



Aymeric MOUTTAPA
Maximilien SOUMAHORO

BUT SD2 – FA VCOD



SAE

Description & Prévision

Série Temporelle



SOMMAIRE

- Introduction
- Analyse de tendance avec moyennes mobiles et régression
- Détermination et visualisation des coefficients saisonniers
- Évaluation de la série temporelle désaisonnalisée
- Prévisions de production d'électricité pour 2023
- Comparaison des méthodes de prévision et calcul d'erreur pour 2022
- Conclusion
- Summary

INTRODUCTION

Dans ce projet, nous nous intéressons à une analyse approfondie de la production d'électricité aux États-Unis, en mettant spécifiquement l'accent sur l'utilisation du Petrol Coke comme source d'énergie. Notre étude s'étend sur une période de 22 ans, de 2001 à 2022, une période significative pour observer les tendances et les variations dans l'utilisation de cette ressource énergétique.

Le Petrol Coke, un sous-produit de la transformation du pétrole, est un acteur clé dans le paysage énergétique américain. Il représente une source d'énergie à la fois controversée et cruciale, notamment en raison de ses impacts environnementaux et de son efficacité énergétique. Notre analyse vise à dégager des tendances claires quant à l'évolution de son utilisation pour la production d'électricité, en utilisant des séries temporelles fournies par l'Agence américaine d'information sur l'énergie (EIA).

Nous appliquerons diverses méthodologies statistiques, telles que le filtrage par moyennes mobiles, la régression, et l'analyse de décomposition saisonnière, pour étudier les tendances, les saisons et les prévisions futures concernant l'utilisation du Petrol Coke. Cela nous permettra non seulement de comprendre son rôle passé et présent dans le secteur de l'énergie, mais aussi d'anticiper son utilisation future.

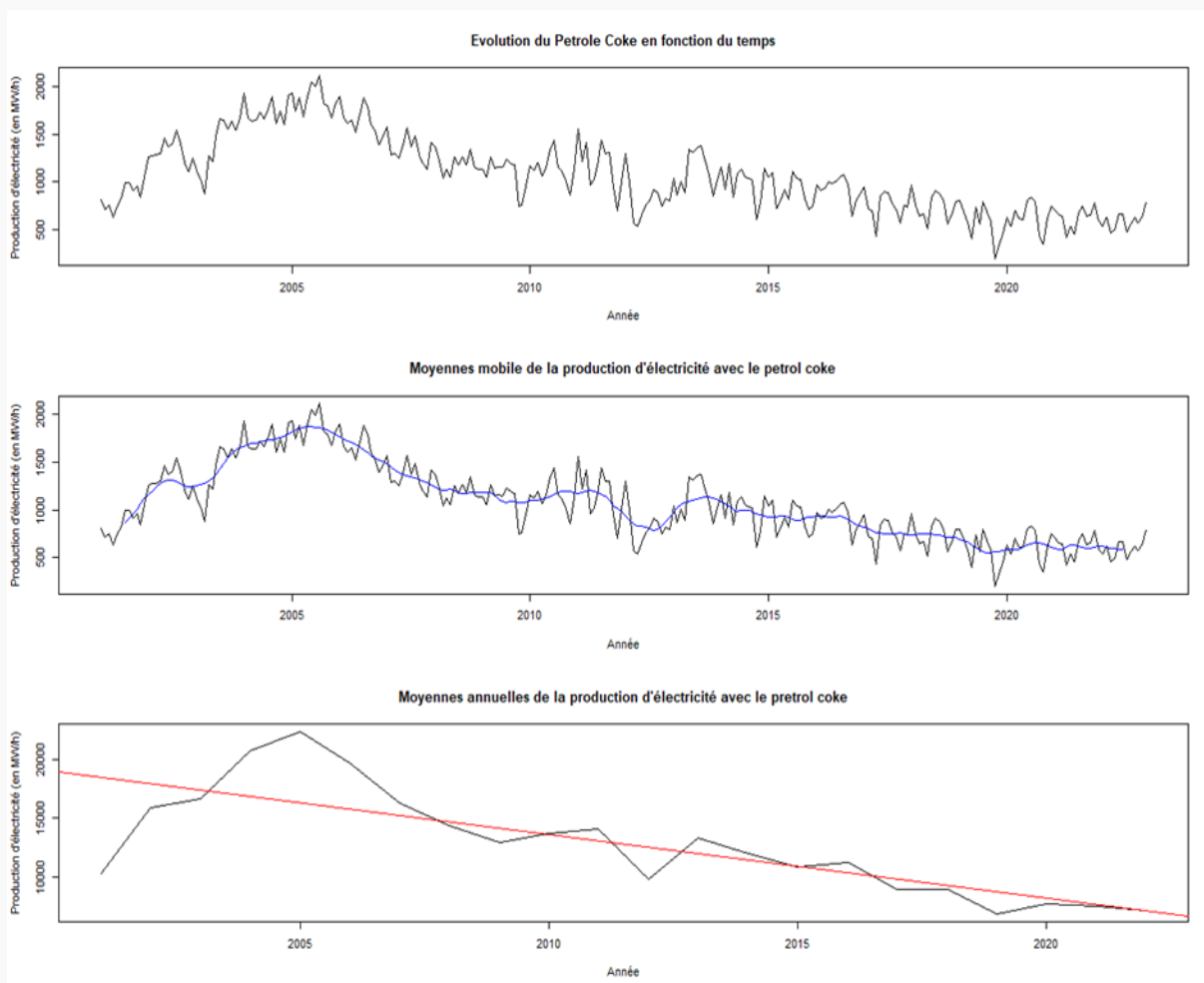
Ce rapport synthétise nos découvertes et analyses, offrant ainsi une vue complète de l'impact et de l'évolution du Petrol Coke dans le secteur énergétique américain sur une période significative. Notre objectif est de fournir une analyse détaillée et nuancée, qui pourra servir de base à des décisions éclairées en matière de politique énergétique et de pratiques de durabilité.

