

La production d'électricité par combustion du charbon

SAE « Description et Prédiction de séries temporelles » - Dedecker

Introduction

Nous sommes un groupe d'étudiants en Bachelor Universitaire de Technologie en Science de Données à l'IUT de Paris - Rives de Seine, composé d'Alexandre Chen, Melvyn Corvo et Ruth Jeyaranjan. Notre projet a pour objectif d'analyser une série temporelle spécifique et d'appliquer des méthodes de prévision afin de déterminer les tendances futures. La série temporelle que nous allons décomposer est la production d'électricité par la combustion du charbon, entre 2001 et 2021. Le charbon est l'une des sources les plus courantes d'énergie utilisées pour produire de l'électricité dans le monde, en raison de sa disponibilité et de son faible coût. Les données dont nous disposons proviennent de l'Agence américaine d'information sur l'énergie (EIA).

Nous allons réaliser une étude descriptive préparatoire pour mieux comprendre la série temporelle. Pour atteindre notre objectif, nous allons appliquer des méthodes de décomposition de série temporelle pour comprendre les composants saisonniers, les tendances à long terme et les irrégularités dans la série temporelle. Ensuite, nous utiliserons des méthodes de prévision telles que trend+season, Holt-Winters et ARMA. Nous allons également évaluer la fiabilité de chaque modèle de prévision en utilisant des mesures d'erreur de prévision et en comparant les résultats de chaque méthode.

Les descriptions et les prévisions ont été réalisées sur le logiciel R.

Sommaire

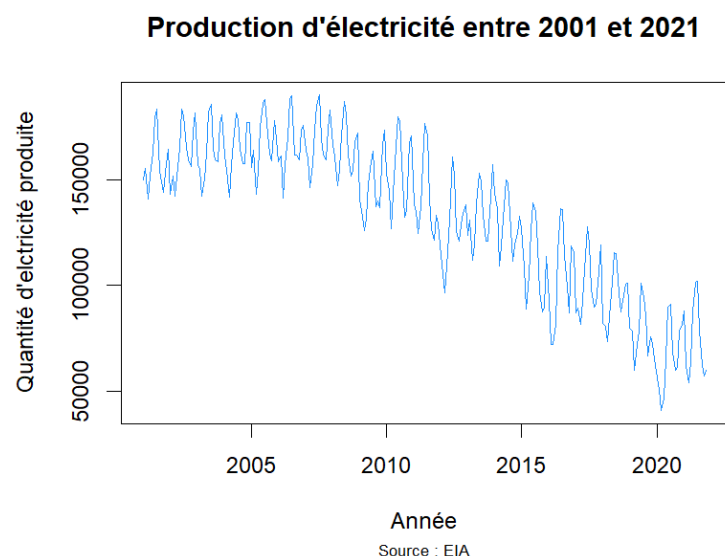
- Etude descriptive.....1-2
 - Courbe des données.....1
 - Modélisation de la série.....1-2
- Désaisonnalisation.....2-3
 - coefficients saisonniers.....2
 - décomposition de la série.....3
 - série corrigée des variations saisonnières.....3
- Prédiction de la production d'électricité en 2022 : 3 méthodes.....
 - estimation de tendance + coefficients saisonniers.....
 - prévisions de Holt-Winters.....

- modèles
ARMA.....
- Prédiction de la production d'électricité en 2021 : 3 méthodes.....
 - estimation de tendance + coefficients saisonniers.....
 - prévisions de
Holt-Winters.....
 - modèles
ARMA.....
- Comparaisons des modèles : Erreur quadratique moyenne

1. Etude descriptive de la série temporelle

Dans cette partie, nous allons explorer les caractéristiques temporelles afin de fournir des informations sur les tendances, les variations saisonnières et les fluctuations aléatoires de la série. Nous allons aussi identifier les points aberrants et les valeurs manquantes dans les données.

Il faut commencer par observer les données afin de pouvoir décomposer la série brute et choisir le modèle adapté. Ci-dessous se trouve un graphique linéaire représentant les données. Notre variable d'intérêt est la quantité d'électricité produite qui évolue entre 2001 et 2021. L'amplitude de la composante saisonnière et du bruit étant constante, on choisit le modèle additif.



