
SAE « Description et Prévision de séries temporelles »

Production d'énergie solaire aux Etats-Unis (2001 - 2022)

Introduction :

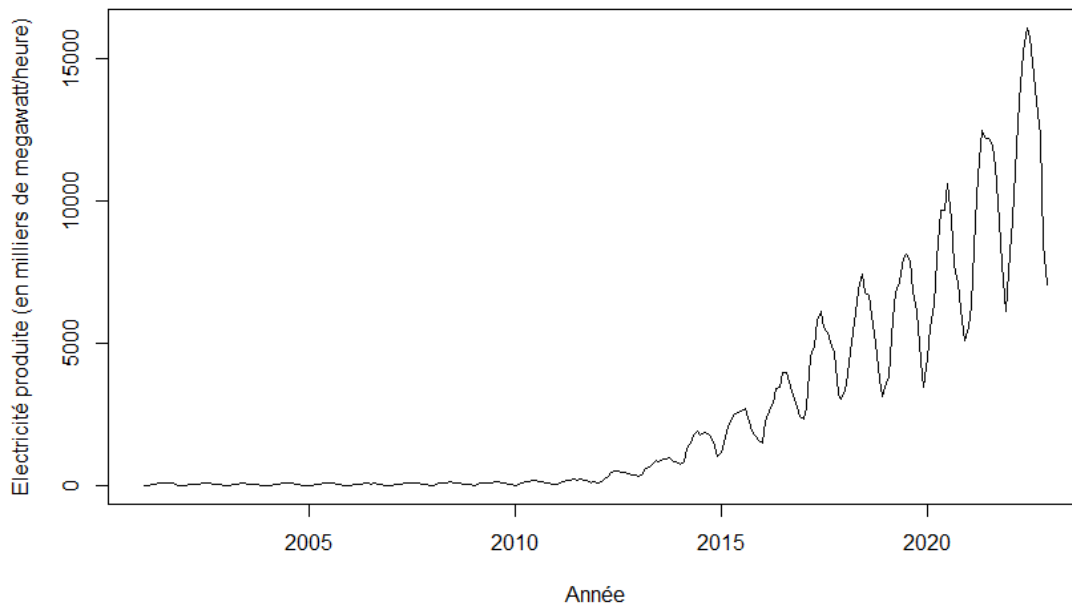
Au sein du contexte énergétique américain, l'Administration de l'Information sur l'Énergie (EIA) se positionne en tant qu'acteur essentiel du Système Statistique Fédéral des États-Unis. Sa mission première consiste à collecter, analyser et diffuser des informations énergétiques pour favoriser l'élaboration de politiques éclairées, assurer l'efficacité des marchés et promouvoir la compréhension publique des interactions entre l'énergie, l'économie et l'environnement. À travers ses programmes diversifiés, l'EIA englobe des données sur des sources variées telles que le charbon, le pétrole, le gaz naturel, l'électricité, les énergies renouvelables et nucléaire, jouant un rôle central au sein du Département de l'Énergie des États-Unis (*source Wikipedia*).

Parmi la multitude de séries temporelles disponibles sur le site de l'EIA, nous nous concentrons sur celles décrivant la production d'électricité sur une période de 22 ans, de 2001 à 2022. Ces données mensuelles détaillent la production par type d'énergie, allant du charbon aux énergies renouvelables telles que l'énergie solaire, éolienne, et géothermique. L'unité de mesure commune à toutes ces séries est le millier de mégawattheures.

Notre mission sera divisée en plusieurs parties. Nous devons tout d'abord présenter la tendance de la série, en utilisant un filtre de moyennes mobiles et la courbe de régression des moyennes annuelles. Ensuite, notre tâche comprend le calcul et la représentation graphique des coefficients saisonniers, ainsi que la décomposition de la série. Nous devons également analyser et commenter la série ajustée saisonnièrement, tout en générant et interprétant les boxplots des résidus.