Afin d'arriver à nos fins, nous utiliserons le logiciel R, avec les packages ggplot2, qui sera utile pour la création de graphiques, et forecast, qui lui sera utile pour analyser et prédire les données.

TENDANCES

Dans le cadre de l'analyse de la production d'électricité à partir du pétrole coke, plusieurs fonctions R ont été mises en œuvre pour traiter et visualiser les données. La série temporelle a été créée avec `ts()`, nous permettant de travailler avec les données dans une séquence chronologique appropriée. Pour examiner les tendances générales, les moyennes annuelles ont été calculées avec `aggregate()`, simplifiant les variations pour mieux percevoir les changements sur le long terme.

Ensuite, `ma()` a produit une moyenne mobile, affichée dans le troisième graphique, lissant les fluctuations à court terme pour mettre en évidence la tendance sous-jacente. Enfin, la ligne de tendance dans le troisième graphique a été ajoutée avec `abline()`, suite à une régression linéaire effectuée avec `lm()`, pour déterminer la stabilité ou la variation de la production au fil du temps.



Le premier graphique montre les données brutes de production d'électricité à partir du Petrol Coke sur les 20 dernières années en mois. Il est essentiel de regarder les hauts et les bas. Si les pics et les creux semblent se répéter chaque année, cela indique une forte composante saisonnière. La présence ou l'absence d'une tendance à la hausse ou à la baisse générale est également importante. Ici, la tendance générale à la baisse indiquerait une baisse de l'utilisation du Petrol Coke au fil du temps.

Le deuxième graphique utilise une moyenne mobile sur 12 mois pour lisser les fluctuations saisonnières et mettre en évidence les tendances à court et moyen terme. La droite de régression linéaire (bleue) appliquée à la moyenne mobile donne un aperçu de la tendance générale sur une période glissante de 12 mois. La pente descendante de cette ligne donne une indication de l'évolution de la tendance sur une base plus régulière et lissée par rapport aux variations mensuelles.

