

Thomas BELAÏD
Baptiste TIVRIER
BUT SD 2AFA - VCOD

SAÉ

"SÉRIES TEMPORELLES"

➤ Introduction :

L'agence d'information sur l'énergie (EIA) des États-Unis est un organisme principal du système statistique fédéral des États-Unis chargé de recueillir, d'analyser et de diffuser l'information sur l'énergie afin de favoriser l'élaboration de politiques éclairées, des marchés efficaces et la compréhension publique de l'énergie et de ses interactions avec l'économie et l'environnement. Les programmes de l'EIA couvrent les données sur le charbon, le pétrole, le gaz naturel, l'électricité, les énergies renouvelables et l'énergie nucléaire. L'EIA fait partie du Département Américain de l'Énergie.

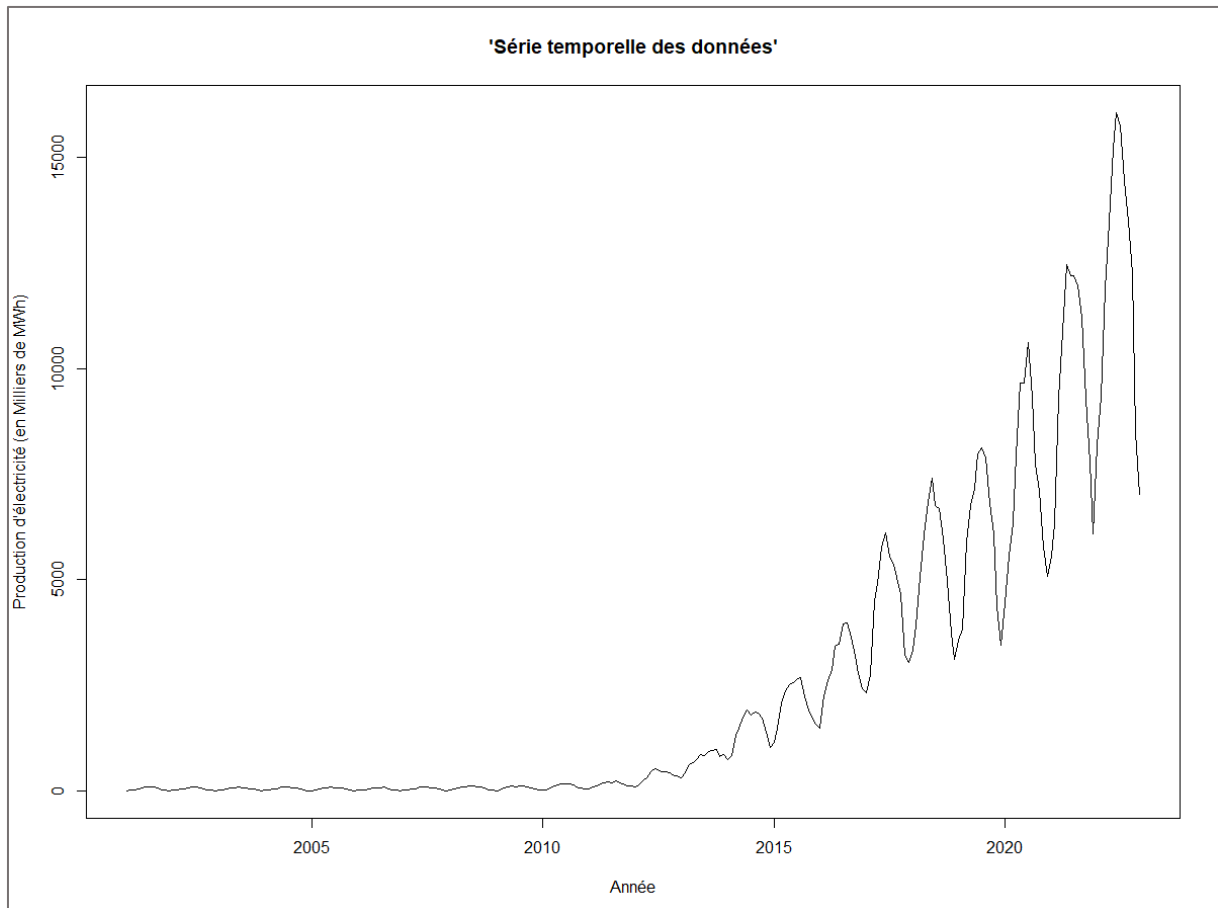
Parmi les centaines de séries chronologiques que l'on peut trouver sur le site Web de l'EIA, nous nous intéressons à celles qui décrivent la production d'électricité entre 2001 et 2022 (donc 22 années complètes d'observation). Il s'agit de données mensuelles ventilées par type de production d'électricité : charbon, liquides de pétrole, coke de pétrole, gaz naturel, autres gaz, nucléaire, hydroélectrique classique, éolien, solaire à grande échelle, géothermique, biomasse. Dans notre cas, notre mission sera d'étudier la série temporelle portant sur la production d'énergie solaire à grande échelle entre 2001 et 2022.