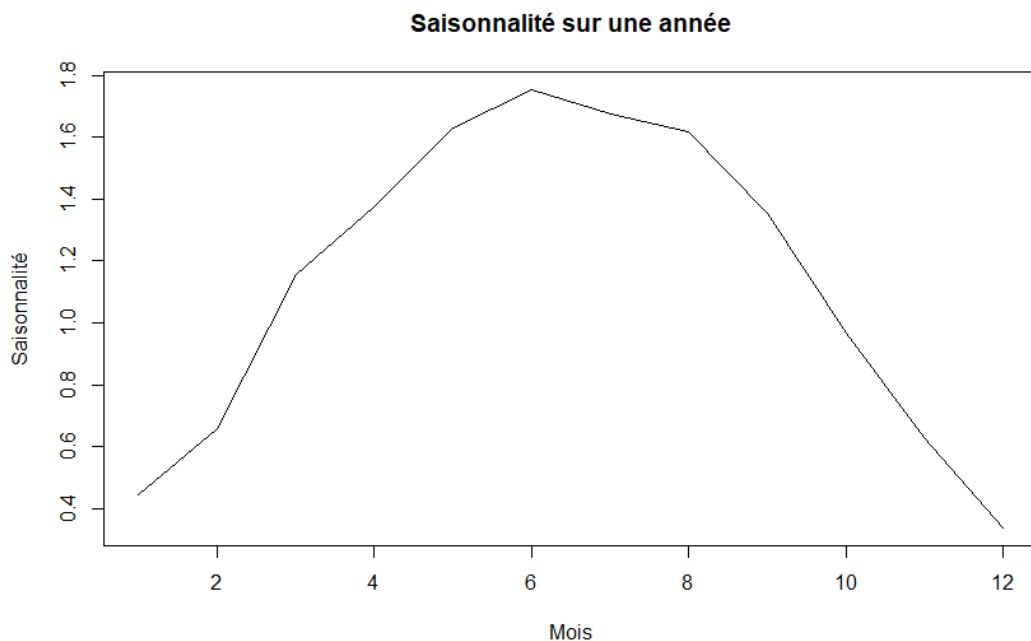
Nous proposerons une estimation de la production d'électricité pour l'année 2023 en utilisant trois méthodes distinctes. De plus, une répétition de ces méthodes est requise pour l'année 2022, en utilisant les données de la période allant de 2001 à 2021. Cela nous permet de tester la fiabilité des méthodes utilisées, car nous connaissons déjà les valeurs pour l'année 2022. Enfin, nous devons calculer l'erreur quadratique moyenne de la prévision pour chacune des trois méthodes utilisées.

Evolution de la production énergétique solaire aux USA entre 2001 et 2022



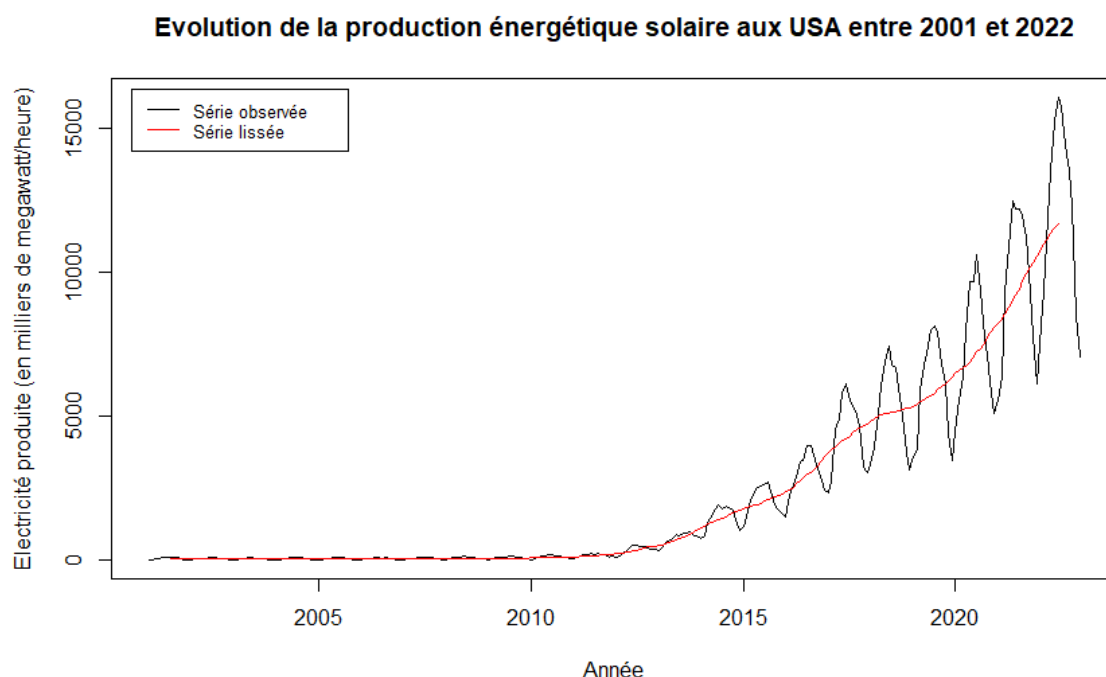
Ce graphique représente la série temporelle étudiée, soit la production énergétique solaire aux Etats Unis entre 2001 et 2022. On constate une très rapide augmentation à partir de 2012, car avant cette date, la production approchait les 0 megawatt/heure. Ainsi, l'augmentation brutale avec une forte saisonnalité en constante augmentation amène à conclure qu'il s'agit d'une série multiplicative.