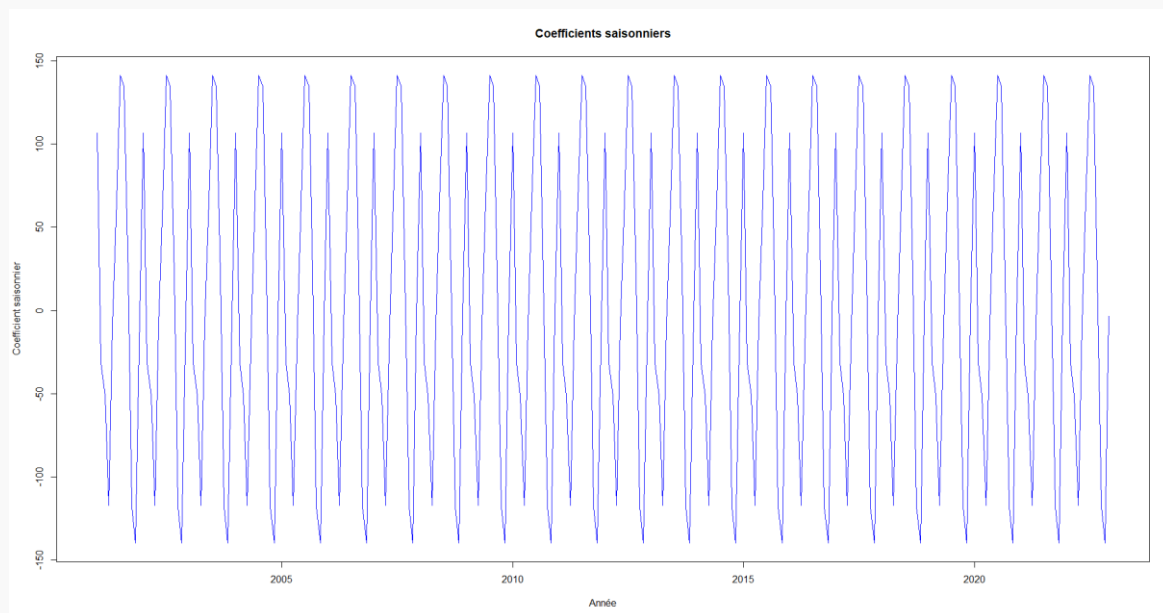
Le troisième graphique simplifie les deux derniers, ce qui aide à visualiser la tendance globale sur plusieurs années. La droite de régression linéaire (rouge) fournit une indication visuelle de cette tendance. La pente descendante de la droite confirme la baisse annuelle moyenne de la production. L'angle et la direction de cette ligne sont cruciaux pour comprendre comment la production du Petrol Coke a évolué au fil des années.

En résumé, l'analyse des tendances dans ces graphiques permet de comprendre comment la production d'électricité à partir du Petrol Coke a varié au fil du temps, à la fois en termes de fluctuations saisonnières et de tendances à long terme. Cela fournit des informations précieuses sur l'évolution de l'utilisation de cette ressource.