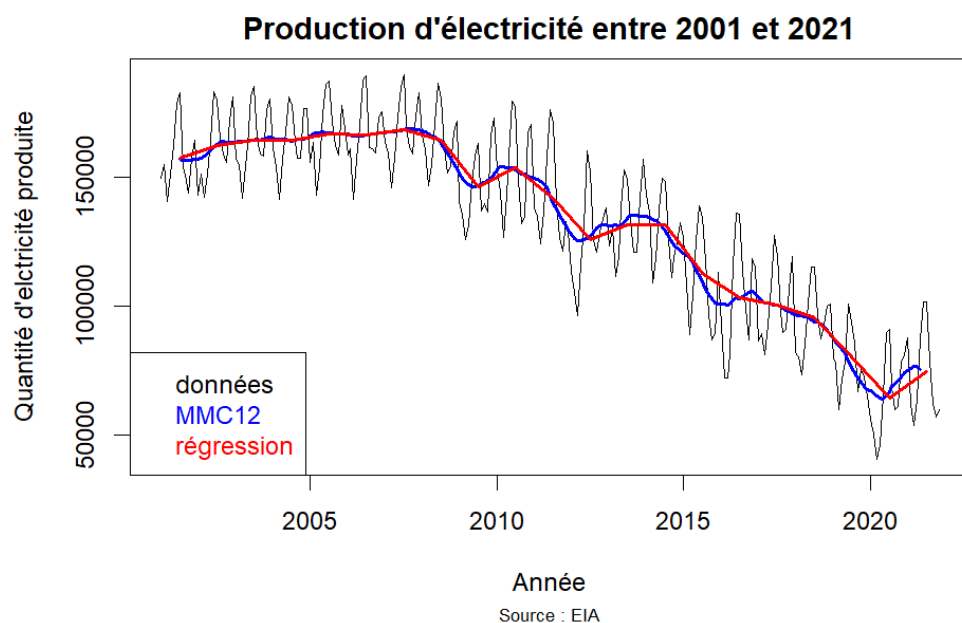
La crise financière de 2008 a pu avoir un impact sur la production d'électricité dans de nombreux pays, en particulier dans les économies avancées tel que les Etats-Unis. Les perturbations économiques liées à la crise ayant entraîné une diminution de la demande d'électricité, les consommateurs ont réduit leur consommation d'énergie. La crise a également eu un impact sur les investissements dans les infrastructures énergétiques qui a pu entraîner une réduction de la capacité de production d'électricité. En outre, bien qu'une grande partie de la production énergétique était issue de la combustion au charbon, les préoccupations environnementales ont commencé à croître. Et les méthodes non renouvelables ont commencé à laisser place à des sources d'énergie plus durables et respectueuses de l'environnement.

Modélisation de la série

A présent, nous devons modéliser la série pour comprendre sa structure. Pour modéliser la tendance de la série, nous avons deux méthodes : les moyennes mobiles et les moyennes annuelles.

Le graphique ci-dessous modélise la courbe des données avec les courbes de moyennes mobiles (en bleu) et de régression des moyennes annuelles (en rouge). La courbe bleu met en évidence la tendance de la série en utilisant un filtre de moyennes mobiles, qui consiste à faire des moyennes partielles de proche en proche, donc les fluctuations irrégulières chroniques sont atténuées. Donc, le filtre des moyennes mobiles va permettre de lisser la série en supprimant la composante saisonnière et révéler la tendance générale. Entre 2001 et 2008, la production d'électricité au charbon est restée relativement stable. Néanmoins, à partir de 2008, on observe une diminution de la production avec des fluctuations.

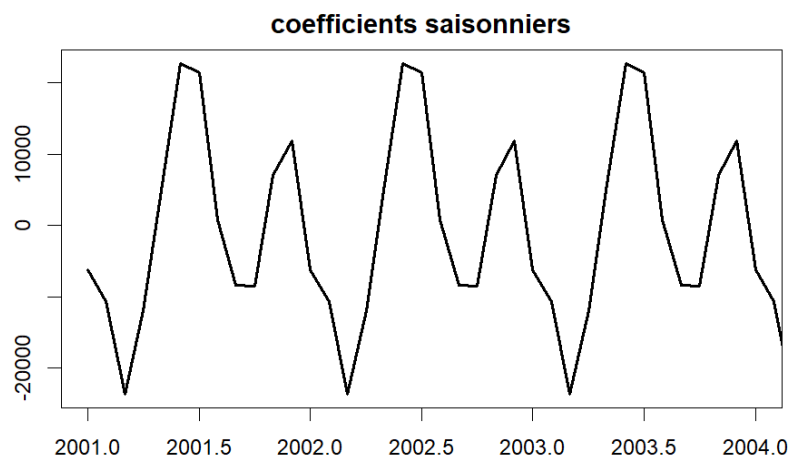
La modélisation par les moyennes annuelles de la production d'électricité au charbon représenté par la courbe rouge suit la même tendance. Cependant, la courbe est plus lisse. Comme son nom l'indique, cette méthode consiste à calculer la moyenne des données sur chaque année.



2. Désaisonnalisation

Calcul des coefficients saisonniers

D'après le graphique sur le calcul des coefficients saisonniers, nous pouvons voir que les coefficients varient au cours de l'année, on peut constater que la production d'électricité est bien plus supérieure en juin par rapport à la moyenne annuelle de la production.



A l'inverse, le mois de mars est celui où la production est la plus faible.