Saisonnalité :



La saisonnalité s'obtient grâce à la méthode 'decompose'. La série étant multiplicative, il faut utiliser le logarithme des données pour avoir accès à la méthode 'decompose', uniquement compatible avec les données de type additives. On observe ici que la production énergétique solaire augmente de janvier à juin, où elle atteint un pic, puis elle redescend jusqu'à décembre. Cela s'explique simplement par le fait qu'il y a plus de soleil en été, et la production énergétique solaire augmente donc jusqu'à cette période avant de redescendre en même temps que la quantité de soleil.

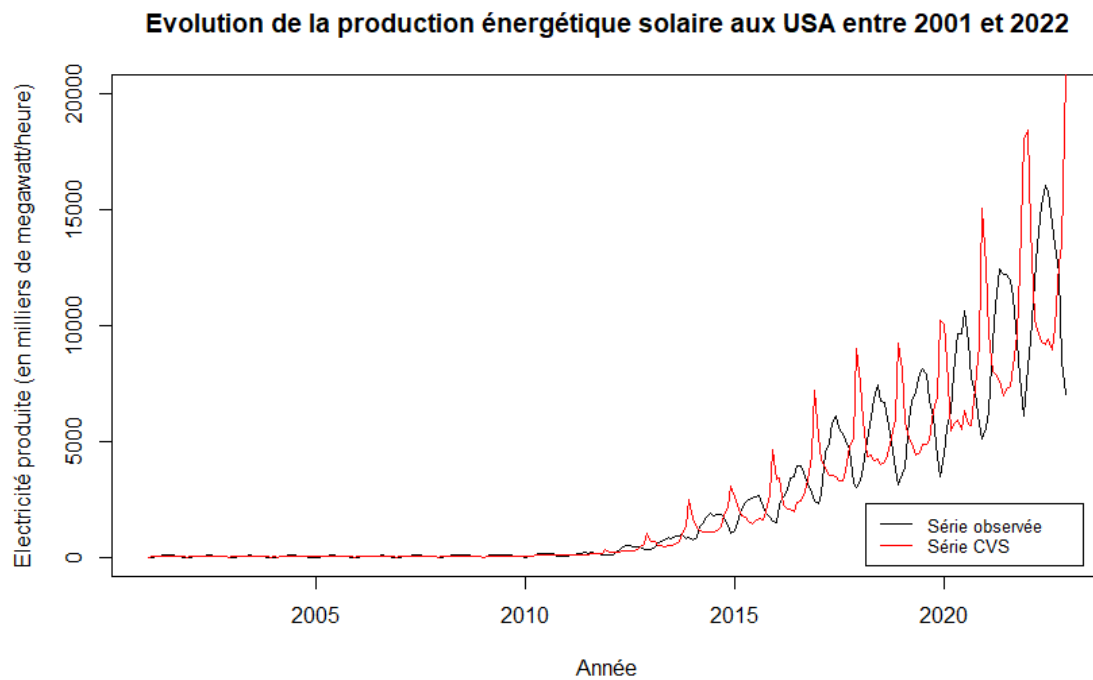
Série lissée :



Comme la saisonnalité, la série lissée s'obtient grâce à la méthode 'decompose'. Cette série lissée est utile pour analyser plus facilement les tendances et les motifs temporels, car elle réduit l'impact des variations multiplicatives qui peuvent rendre la série originale plus difficile à interpréter. On constate que la série lissée commence un peu après le début de la série observée et s'arrête un peu avant la fin. Cela s'explique par le fait qu'elle est calculée à l'aide des moyennes mobiles centrées d'ordre 12, ce qui cause donc une absence de données sur les 6 premiers et derniers mois.

Sur ce graphique, on observe comme précédemment une augmentation de la production énergétique solaire à partir de 2012. Avant cette date, la série lissée est linéaire, car les variations sont très minimales à cette période. Enfin, on constate également un ralentissement de l'augmentation en 2019, ce qui n'était pas forcément visible sur le graphique original.

