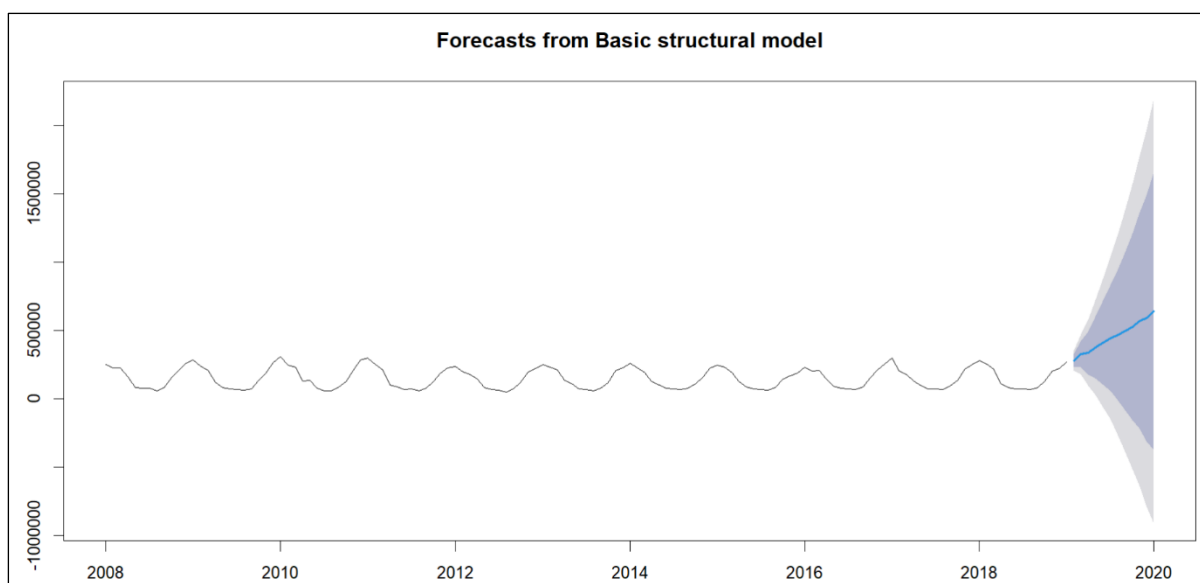
Initial states:
l = 152165.2002

sigma: 0.1012

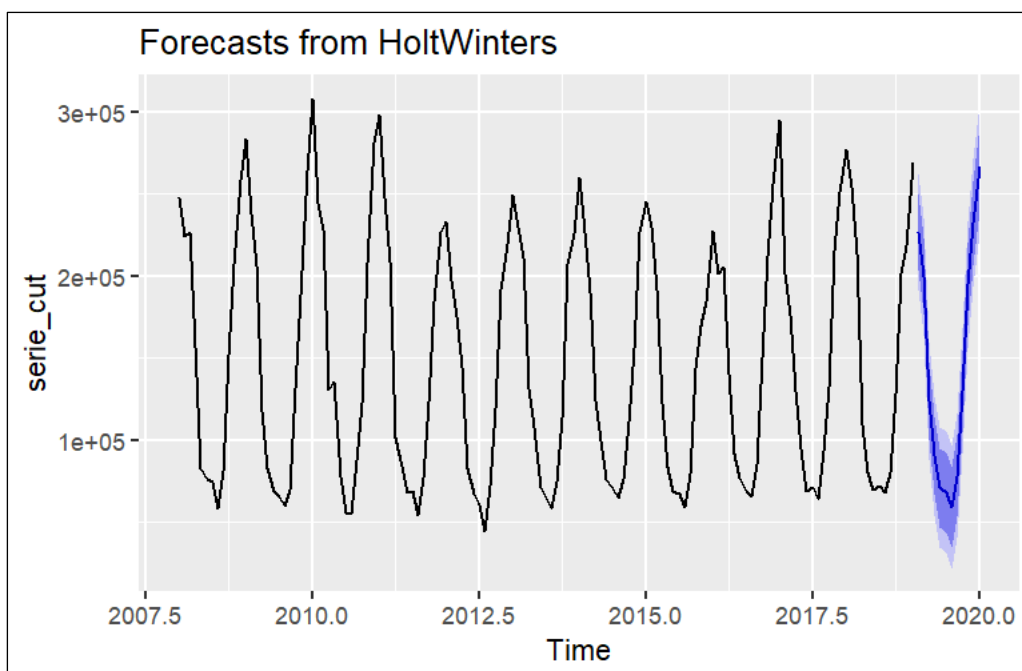
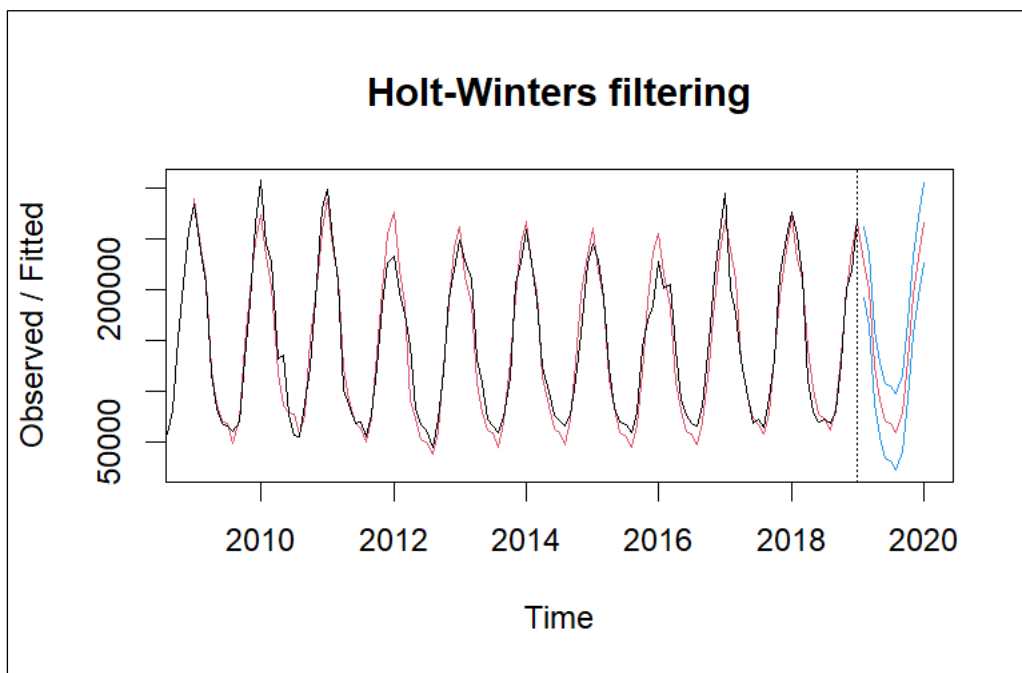
AIC      AICC      BIC
3208.813 3209.000 3217.485