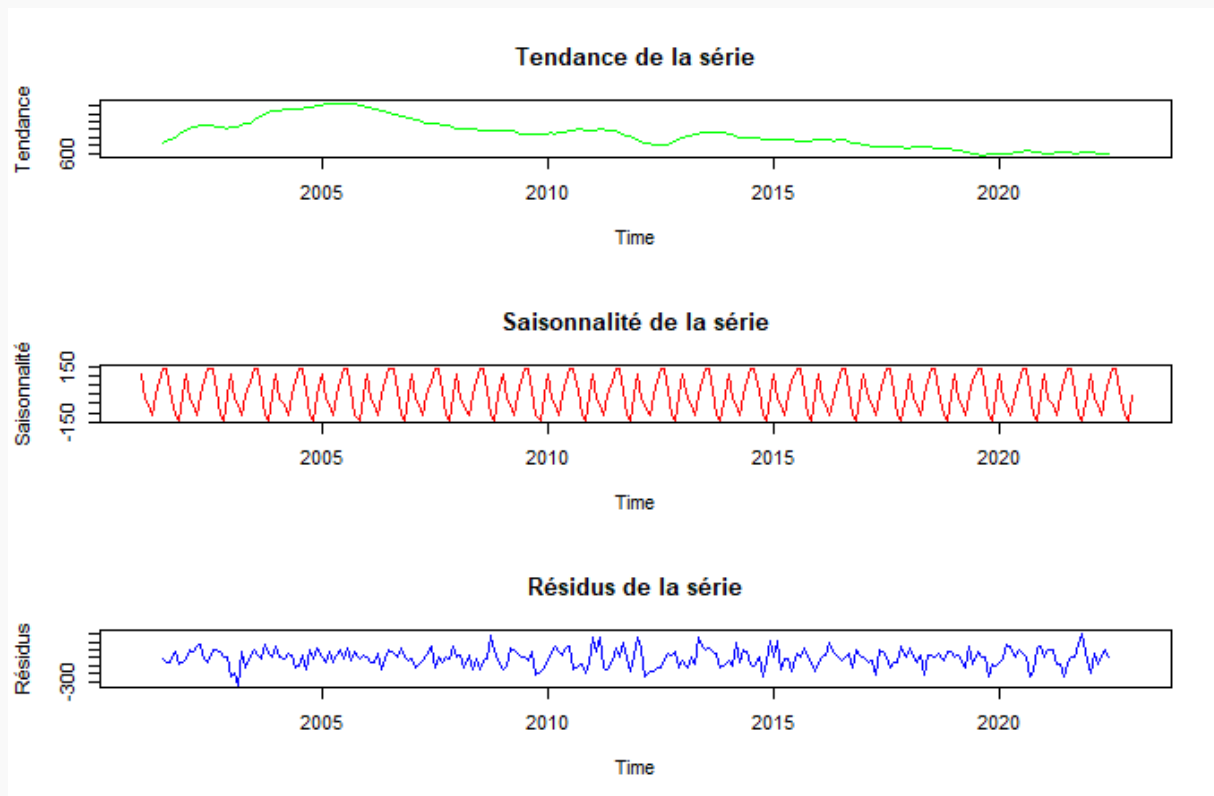
COEFFICIENTS SAISONNIERS

Les coefficients saisonniers mesurent les variations périodiques, comme celles observées d'une saison à l'autre dans une série temporelle. Ils quantifient ces fluctuations spécifiques à chaque période, aidant à distinguer les variations régulières de la tendance générale.

La fonction `decompose()` dans R est utilisée pour décomposer une série temporelle en ses composantes saisonnière, tendancielle et irrégulière (ou résiduelle). Ce processus est particulièrement utile pour analyser les séries temporelles qui présentent des motifs saisonniers clairs, comme dans le cas de données économiques, météorologiques ou de consommation.



Le graphique montre les coefficients saisonniers d'une série temporelle, indiquant une saisonnalité claire et constante au fil du temps. Chaque pic et creux récurrent représente un cycle saisonnier reflétant une périodicité annuelle. La stabilité de l'amplitude des cycles suggère que l'effet saisonnier reste uniforme d'année en année, sans tendance croissante ou décroissante notable dans les coefficients saisonniers. Cela impliquerait que les facteurs saisonniers qui influencent les données sont restés cohérents tout au long de la période observée.



Tendance de la série : est relativement plate et montre peu de variation au fil du temps. Cela suggère que la série a une tendance de long terme stable sans croissance ou décroissance significative.