Tableau des 12 premières valeurs

jan.	fév.	mar.	avr.	mai	juin	jui.	aou.	sept.	oct.	nov.	déc.
-6219.45 92	-10710.0 008	-23713.4 175	-12033.2 300	5818.85 54	22788.8 927	21482.1 971	731.809 6	-8429.66 96	-8561.03 83	7003.29 71	11841.7 637

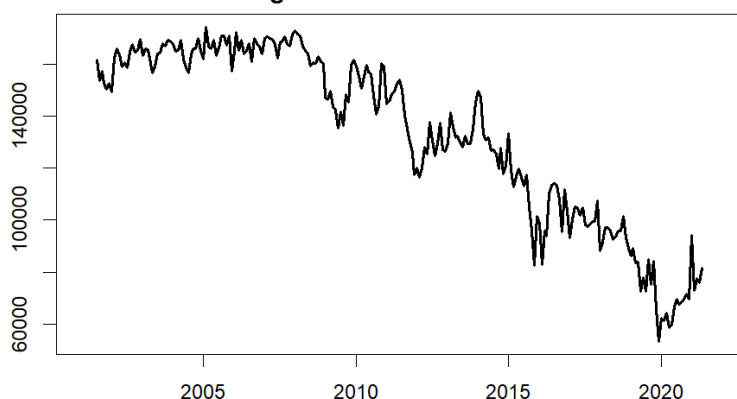
Décomposition de la série

La décomposition d'une série temporelle additive consiste à séparer les différentes composantes d'une série temporelle, notamment la tendance, la saisonnalité et le bruit, en les modélisant séparément. La tendance qui représente l'évolution générale de la série chronologique sur une longue période, est croissante entre 2001 et 2021. Cela indique une augmentation de la production d'électricité au fil du temps. Aussi, la série présente des fluctuations régulières sur tous les 2 ans. Les coefficients saisonniers sont utilisés pour modéliser la saisonnalité d'une série temporelle. + commentaires sur les résidus

Série corrigée des variations saisonnières (CVS)

En utilisant la SCVS, nous pouvons identifier les tendances à long terme et prévoir les tendances futures. En éliminant les fluctuations saisonnières, nous obtenons des données plus précises.

série corrigée des variations saisonnières

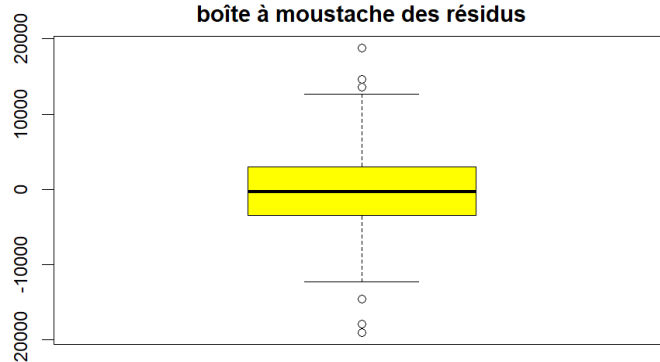


La série corrigée des variations saisonnières montre bien que la production d'électricité grâce au charbon diminue drastiquement, mais augmente pendant les mois d'hiver à cause du principalement au chauffage des ménages. Grâce à la série corrigée des variations saisonnières on obtient une vision plus claire de la tendance à long terme de la production d'électricité dû au charbon, en éliminant l'effet

saisonnier.

La boîte à moustache (à droite) représente la distribution des résidus de la série corrigée des variations saisonnières. La boîte est centrée autour de zéro et les moustaches et les valeurs aberrantes sont équilibrées de manière symétrique, cela indique une distribution normale et une bonne qualité de l'ajustement du modèle.

boîte à moustache des résidus



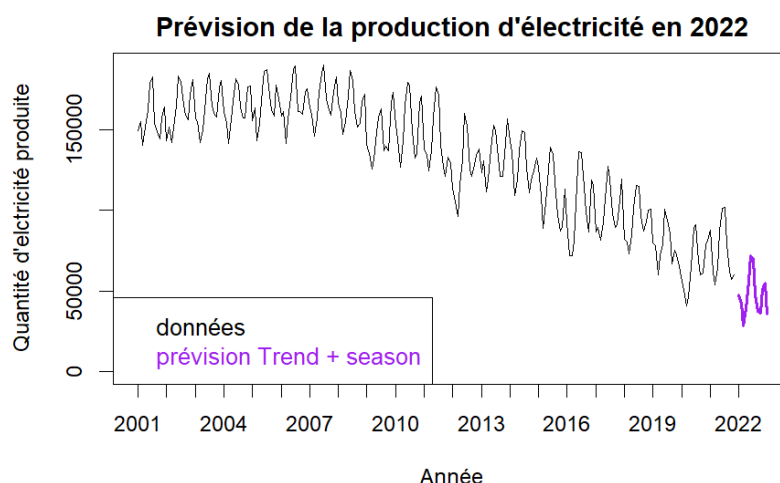
3. Prédiction de la production d'électricité en 2022 : 3 méthodes

Dans cette partie nous allons prédire la production d'électricité pour 2022, en utilisant trois méthodes différentes :

- estimation de tendance + coefficients saisonniers
- prévisions de Holt-Winters
- modèles ARMA : en prenant les coefficients saisonniers pour les 3 dernières années

Première méthode : estimation de tendance + coefficients saisonniers

La première méthode de prévision par estimation de tendance + coefficients saisonniers est constituée en 3 étapes : le calcul des coefficients saisonniers \hat{S}_i , la construction de la série corrigées des variations saisonnières et ajustement de la tendance des moindres carrés sur la CVS.