Error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -171.7635 14786.57 10867.13 -1.159659 8.082301 0.6933652 0.2761884
```

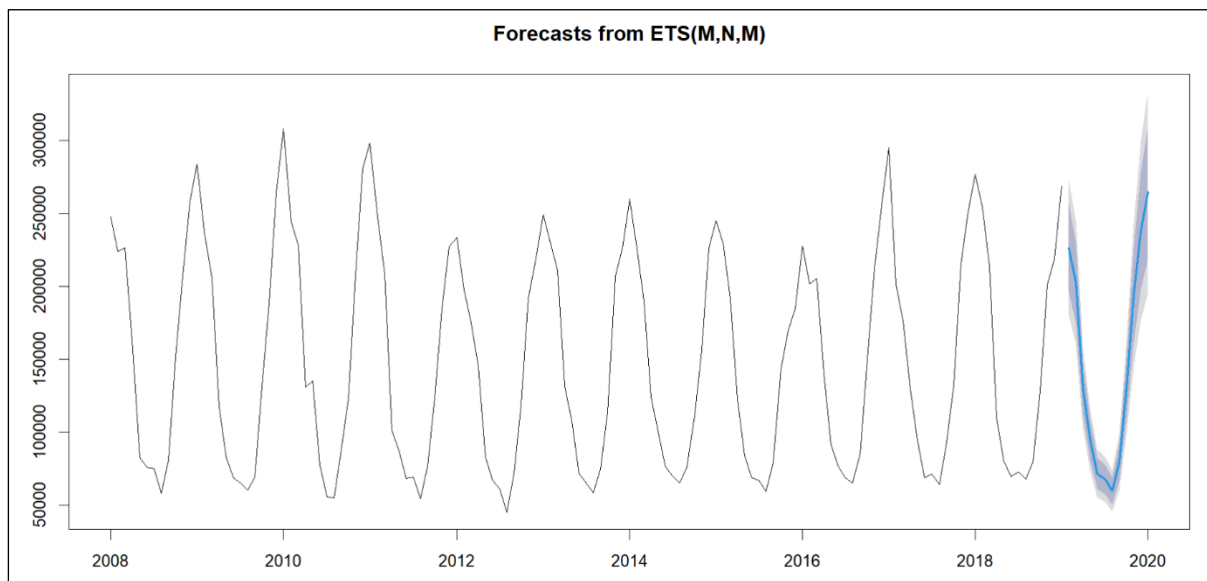
## ANNEXE 13 : Pr vision avec la m thode STS



## ANNEXE 14 : Prévisions avec la méthode Holt-Winters



## ANNEXE 15 : Prévision avec la méthode ETS



## ANNEXE 16 : Estimation avec la méthode ETS

ETS(M,N,M)

Call:

```
ets(y = serie_cut)
```

Smoothing parameters:

alpha = 0.2423

gamma = 1e-04

Initial states:

l = 149414.3272

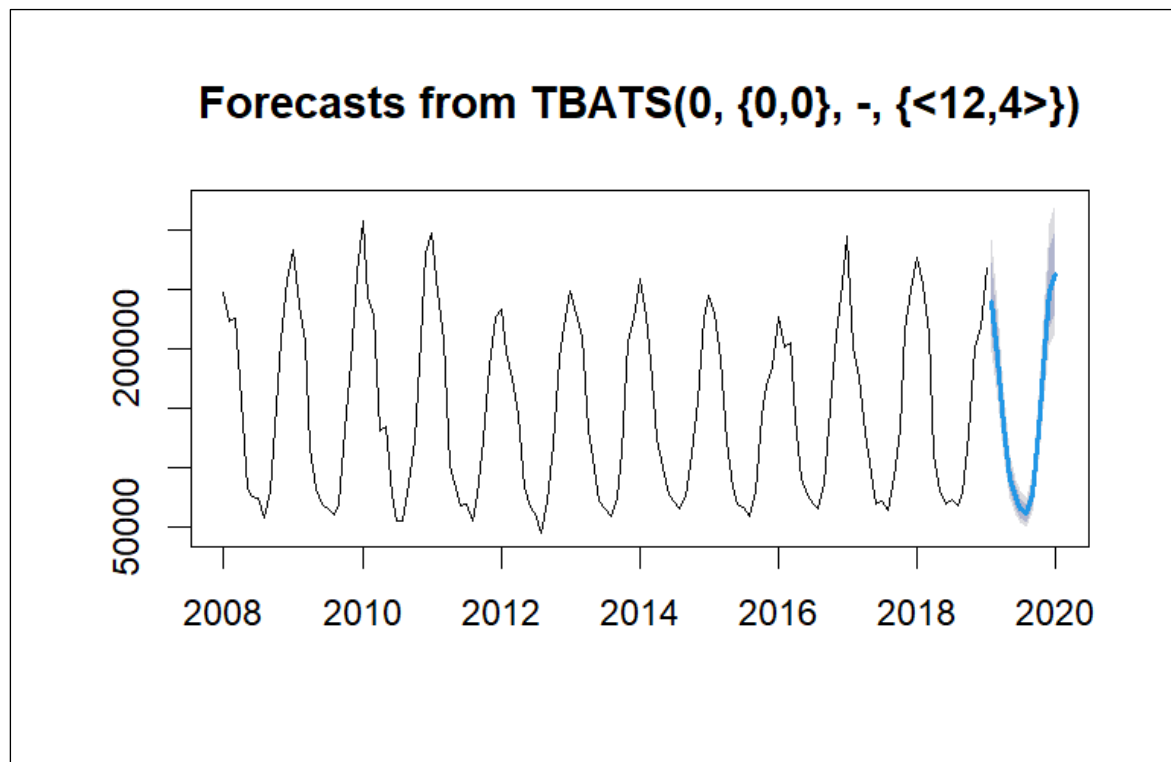
s = 1.6241 1.3223 0.8916 0.5439 0.4064 0.461  
0.4879 0.6457 0.8856 1.3817 1.5448 1.8051

sigma: 0.1035

AIC	AICc	BIC
3192.504	3196.607	3235.859



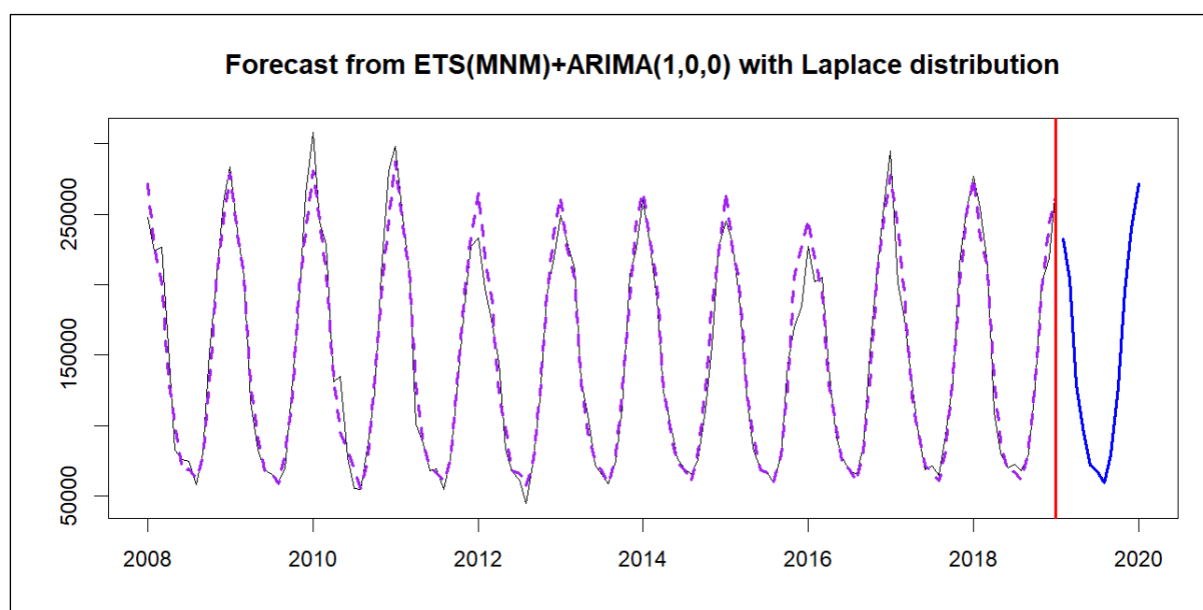
## ANNEXE 17 : Prévision avec la méthode TBATS



## ANNEXE 18 : Résultats de la prévision avec TBATS