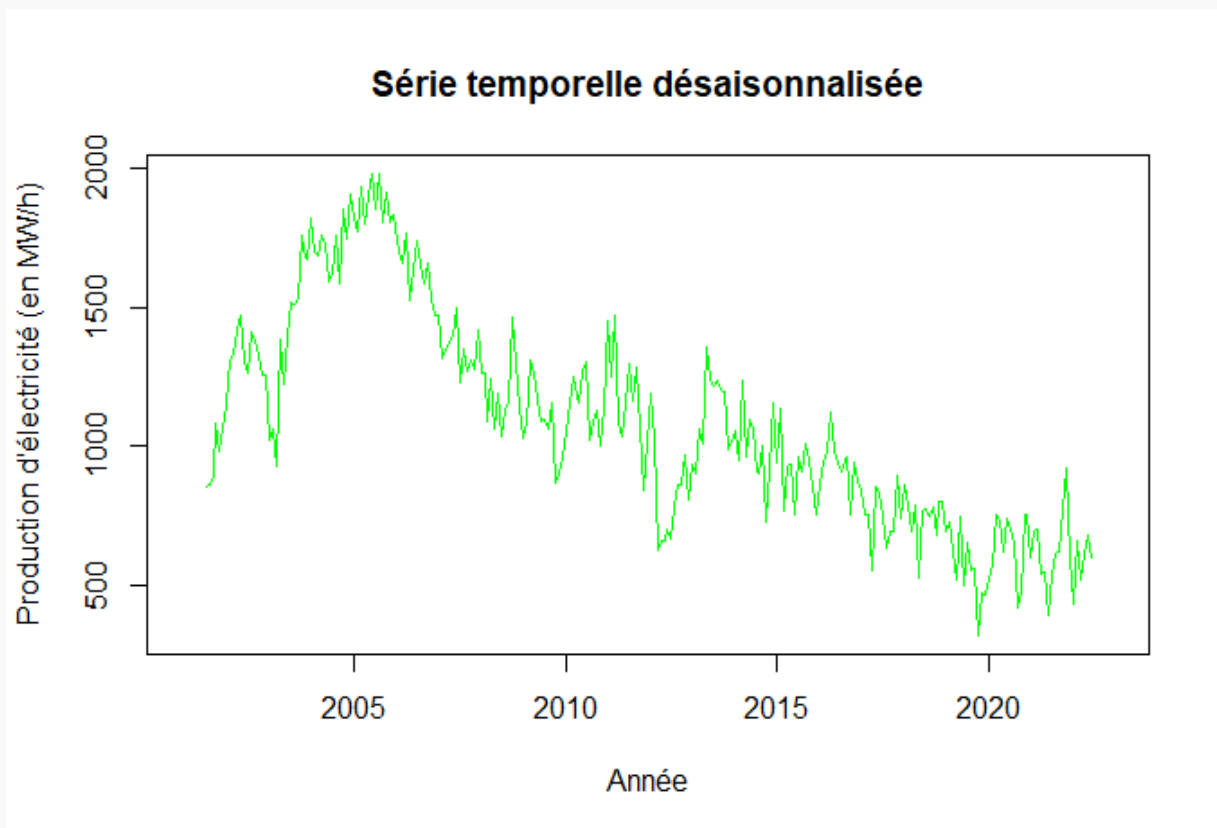
Saisonnalité de la série : des pics et des creux réguliers et symétriques qui se répètent avec une fréquence constante. Cela indique un motif saisonnier clair et bien défini, ce qui est typique pour les données ayant des influences saisonnières, comme la demande d'énergie ou les ventes au détail.

Résidus de la série : Le troisième graphique montre les résidus, c'est-à-dire les variations de la série temporelle qui ne sont ni expliquées par la tendance ni par la saisonnalité. Les fluctuations semblent aléatoires et ne présentent aucun motif systématique ou périodique discernable, ce qui indique que la décomposition a capturé avec succès la majorité de la structure systématique de la série dans les composantes de tendance et de saisonnalité.

SERIE DESAISONNALISEE

Une série temporelle désaisonnalisée est une série où les fluctuations saisonnières prévisibles ont été ajustées ou retirées. Cela permet de mettre en évidence la tendance sous-jacente ou les variations non saisonnières, en éliminant les effets récurrents associés à des périodes spécifiques de l'année.