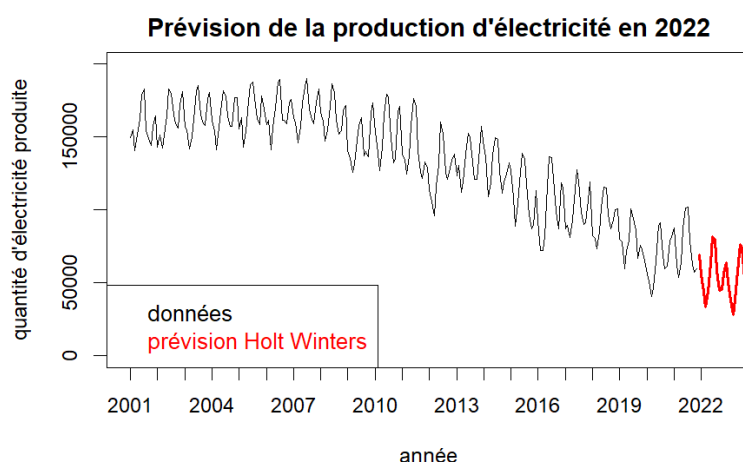


Sur cette courbe, on voit que la production prévue suit la tendance en diminuant. On retrouve également le motif saisonnier.

Cette méthode repose sur l'hypothèse que toute série chronologique peut être décomposée en une tendance, des effets saisonniers et des résidus.

Deuxième méthode : les prévisions de Holt-Winters

Le principe de cette méthode suppose que l'on peut prévoir la quantité d'électricité produite avec les observations antérieures en affectant un poids qui décroît exponentiellement avec la distance à l'indice. De même, pour le comportement saisonnier. Par ailleurs, cette méthode étant récursive, lorsque de nouvelles observations sont soumises, les paramètres se mettent à jour sans recommencer l'optimisation.



Le graphique montre les données historiques de la production d'électricité jusqu'à fin 2021, ainsi que les prévisions de Holt-Winters pour l'année prochaine, représentées par la ligne rouge. On peut

voir que les prévisions de Holt-Winters pour la production d'électricité sont à la baisse pour les années à venir avec une périodicité similaire aux années précédentes. Les prévisions de Holt-Winters suggèrent donc une tendance décroissante à l'avenir.

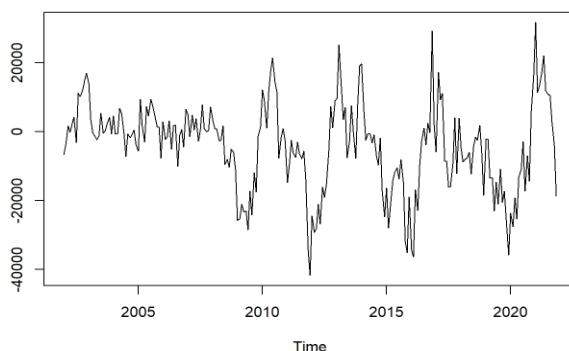
Troisième méthode : modèles ARMA

Les modèles ARMA (Auto Regressive Moving Average) fournissent une approche complémentaire à la prévision des séries chronologiques. Mais cette méthode peut être mise en œuvre seulement pour un modèle additif.

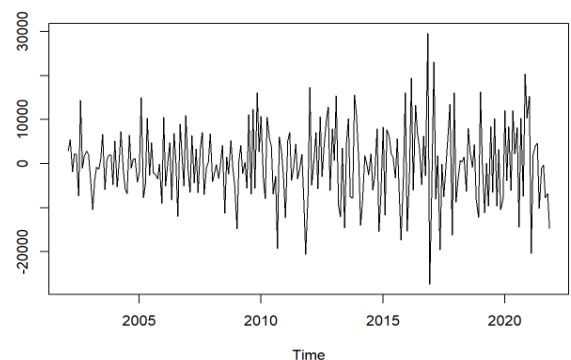
Différenciation de la série et vérification de la stationnarité

On crée la série des différences. Tout d'abord à l'ordre de la période p_0 puis à l'ordre 1. Graphiquement, on perçoit que la série des x_i différenciée à l'ordre 1 présente bien les caractéristiques d'une suite stationnaire.