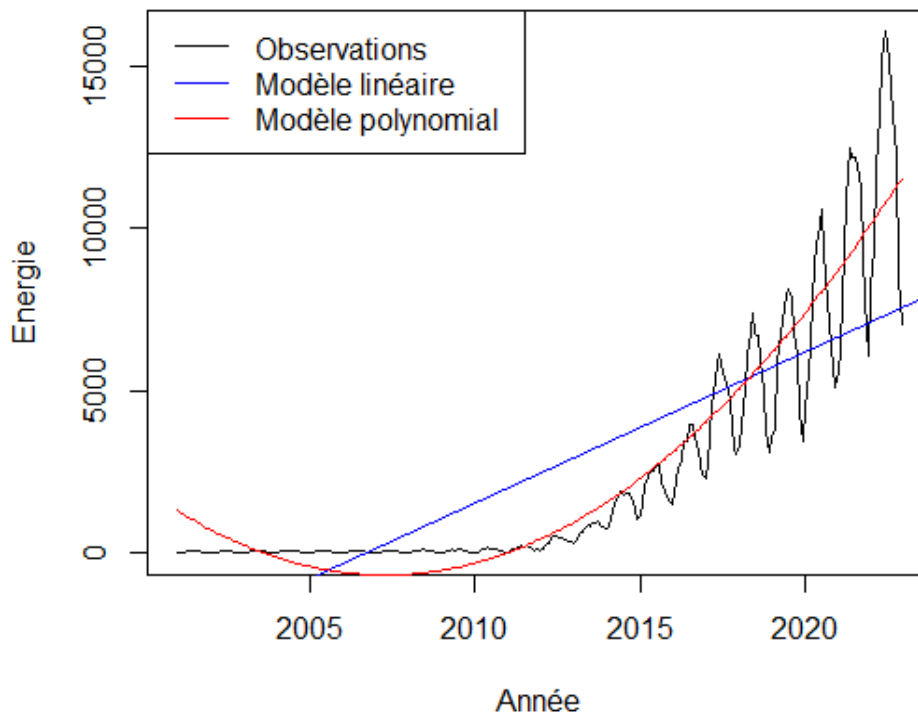
Série corrigée des variations saisonnières :



La série corrigée des variations saisonnières (CVS) permet d'annuler les effets des variations saisonnières sur une série temporelle pour faire mieux apparaître la tendance. Pour une série de type multiplicative, la CVS s'obtient en divisant les valeurs de la série temporelle par les variations saisonnières.

Ici, les variations saisonnières sont telles que l'on a obtenu une symétrie avec la série originelle. Ceci ne permet donc pas une meilleure lisibilité de la tendance.

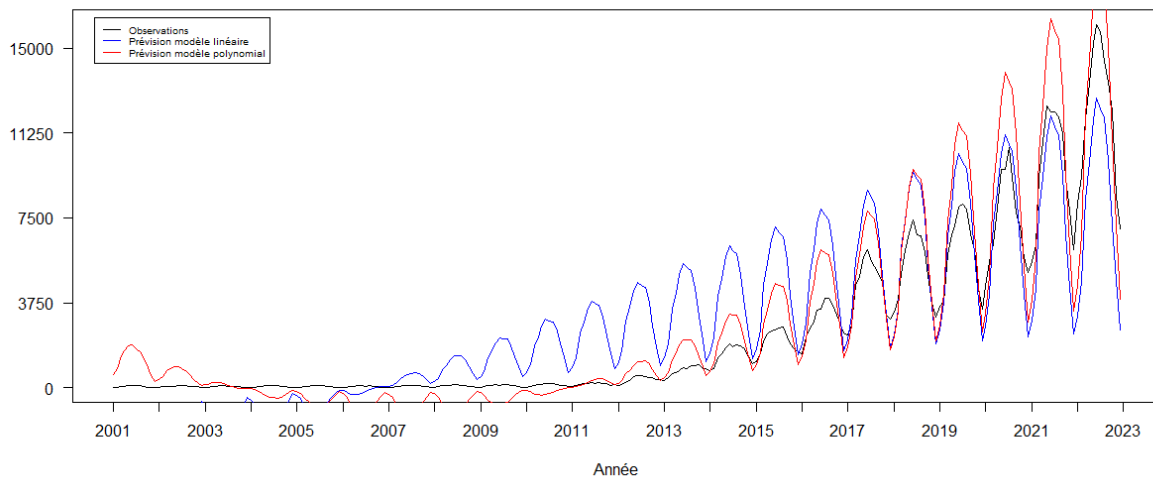
Tendances de la production énergétique solaire aux US



Le graphique ci-dessus regroupe un modèle linéaire et polynomial, privé de saisonnalité, de la série.