```
TBATS(0, {0,0}, -, {<12,4>})  
  
Call: tbats(y = serie_cut)  
  
Parameters  
  Lambda: 0  
  Alpha: 0.07851826  
  Gamma-1 Values: 0.007687197  
  Gamma-2 Values: 0.01780079  
  
Seed States:  
      [,1]  
[1,] 11.791191540  
[2,]  0.735299119  
[3,] -0.006697894  
[4,] -0.050829162  
[5,]  0.023286971  
[6,]  0.038047061  
[7,] -0.050499211  
[8,]  0.027450712  
[9,] -0.018356212  
attr("lambda")  
[1] 6.077862e-09  
  
Sigma: 0.1001497  
AIC: 3193.354
```

## ANNEXE 19 : Prévision avec la méthode ADAM ETS



## ANNEXE 20 : Résultats de la prévision avec la méthode ADAM ETS

```

Model estimated using auto.adam() function: ETS(MNM)+ARIMA(1,0,0)
With optimal initialisation
Distribution assumed in the model: Laplace
Loss function type: likelihood; Loss function value: 1435.953
Persistence vector g:
  alpha  gamma
0.0000  0.0034

ARMA parameters of the model:
  Lag 1
AR(1)  0.3693

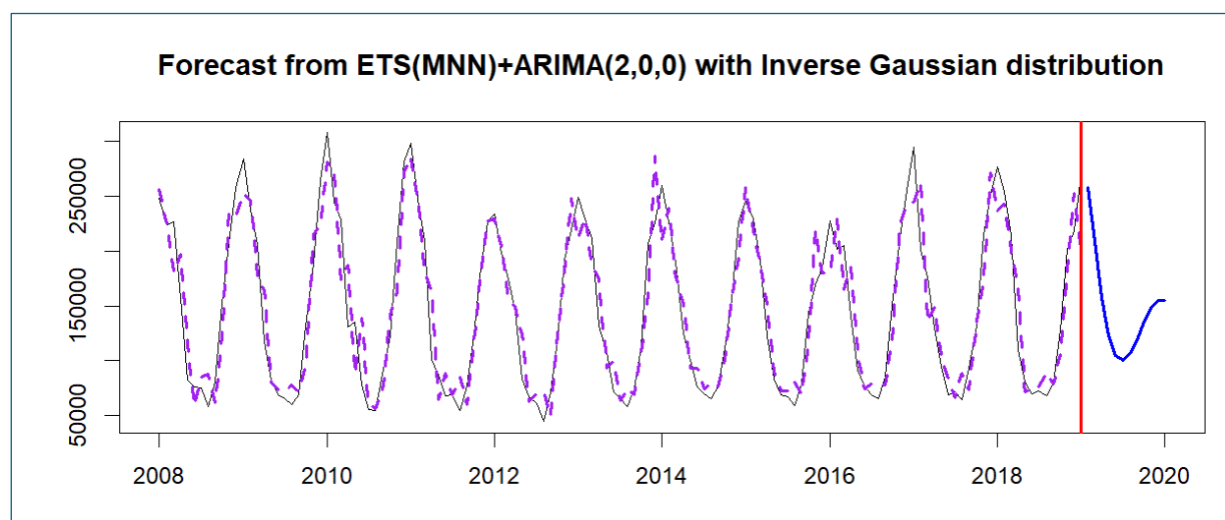
Sample size: 133
Number of estimated parameters: 17
Number of degrees of freedom: 116
Information criteria:
      AIC      AICc      BIC      BICc
2905.907 2911.228 2955.042 2968.055

```

## ANNEXE 21 : Valeurs des prévisions sur 12 mois avec la méthode ADAM ETS

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2019		232456.44	204236.63	129032.10	94471.76	71587.99	67847.39	59597.46
2020	271440.81							
	Sep	Oct	Nov	Dec				
2019	79346.54	127419.56	195585.78	240034.73				
2020								

## ANNEXE 22 : Prévision avec la méthode ADAM ETS ARIMA



## ANNEXE 23 : Résultats de la prévision avec la méthode ADAM ETS ARIMA