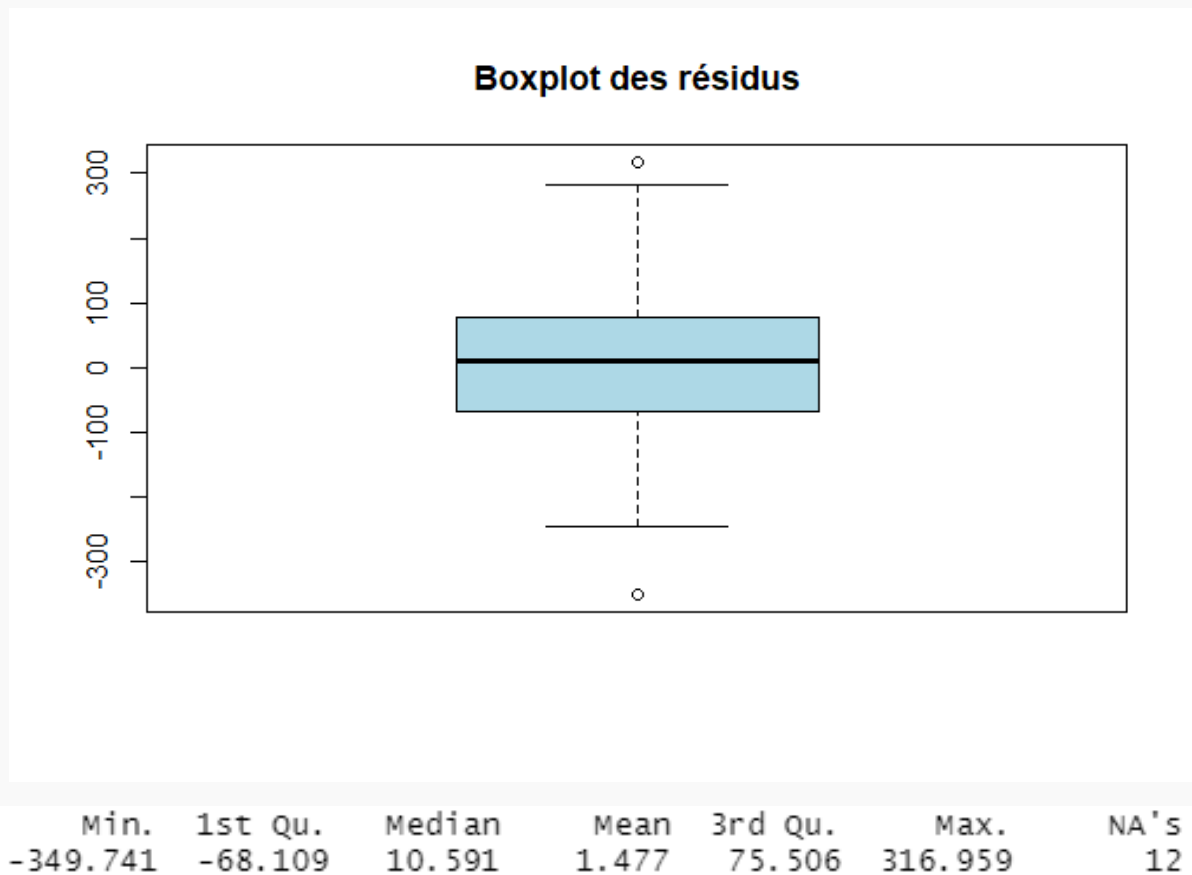
Code : `serie_desaisonnalisee <- decomposition$trend + residus`



On observe une tendance décroissante au fil du temps, avec la production d'électricité atteignant un pic autour de l'année 2005, suivie d'une tendance générale à la baisse.

Vers la fin de la série, autour de 2020, la production semble se stabiliser un peu, avec moins de baisse prononcée que dans les années précédentes.

La fonction `boxplot()` est utile pour la création d'une boîte à moustache et `summary()` permet d'avoir un résumé des données (avec min, max, quartiles).



La boîte à moustache montre que la distribution des résidus n'est pas parfaitement symétrique, la médiane étant légèrement décalée vers le haut de la boîte. Cependant, le fait que la moyenne soit très proche de zéro montre néanmoins une certaine symétrie du modèle. Il est important de noter les points atypiques, car ils peuvent indiquer des points de données qui ne correspondent pas bien au modèle.

PREVISION 2023

La méthode de prédiction naïve anticipe la prochaine valeur d'une série temporelle en se basant sur la dernière observation, sans tenir compte des tendances ou schémas plus complexes. C'est une approche rapide mais peut manquer de précision pour des séries temporelles complexes.

La méthode de Holt-Winters est une technique de prévision pour les séries temporelles qui inclut la tendance, la saisonnalité et les fluctuations. Elle s'appuie sur les données passées pour estimer les tendances futures, ainsi que les modèles saisonniers, aidant ainsi à prédire les valeurs à venir en tenant compte des schémas saisonniers observés.