Série des différences à l'ordre p_0



Série des différences à l'ordre 1



Afin de s'assurer que la série des x_i présente bien les caractéristiques d'une suite stationnaire, nous effectuons le test de Dickey-Fuller augmenté (DFA) avec les hypothèses de test suivantes:

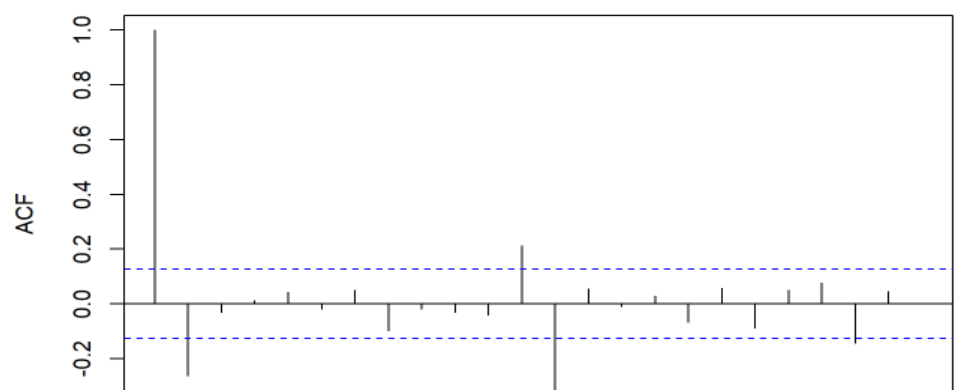
H0: "la série n'est pas stationnaire", H1: "la série est stationnaire"

À l'ordre 1, en utilisant la fonction `adf.test`, nous trouvons une p-valeur égale à 1%. La série différenciée peut donc être considérée comme stationnaire.

Analyse de l'ACF et de la PACF de la série stationnaire

ACF

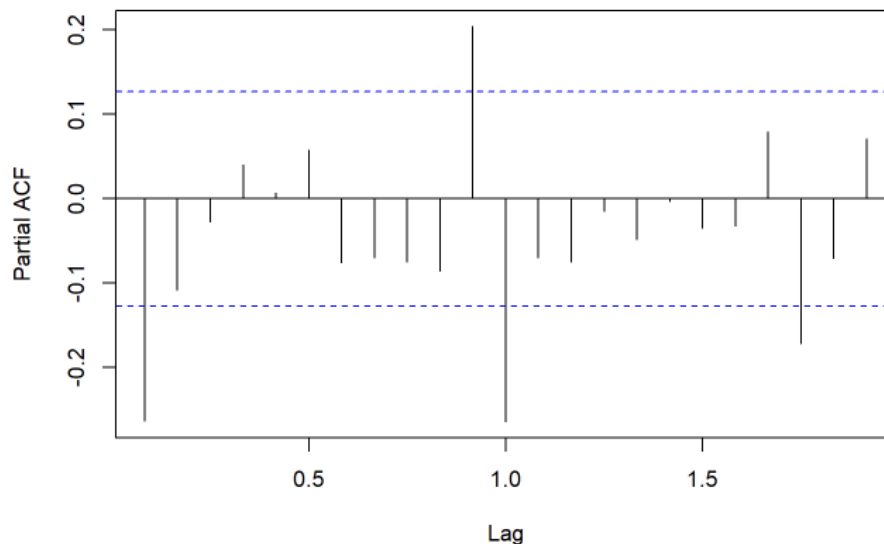
On trace les graphiques de l'ACF (Auto-Correlation Function) et de la PACF (Partial Auto-Correlation)



Function) de la série stationnaire.

Après observation du graphique, on détermine que $q = 2$ et $Q = 1$ est un choix possible.

PACF :



Après observation du graphique, on détermine que $p = 1$ et $P = 1$ est un choix possible.

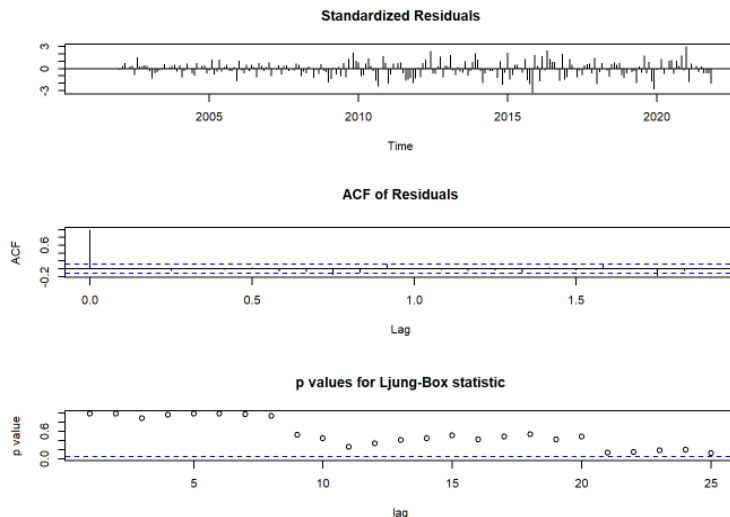
Sélection de modèle grâce à l'AIC

Pour effectuer un graphique le plus fidèle que possible, nous devons utiliser la combinaison de coefficients p , P , q et Q avec l'AIC le plus bas. À l'aide de la fonction `arima`, et à la suite de plusieurs tentatives, nous arrivons à la conclusion qu'il faut conserver le modèle SARMA(0,2,2,3).