```

Model estimated using auto.adam() function: ETS(MNN)+ARIMA(2,0,0)
With optimal initialisation
Distribution assumed in the model: Inverse Gaussian
Loss function type: likelihood; Loss function value: 1538.813
Persistence vector g:
alpha
1e-04

ARMA parameters of the model:
      Lag 1
AR(1)  1.4574
AR(2) -0.7445

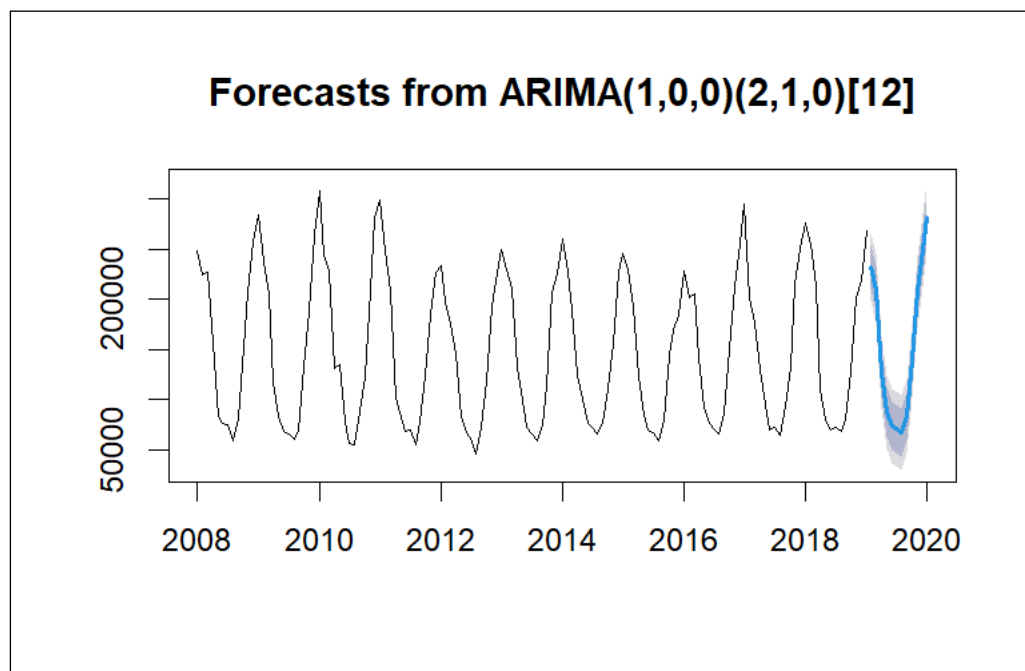
Sample size: 133
Number of estimated parameters: 7
Number of degrees of freedom: 126
Information criteria:
      AIC      AICc      BIC      BICc
3091.625 3092.521 3111.858 3114.048

```

## ANNEXE 24 : Valeurs des prévisions sur 12 mois avec la méthode ADAM ETS ARIMA

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
2019		258017.3	208350.8	157315.4	122478.4	104825.3	100667.4	106559.9	119312.4
2020	154740.7								
	Oct	Nov	Dec						
2019	134846.6	148170.1	155176.4						
2020									

## ANNEXE 25 : Pr vision avec la m thode SARIMA



## ANNEXE 26 : R gression de la m thode SARIMA

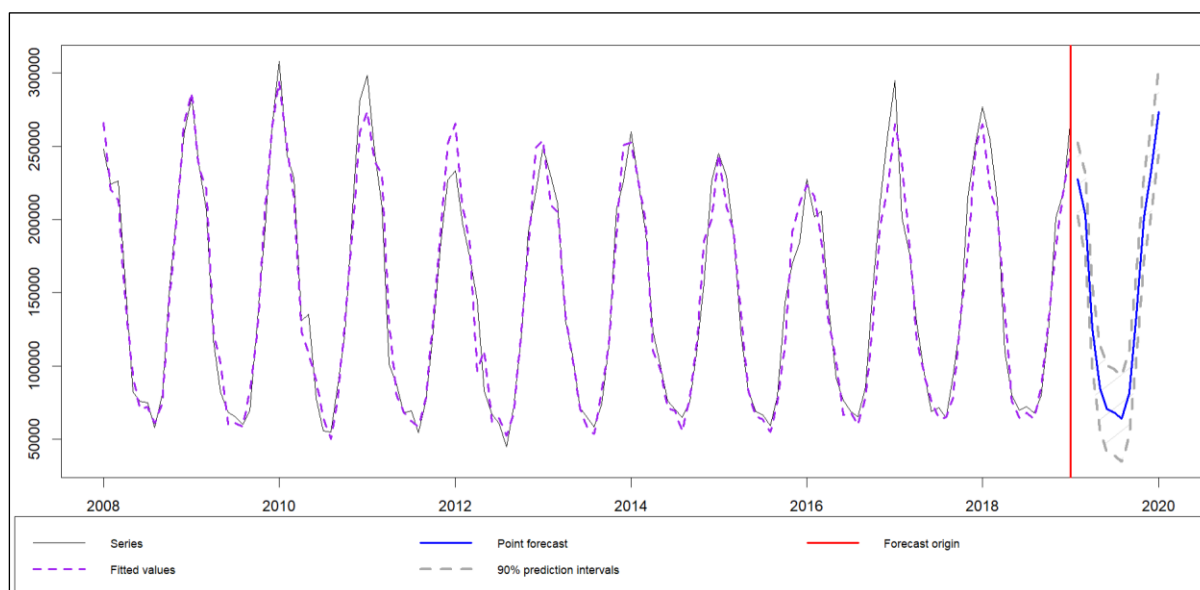
```
Series: serie_cut
ARIMA(1,0,0)(2,1,0)[12]