Ces deux modèles mettent en évidence la croissance de la production énergétique solaire aux USA, qui a commencé à se développer sérieusement vers 2012. La courbe de tendance polynomiale est plus précise et montre une croissance qui tend à s'amplifier au fil des années.

Production énergétique solaire entre 2000 et 2022 aux USA (en milliers de megawatt/heure)

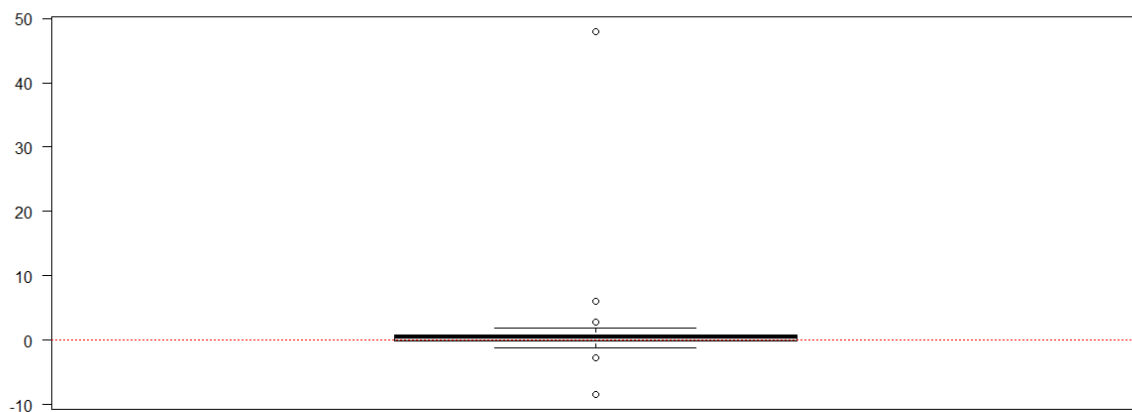


En ajoutant la saisonnalité aux modèles linéaire et polynomial, on obtient une meilleure représentation de la série observée.

On remarque que ces deux modèles représentent avec quelques limites la tendance de la courbe, avec de grosses différences avant 2015 pour le modèle linéaire, et entre 2005 et 2010 pour le modèle polynomial. Les pics de la série sont dans l'ensemble assez sur-estimés par les deux modèles.

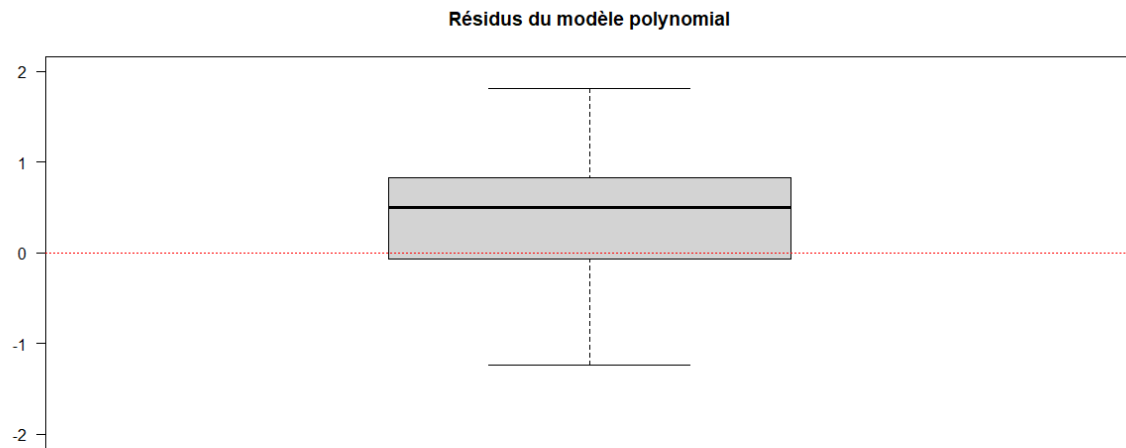
Cela peut s'expliquer par le fait que la production électrique solaire reste presque totalement nulle jusqu'en 2012.

Résidus du modèle polynomial



Regardons maintenant la boîte à moustache des résidus du modèle polynomial.