Pour cela, nous aurons 5 différentes étapes à effectuer :

- Tout d'abord, nous regarderons la tendance de la série grâce à 2 courbes obtenues différemment. Pour cela, nous utiliserons un filtre par moyennes mobiles centrées et une régression des moyennes annuelles.
- Ensuite, nous calculerons et tracerons les coefficients saisonniers de la série tout en représentant la décomposition de la série par ses 3 différentes composantes.
- De plus, nous établirons une analyse de la série corrigée des variations saisonnières ainsi qu'une analyse de la boîte à moustaches des résidus.
- Puis, nous prédirons la production d'énergie durant l'année 2023 à l'aide de 3 méthodes différentes, tout en les comparant.

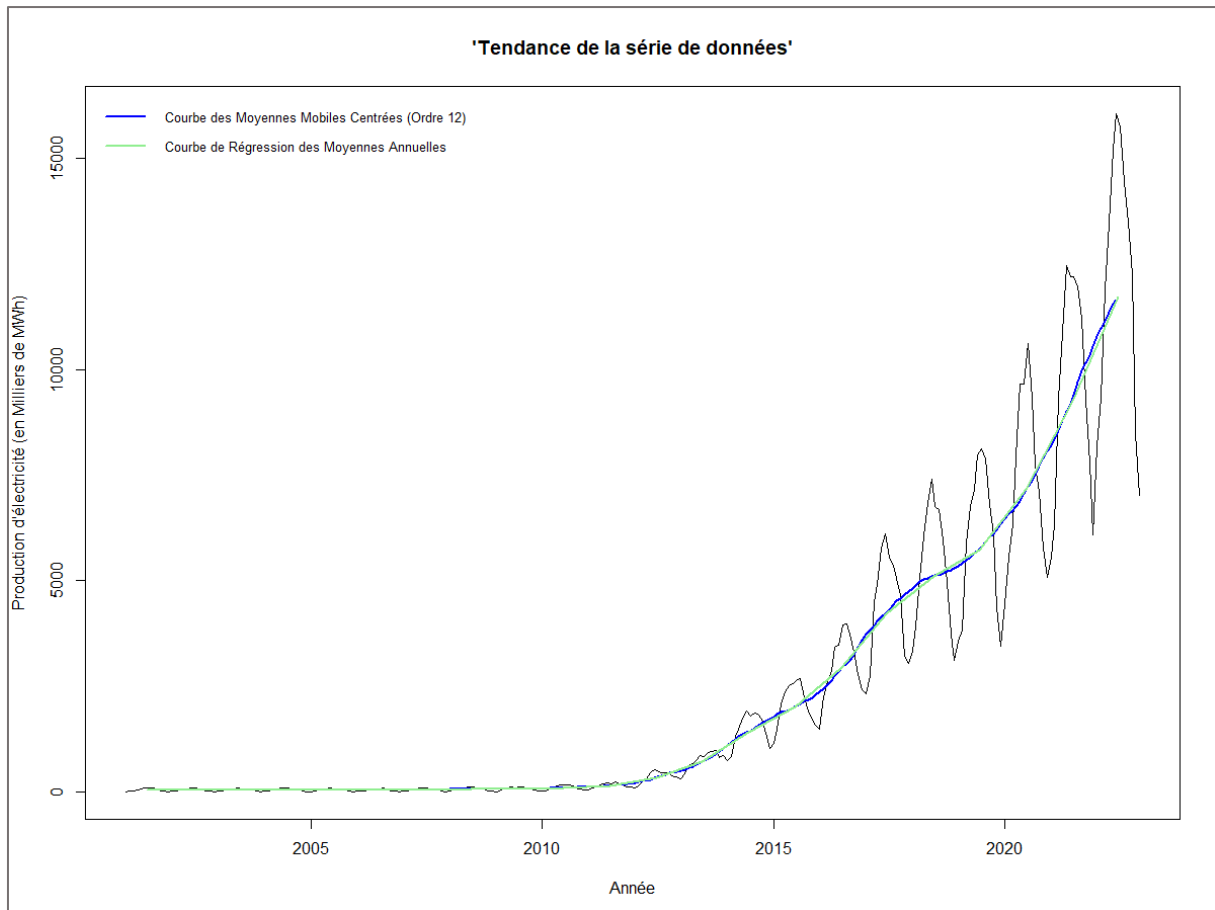
- Enfin, nous reprédirons à l'aide des 3 mêmes méthodes utilisées la production d'énergie durant l'année 2022 grâce aux données de 2001 à 2021. Nous calculerons l'erreur quadratique moyenne de chaque méthode pour savoir laquelle était la plus proche des valeurs réelles.