Coefficients:
      ar1      sar1      sar2
      0.4638  -0.5400  -0.4611
s.e.    0.0828   0.0889   0.0862

sigma^2 = 280858282:  log likelihood = -1350.97
AIC=2709.94  AICc=2710.29  BIC=2721.13

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -283.9782 15785.52 11266.24 -0.7774325 8.058478 0.7188298 0.04391155
```

## ANNEXE 27 : Pr vision avec la m thode SSARIMA

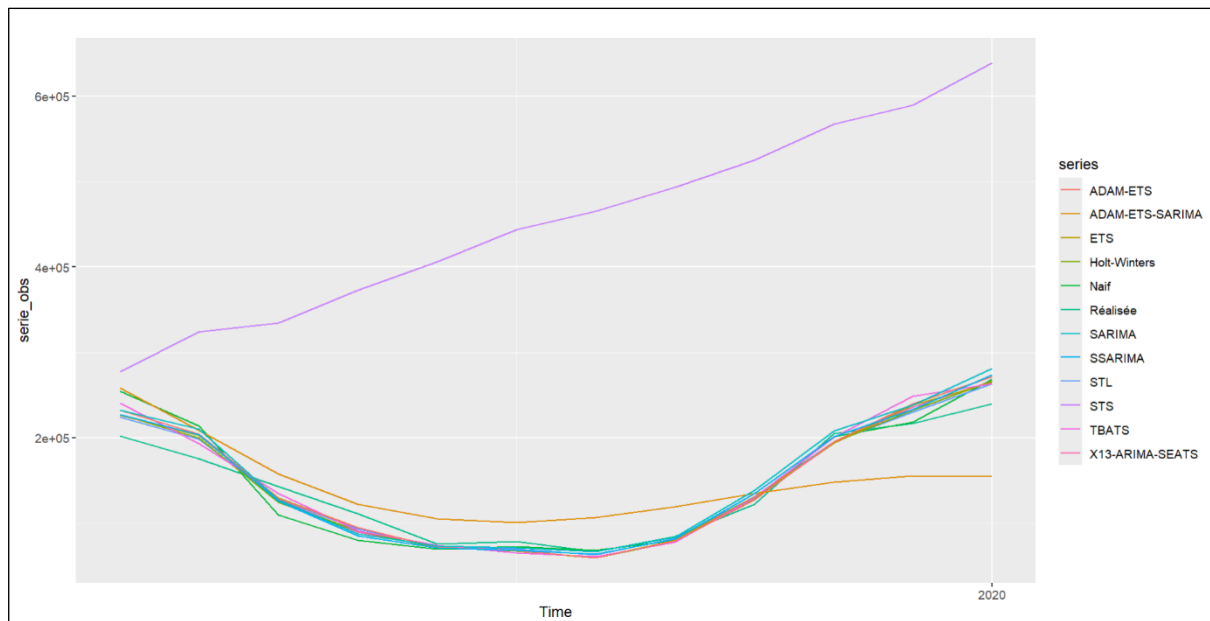


## ANNEXE 28 : Estimation du mod le SSARIMA

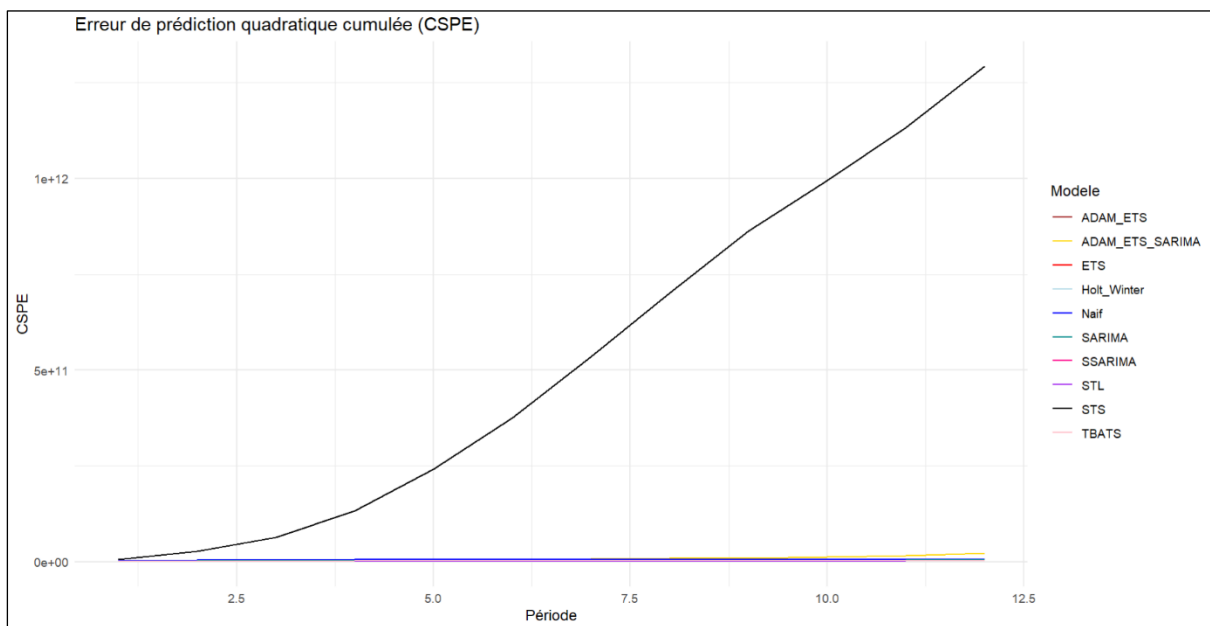
```
Model estimated: SARIMA(2,0,3)[1](3,0,0)[12]
Matrix of AR terms:
      Lag 1 Lag 12
AR(1) -0.1391 0.3759
AR(2)  0.1297 0.1133
AR(3)  0.0000 0.4780
Matrix of MA terms:
      Lag 1
MA(1) 0.7063
MA(2) 0.1982
MA(3) 0.0722
Initial values were produced using backcasting.