- Pour $p = 1$, $P = 1$, $q = 2$, $Q = 1$ on trouve AIC= 4887.43
- Pour $p = 1$, $P = 1$, $q = 1$, $Q = 1$ on trouve AIC= 4889.42
- Pour $p = 1$, $P = 1$, $q = 2$, $Q = 0$ on trouve AIC= 4947.03
- Pour $p = 1$, $P = 1$, $q = 1$, $Q = 0$ on trouve AIC= 4945.08
- Pour $p = 0$, $P = 2$, $q = 2$, $Q = 3$ on trouve AIC= 4878.09 (meilleur AIC)
- Pour $p = 0$, $P = 2$, $q = 2$, $Q = 4$ on trouve AIC= 4878.72
- Pour $p = 0$, $P = 2$, $q = 2$, $Q = 1$ on trouve AIC= 4887.91
- Pour $p = 0$, $P = 2$, $q = 2$, $Q = 0$ on trouve AIC= 4933.24
- Pour $p = 1$, $P = 1$, $q = 2$, $Q = 2$ on trouve AIC= 4888.44
- Pour $p = 1$, $P = 2$, $q = 2$, $Q = 2$ on trouve AIC= 4880.22

On choisira donc le modèle avec $p = 0$, $P = 2$, $q = 2$, $Q = 3$, pourvu que le modèle soit valide, c'est-à-dire, que les résidus soient corrects.

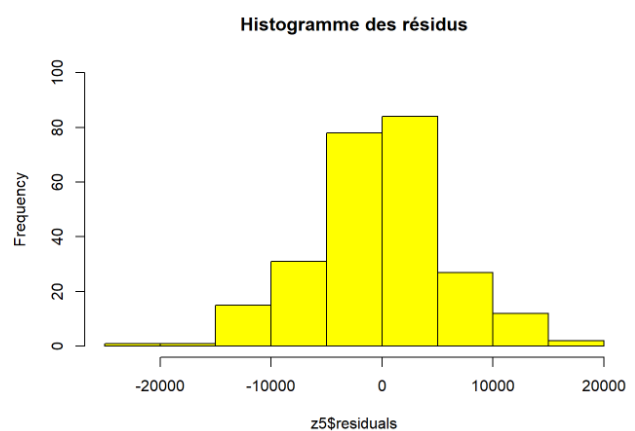
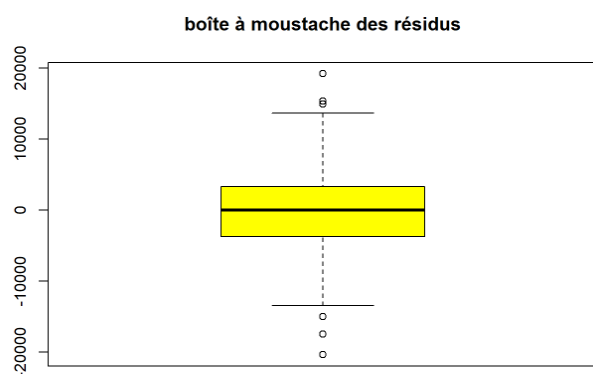
Validation du modèle grâce à l'analyse des résidus



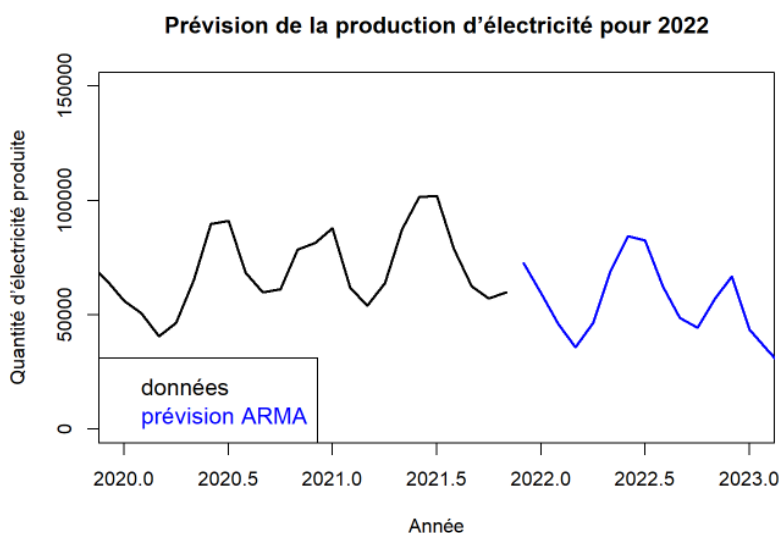
En utilisant la fonction `tsdiag` nous pouvons observer que les résidus sont corrects. En effet, ils ne sont pas écrasés proche de 0.

Le modèle $SARMA(0,2,2,3)$ est donc un modèle utilisable pour nos prévisions.

Nous complétons notre analyse en traçant la boîte à moustache et l'histogramme des résidus.