On peut relever la présence d'un résidu extrêmement haut (presque 50), en plus de quelques-uns assez éloignés de 0. Nous pouvons restreindre l'axe des ordonnées pour voir plus précisément ce qu'il se passe pour les résidus plus proches de 0.



Cette boîte à moustache est identique à celle au-dessus, seuls les valeurs entre -2 et 2 ont été conservées, pour voir leur répartition plus facilement.

La tendance à surestimer, relevée plus haut, du modèle polynomial se confirme ici, une grande partie des résidus se trouvent au-dessus de 0.

Prédiction de la production d'électricité pour l'année 2023 :

Afin de faire des prédictions pour l'année de 2023, nous avons utilisé trois méthodes : ARMA, Holt-Winters et la méthode polynomiale.

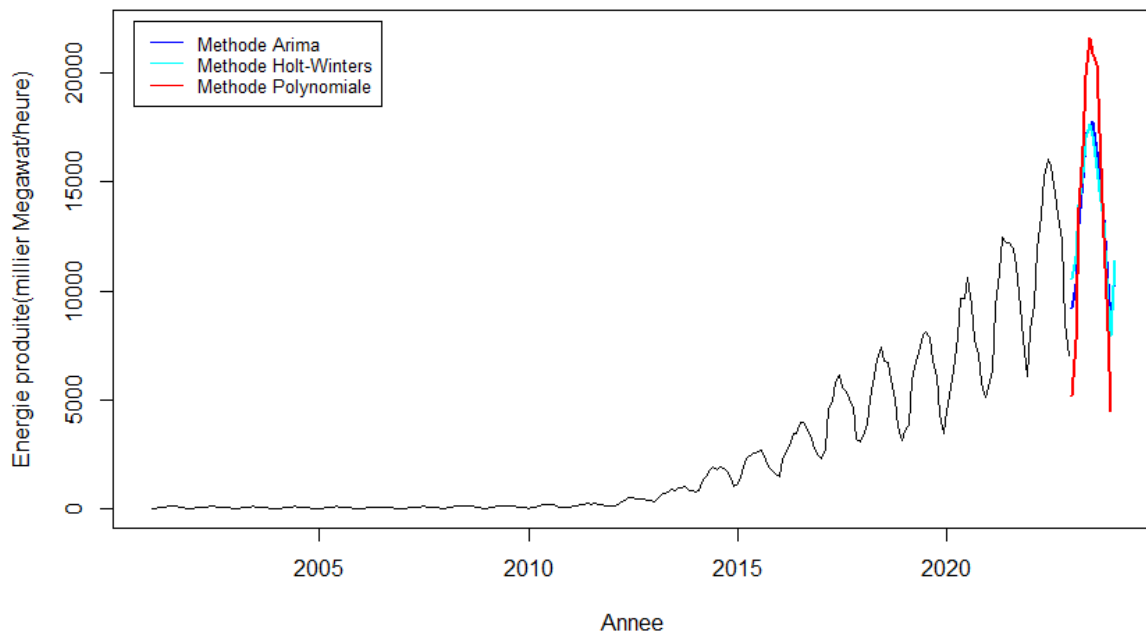
Pour la première méthode, nous avons d'abord cherché cinq modèles ARIMA différents (z1 à z5), puis nous avons calculé leur AIC. On peut alors voir que s'est z2 qui a le plus petit AIC et s'est celui ci que nous allons choisir pour la suite du projet.

```
> AIC(z1)
[1] 3620.988
> AIC(z2)
[1] 3620.656
> AIC(z3)
[1] 3685.393
> AIC(z4)
[1] 3674.408
> AIC(z5)
[1] 3673.902
```

Pour la méthode Holt-Winters, nous avons ajusté le modèle Holt-Winters sur les données de la série et de générer les prédictions.

Et pour la dernière méthode utilisée, la méthode polynomiale, nous avons d'abord fait le logarithme de la série temporelle et fait sa décomposition. Ensuite, nous avons initialisé x et y, calculé la CVS, et ajusté un modèle polynomial nous permettant de calculer la tendance polynomiale pour la période future et ainsi en faire les prévisions.