Loss function type: likelihood; Loss function value: 1467.0596
Error standard deviation: 15469.4
Sample size: 133
Number of estimated parameters: 9
Number of degrees of freedom: 124
Information criteria:
      AIC      AICc      BIC      BICc
2952.119 2953.583 2978.132 2981.711
```

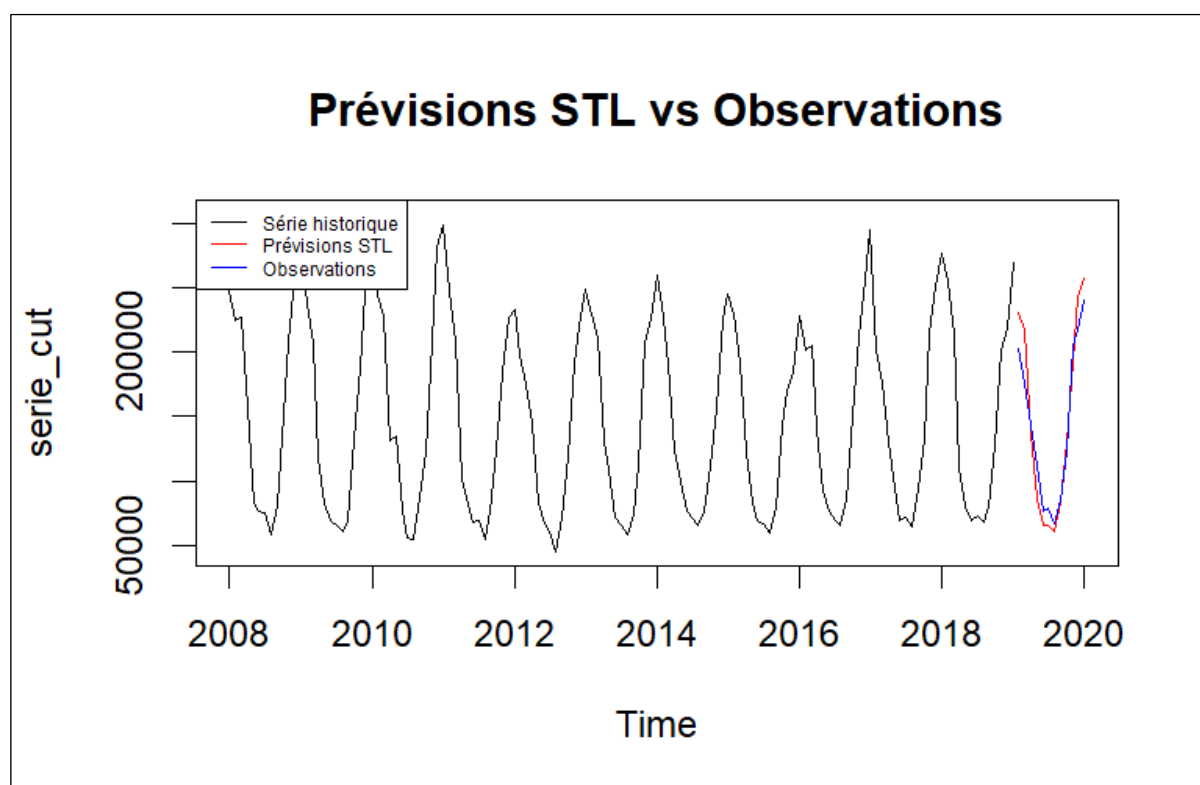
## ANNEXE 29 : Comparaison des différents modèles avec STS



## ANNEXE 30 : Erreur de prédiction quadratique cumulée avec STS



# ANNEXE 31 : Prédiction du meilleur modèle STL avec un pas de 1 mois en rolling windows



# ANNEXE 32 : Valeurs des prévisions en fenêtre roulante pour les 12 mois

	Mois	Prevision	Observation	Erreur
1	févr. 2019	230456.36	202359.0	28097.3609
2	mars 2019	217255.50	175296.0	41959.4969
3	avr. 2019	136818.75	142662.0	-5843.2472
4	mai 2019	86005.36	111302.0	-25296.6418
5	juin 2019	65171.91	75851.0	-10679.0899
6	juil. 2019	65569.81	78875.0	-13305.1855
7	août 2019	61078.19	66801.0	-5722.8117
8	sept. 2019	83838.88	84344.0	-505.1244
9	oct. 2019	129977.26	122412.0	7565.2628
10	nov. 2019	192743.37	204802.0	-12058.6257
11	déc. 2019	242935.52	216863.0	26072.5163
12	janv. 2020	257566.05	239974.4	17591.6699

## Bibliographies

[Don't Make The same Mistake As Me! Almost-Returned\\_2 \(FR\)](#)

[\[nrg\\_cb\\_gasm\] Supply, transformation and consumption of gas - monthly data](#)

[Energy statistics - quantities \(nrg\\_quant\)](#)

[Base de données - Eurostat](#)

[Froid : la consommation de gaz atteint un nouveau sommet historique - ladepeche.fr](#)

[Hausse de la consommation des ménages en février : retour à la confiance ou conséquence de la vague de froid ? - FioulReduc](#)

[L'hiver 2013-2014 parmi les plus doux depuis 1900](#)