Prévisions à partir du modèle retenu



À partir de l'équation du modèle $SARMA(0, 2, 2, 3)$ et de l'équation de différenciation, nous sommes capables de prévoir les futures valeurs de la série pour l'année 2022-2023 et de le représenter ci contre.

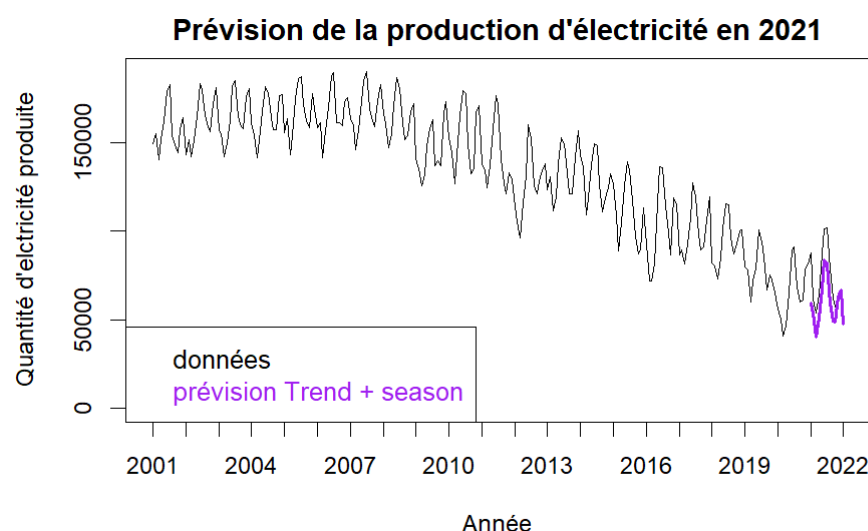
4. Prédiction de la production d'électricité en 2021 : 3 méthodes

Dans cette partie, nous allons prédire la production d'électricité pour 2021, en coupant la série. Donc, les prédictions seront réalisées en utilisant les trois mêmes méthodes différentes :

- estimation de tendance + coefficients saisonniers
- prévisions de Holt-Winters
- modèles ARMA : en prenant les coefficients saisonniers pour les 3 dernières années

Nous ne rentrerons pas dans les détails techniques de ces méthodes, car elles ont déjà été expliquées précédemment lors de la prévision pour 2022.

Première méthode : estimation de tendance + coefficients saisonniers

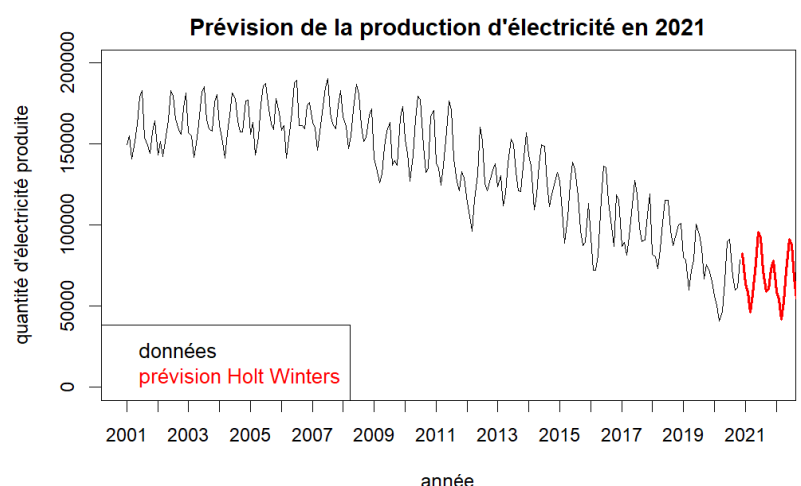


Avec cette première méthode, on retrouve la courbe réelle pour 2021, en utilisant les données de 2001 à 2020. En comparant avec les données réelles, la prédiction ne semble pas juste. En effet, sur l'année 2021, le motif saisonnier est différent.

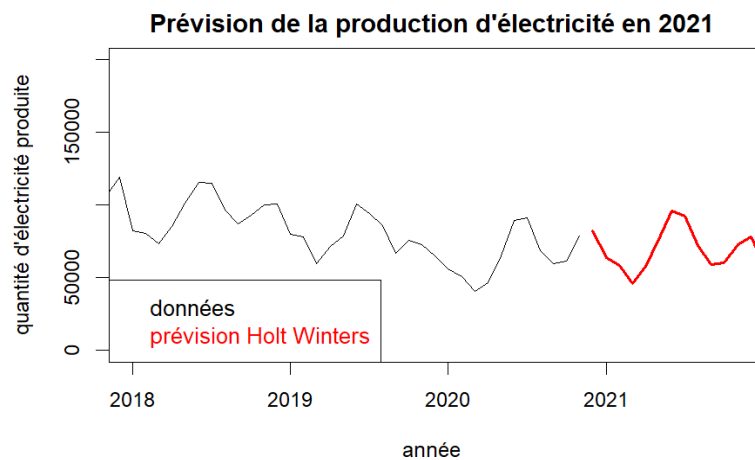
Nous devons essayer d'autres méthodes et la comparer à celles-ci.