Evolution production solaire entre 2001 et 2024 aux USA



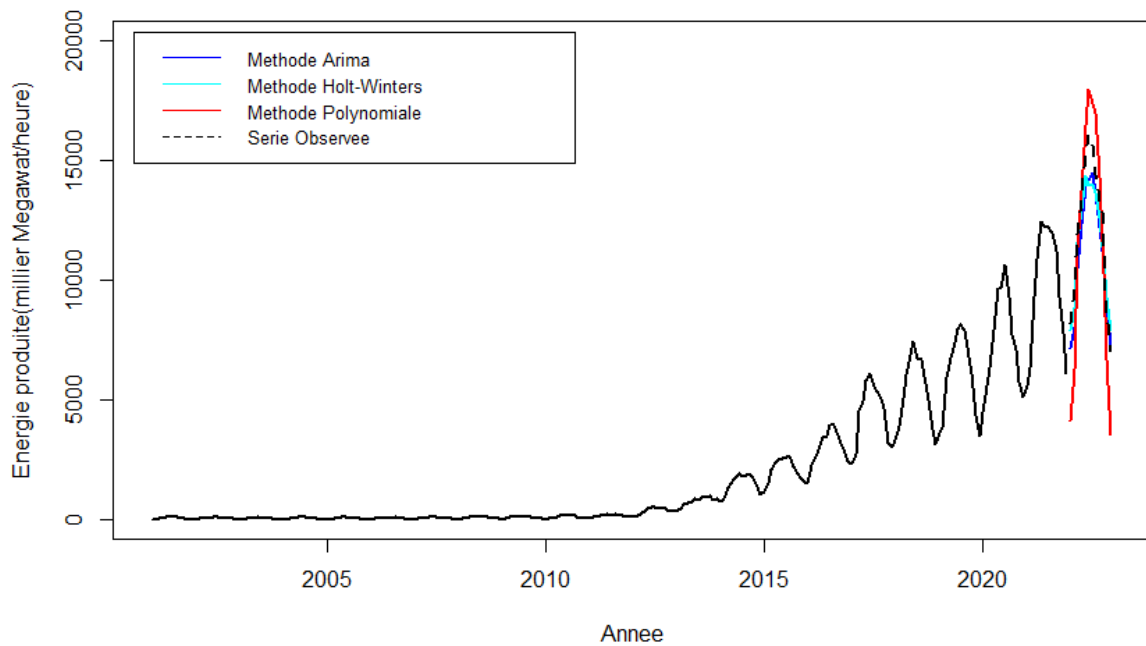
Prédiction de la production d'électricité pour l'année 2022 :

Dans le but de déterminer quelle méthode donne une meilleure prédiction, nous avons fait une prédiction pour l'année 2022.

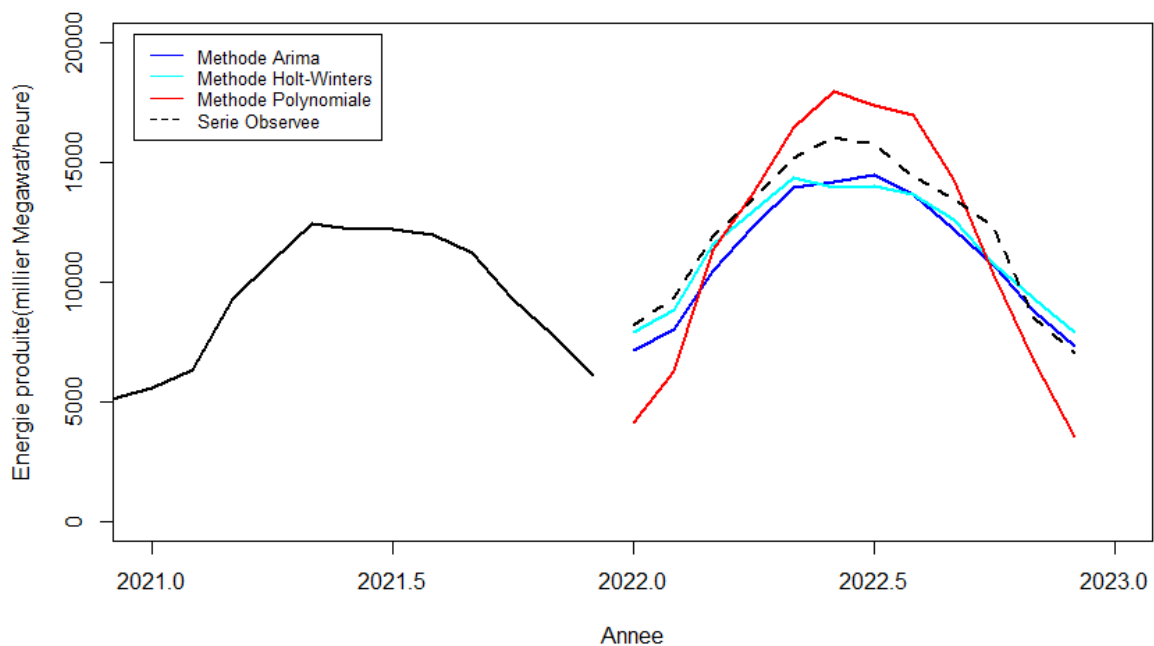
Dans un premier temps, nous avons fait un échantillon d'apprentissage et un échantillon de test : l'échantillon d'apprentissage est sur les années de 2001 à 2021 et celui de test est sur l'année de 2022.