[Gaz : la France a beaucoup moins consommé en 2014](#)

[Bilan météo de janvier 2018](#)



# Table des matières

<b>Introduction.....</b>	<b>3</b>
<b>I) Analyse exploratoire .....</b>	<b>4</b>
a) Présentation et caractérisation de la série étudiée.....	4
b) Détection des points atypiques .....	6
c) Statistiques descriptives sur la série corrigée .....	7
d) Détection de la saisonnalité .....	9
e) Désaisonnalisation et décomposition .....	11
<b>II) Prévission de la série saisonnière corrigée des points atypiques sur une année 13</b>	
a) Estimation et prévisions des modèles.....	13
(i) Les méthodes Naïve, X13-SARIMA-SEATS, STL et STS.....	13
- Méthode Naïve .....	13
- X13-SARIMA-SEATS .....	13
- STL (Seasonal-Trend decomposition Loess) .....	13
- STS (Structural Time Series) .....	14
(ii) Les méthodes de lissage exponentiel : Holt-Winters, ETS, TBATS, ADAM ETS et ADAM ETS SARIMA .....	14
- Holt-Winters .....	14
- ETS (Error, Trend, Seasonality) .....	15
- TBATS (Trigonometric, Box-Cox, ARMA errors, Trend and Seasonal).....	15
- ADAM ETS .....	16
- ADAM ETS SARIMA .....	16
(iii) les modèles SARIMA(p,d,q)(P,D,Q)[12] et SSARIMA. ....	17
- SARIMA(p,d,q)(P,D,Q)[12].....	17
- SSARIMA .....	17
b) Présentation sous forme de tableau des paramètres des modèles .....	18
<b>III) Représentation graphique de l'évolution des prévisions des différents modèles .....</b>	<b>18</b>
<b>IV) Qualité de prévision .....</b>	<b>20</b>
<b>V) Test de précision .....</b>	<b>22</b>

<b>VI) Pr�vision sur une ann�e avec un pas de 1 mois avec le meilleur mod�le ...</b>	<b>23</b>
<b>Conclusion .....</b>	<b>25</b>
<b>Annexes .....</b>	<b>26</b>
<b>Bibliographies.....</b>	<b>40</b>
<b>Table des mati�res .....</b>	<b>41</b>