Deuxième méthode : les prévisions de Holt-Winters

Le graphique montre les données historiques de taux de charbon jusqu'à fin 2020, ainsi que les prévisions de Holt-Winters pour les années prochaine, représentées en rouge. La prévision de Holt-Winters suggère donc que cette périodicité se

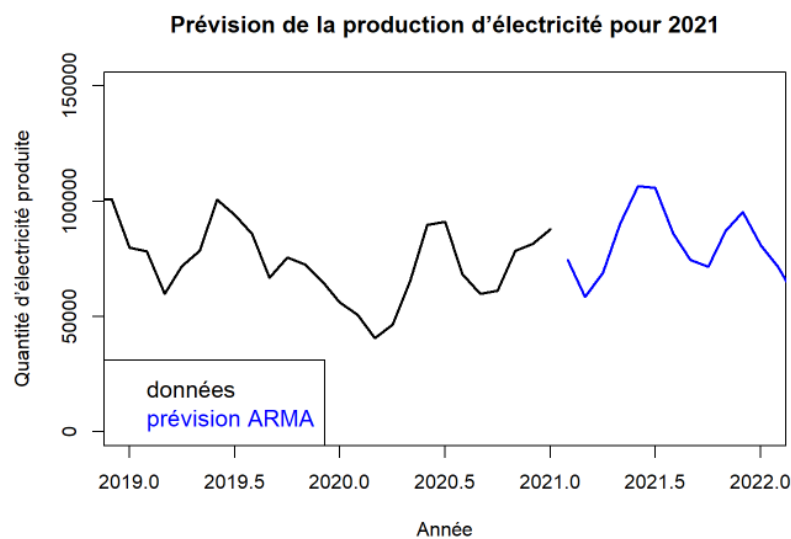


poursuivra à l'avenir, et que la production d'électricité liée au charbon sera très proche de celle de 2020 avec des valeurs très proches.

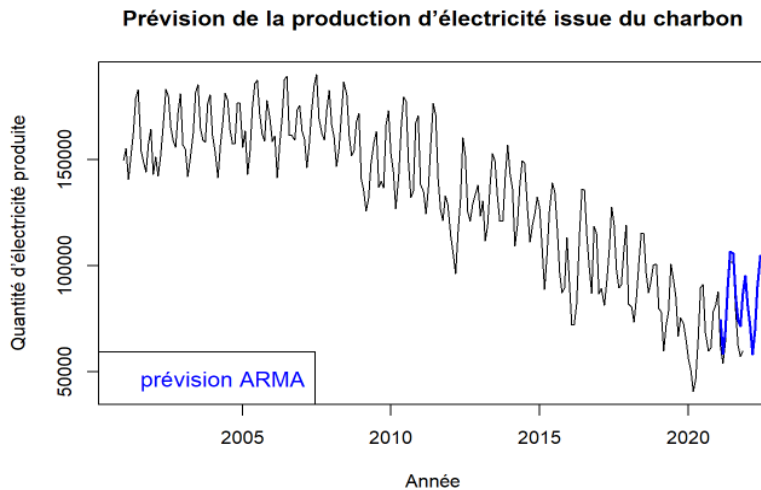


Ici on a réduit les données historiques de 2018 à fin 2020 pour avoir des données plus précises et on voit bien que la prévision nous montre une production très similaire à celle de l'année 2020. avec des quantités de production presque égales.

Troisième méthode : les modèles ARMA



Le modèle ARMA nous fournit un résultat similaire avec des prévisions de production similaire à celle de l'année 2020.



Comparaison des modèles : Erreur quadratique moyenne (EQM)

L'EQM est une mesure de l'erreur moyenne de prédiction pour les modèles de prévision. Nous allons l'utiliser pour comparer différents modèles et évaluer leur performance de prédiction. Donc, celle-ci est calculée en prenant la différence entre les valeurs observées et les valeurs prédites pour chaque période de temps, en élevant cette différence au carré et en prenant la moyenne de ces carrés. Ainsi, on trouve une mesure de l'erreur moyenne de prédiction pour l'ensemble de la série chronologique. Et le modèle renvoyant la plus petite EQM est dit meilleur modèle de prévision.

$$EQM = \frac{\sum_{i=0}^n (\hat{y}_i - y_i)^2}{n}$$

Prévisions tendance + coefficients saisonniers :

EQM : 473470237

Prévisions Holt-Winters :

EQM : 198014493

Prévisions ARMA :

EQM : 127637923

On choisit le modèle ARMA, car il correspond à l'EQM le plus faible.

Conclusion

Our analysis and time series forecasting results could help guide investment decisions and long-term energy development strategies by providing accurate forecasts of electricity production.

Additionally, the time series analysis could help identify past trends and key factors that have impacted electricity production over time. Even though the electricity production from coal has drastically reduced during the last decade. Our previsions show that it will maintain its current place in the near future of electricity production. Based on our forecasts and their comparisons, we choose to adopt the ARMA model.