Ensuite nous avons refait comme pour la partie pour les prédictions pour l'année 2023 mais en utilisant les échantillons créés précédemment.

Evolution production solaire entre 2001 et 2023 aux USA



Evolution production solaire entre 2021 et 2023 aux USA



Sur le premier graphique nous pouvons voir les prédictions pour les différentes méthodes et la série observée, afin de voir graphique les différences entre les méthodes et les données d'origine. Toutefois, nous avons fait un graphique faisant un zoom sur l'année 2021 et 2022, dans le but d'avoir une représentation graphique plus lisible et précise.

Graphiquement, nous pouvons voir que la prévision représentée en cyan est plus proche de l'échantillon observé que les autres prévisions sur les minimums de la série. Alors que pour les valeurs maximales de la série observée, nous sommes plus proche de la prévision faite par la méthode Polynomiale.

Calcul des Erreurs Quadratiques Moyennes :

Pour déterminer quelle méthode permet de faire les meilleures prédictions. La plus petite valeur d'EQM détermine la méthode à choisir.

Pour calculer les EQM, nous avons d'abord calculé le logarithme de l'échantillon test. Et, pour chaque EQM, nous avons calculé la moyenne du logarithme de la prévision de la méthode, moins le log test au carré.

```
> mse_arima
[1] 0.01086636
> mse_hw
[1] 0.007532279
> mse_poly
[1] 0.103816

ltest=log(test)
mse_arima = mean((log(pr$pred)-ltest)^2)
mse_hw = mean((log(pr_hw)-ltest)^2)
mse_poly = mean((log(prpoly2)-ltest)^2)
```

On peut voir alors que pour la méthode Holt-Winters a la valeur la plus petite, et, est la méthode qui permet d'avoir la prévision la plus précise.

Conclusion :

Pour conclure, suite à l'analyse de la série temporelle sur la production d'énergie solaire aux États-Unis entre 2001 et 2022, nous avons mis en évidence une croissance à partir de 2012. La saisonnalité a été déterminée avec la méthode de décomposition, montrant une augmentation de la production de janvier à juin, et une diminution jusqu'au mois de décembre en raison de la variation d'ensoleillement.

Les modèles linéaire et polynomial ont été ajustés pour inclure cette composante, améliorant la représentation de la série observée. Cependant, ces modèles ont montré des limites, particulièrement avant 2015 pour le modèle linéaire et entre 2005 et 2010 pour le modèle polynomial, surestimant les pics de production. L'examen des résidus du modèle polynomial a confirmé nos analyses.

En vue de faire des prédictions pour l'année 2023, trois méthodes ont été utilisées : ARMA, Holt-Winters et la méthode polynomiale. Pour vérifier la fiabilité des méthodes, nous avons fait la comparaison des prédictions pour l'année 2022. Ceci nous a révélé que la méthode Holt-Winters présentait la plus petite erreur quadratique moyenne (EQM), indiquant une meilleure précision dans les prévisions.

Nous pouvons faire cette conclusion en s'appuyant sur une évaluation des différents modèles, de leur ajustement saisonnier, et de la comparaison des prédictions avec les données réelles.

Summary:

This analysis focuses on the production of solar energy in the US between 2001 and 2022. The data is provided by the U.S. Energy Information Administration (EIA).

We focused our analysis on several points. First, we presented the trend of the series using moving averages and annual regression curves. The series exhibits a rapid increase from 2012 onwards, indicating a multiplicative nature. We had to calculate and graphically represent seasonal coefficients, by decomposing the series. We then presented the seasonally adjusted series.

We also tried to get the trend of the series by representing it with linear and polynomial models. Then, we generated and interpreted residuals boxplots, to see if they were accurate representations and found that there were some imperfections in the models.

Estimates for electricity production in 2023 were provided through three distinct methods (ARMA, Holt-Winters, and polynomial), with a repetition of these methods for 2022 to test their reliability. The mean squared error for each forecasting method was also calculated, and we found that the Holt-Winters method was the better approach.