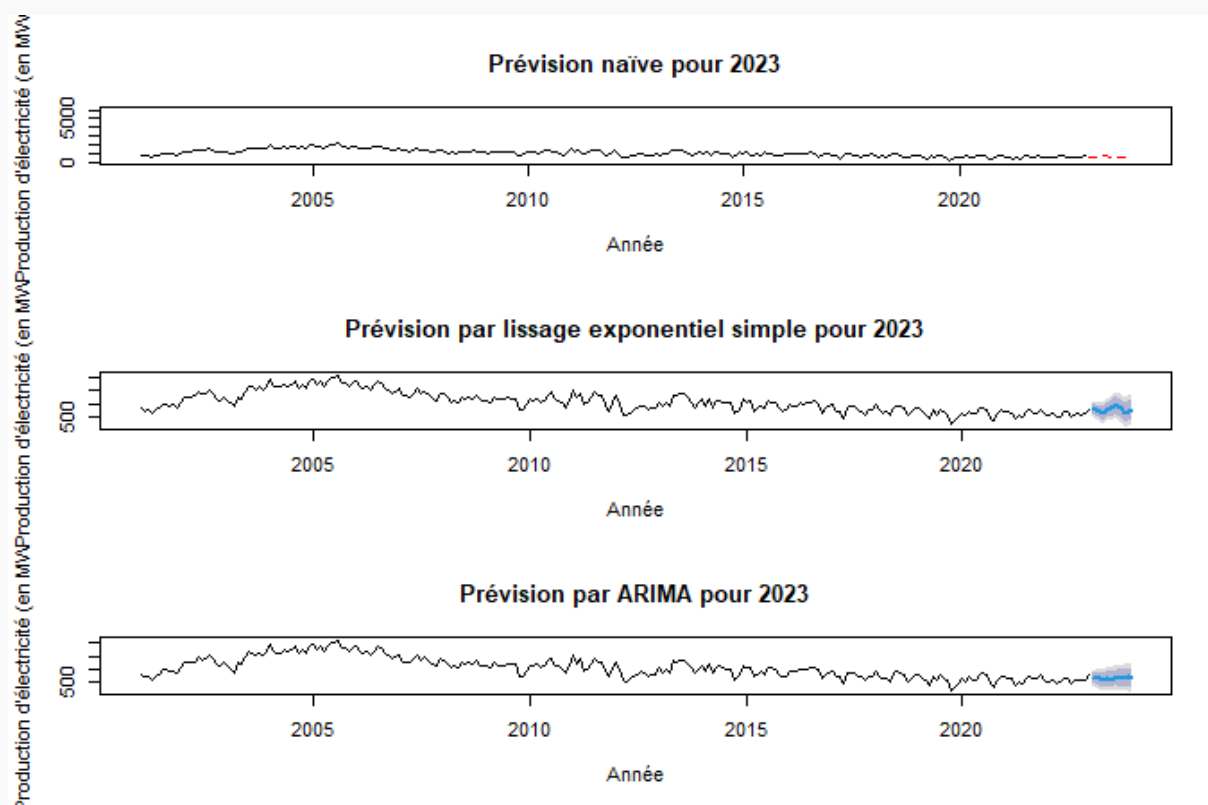
ARIMA est une méthode de prévision des séries temporelles qui intègre les tendances et les schémas passés pour prédire les valeurs futures. Elle utilise l'historique des données pour estimer les changements et les modèles temporels, aidant ainsi à anticiper les évolutions à venir.

Voici les codes nous permettant de tester les méthodes ci-dessus :

`window()` : Permet de sélectionner et d'isoler une période spécifique au sein d'une série temporelle pour une analyse ciblée.

`HoltWinters()` : Applique le lissage exponentiel pour modéliser et prévoir des séries temporelles avec tendance et saisonnalité.

`auto.arima()` : Automatise la sélection du modèle ARIMA le plus adapté pour la modélisation et la prévision de séries temporelles, en évaluant plusieurs combinaisons de paramètres.



Prévision naïve pour 2023: Ici, la dernière valeur observée est simplement projetée dans le futur. Il n'y a pas de changement attendu dans la variable de série temporelle, ce qui se traduit par une ligne horizontale.

Prévision par lissage exponentiel simple pour 2023: La prévision semble réagir aux fluctuations passées des données, suggérant une certaine adaptation au niveau de la série temporelle, mais sans prendre en compte de tendance ou de saisonnalité. La zone ombrée représente l'intervalle de confiance, indiquant le degré d'incertitude associé aux prévisions.

Prévision par ARIMA pour 2023: Semblable au lissage exponentiel, une zone ombrée indique l'intervalle de confiance pour les prévisions. Les prévisions ARIMA semblent suivre les motifs passés des données en ajustant les tendances et les effets saisonniers.

En résumé, les trois graphiques illustrent différentes approches de prévision pour 2023, avec des degrés croissants de complexité dans les méthodes de modélisation, allant de l'approche la plus simple (naïve) à une approche statistiquement plus élaborée (ARIMA). Les trois prévoient une stabilisation en 2023.

DIFFERENTES METHODES & EQM

L'erreur quadratique moyenne (EQM) est une mesure qui calcule la moyenne des carrés des différences entre les prévisions et les valeurs réelles. Plus l'erreur quadratique moyenne est faible, plus les prévisions sont proches des valeurs réelles, indiquant ainsi une meilleure performance de la méthode de prévision utilisée.

```
eqm_naive <- mean((petrol_coke_2022 - prevision_naive_2022)^2) #Naïve  
eqm_holtwinters <- mean((petrol_coke_2022 - prevision_holtwinters_2022$mean)^2) #Lissage exponentiel simple  
eqm_arima <- mean((petrol_coke_2022 - prevision_arima_2022$mean)^2) #ARIMA
```

EQM pour la méthode naïve : 22601.67.

Cette méthode, qui consiste à prendre la dernière observation connue comme prévision pour la prochaine période, donne une EQM relativement élevée. Cela signifie que, bien que facile à calculer, elle n'est pas très précise pour les données en question, probablement parce que cette méthode ne tient pas compte des tendances ou des modèles saisonniers qui pourraient exister.

EQM pour la méthode de lissage exponentiel simple : 30461.76.

Le lissage exponentiel simple donne une EQM encore plus élevée, ce qui est quelque peu surprenant car cette méthode est censée réagir aux changements dans les données plus récentes.

EQM pour la méthode ARIMA : 18327.01.

La méthode ARIMA, qui est, comme vu précédemment, plus sophistiquée, obtient la meilleure (la plus basse) EQM parmi les trois méthodes. Cela indique que la méthode ARIMA est la plus précise pour les prévisions de cette série temporelle spécifique, probablement parce qu'elle a pu capturer efficacement la structure sous-jacente des données.

En conclusion, sur la base des EQM fournies, la méthode ARIMA serait à privilégier pour la prévision de cette série temporelle, car elle offre la plus grande précision par rapport aux méthodes naïve et de lissage exponentiel simple.

CONCLUSION

Ce projet a détaillé une analyse statistique approfondie de la production d'électricité issue du pétrole coke, révélant une tendance globale à la diminution sur la période étudiée.

L'utilisation de méthodes de moyennes mobiles et de régression a démontré une capacité à discerner les tendances à long terme, malgré les fluctuations inhérentes aux données saisonnières.

La décomposition de la série a clarifié les schémas saisonniers, et l'ajustement saisonnier a permis de mieux comprendre les variations résiduelles.

La prévision pour 2023, effectuée par trois méthodes distinctes, a mis en avant la supériorité de l'approche ARIMA en termes de précision prédictive, comme en témoignent les erreurs quadratiques moyennes les plus basses.

Cette conclusion met en perspective la transition vers des alternatives plus durables ou efficaces en matière de production d'énergie, et souligne l'importance de telles analyses pour guider les décisions stratégiques dans le domaine de l'énergie.

SUMMARY

This study presents a comprehensive statistical analysis of Petrol Coke's role in U.S. electricity production from 2001 to 2022.

Employing sophisticated data analysis techniques, the study reveals significant long-term trends and notable seasonal fluctuations.

The application of moving averages, linear regression, and seasonal decomposition methodologies offer deep insights into the data.

Key findings include the impact of Petrol Coke on environmental sustainability and its efficiency in energy production.

The use of R software and its specific packages, like ggplot2 and forecast, facilitates effective data visualization and prediction, enhancing the understanding of Petrol Coke's evolving role in the energy sector.

This analysis contributes to informed decisions in energy policy and sustainability initiatives.