➤ Série temporelle des données :



Le graphique ci-dessus représente la série temporelle des données sur la période de 2001 à 2022. Les données initiales, couvrant la période de 2001 jusqu'à environ 2012, présentent des résultats quasi nuls, suivies par des phases successives d'augmentation et de diminution de plus en plus prononcées à partir de cette année-là. Nous pouvons déduire qu'il s'agit d'une série multiplicative dû à la différence d'amplitude de la production d'énergie au fil des années.

➤ Tendence de la série de données :



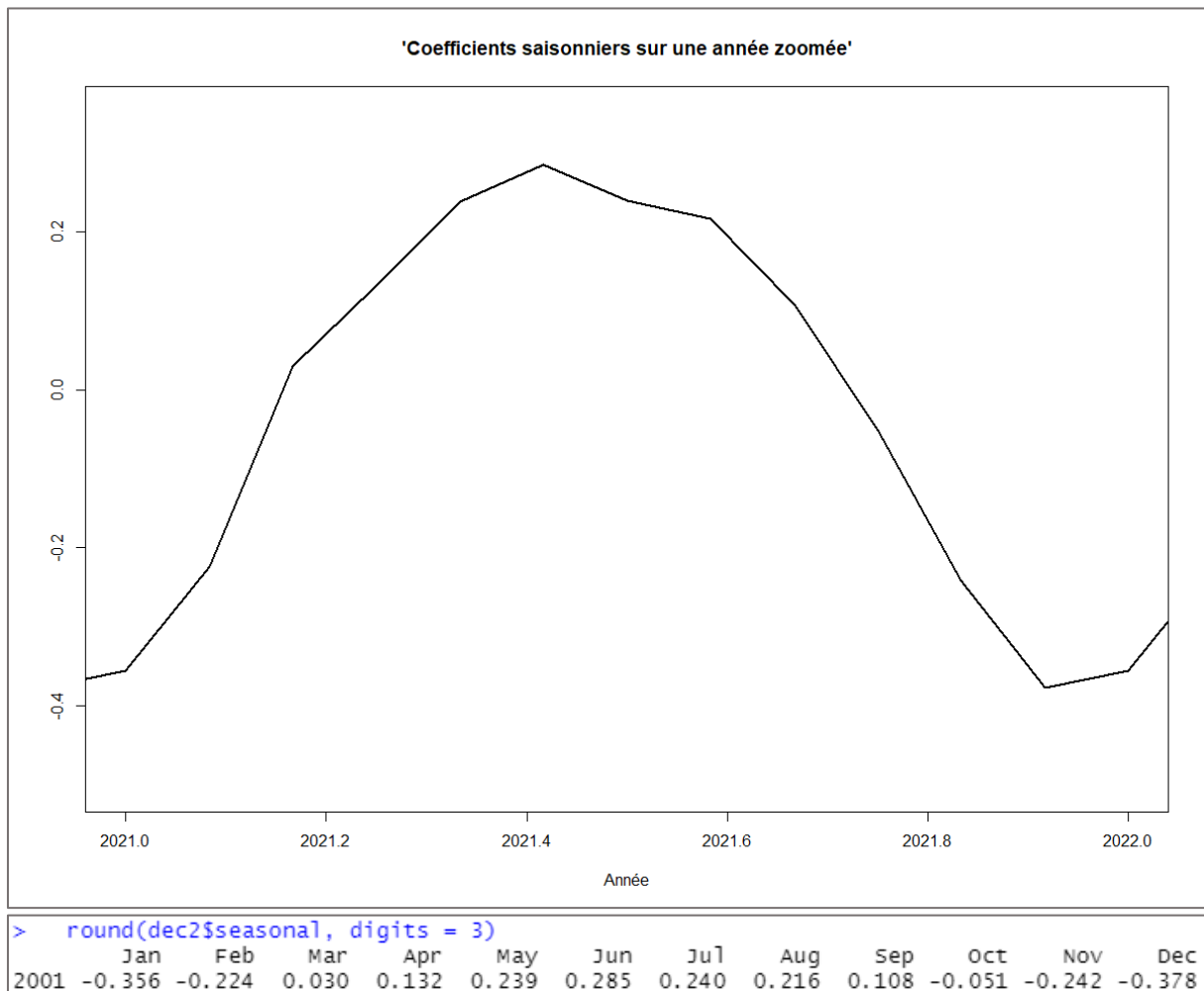
Le graphique ci-dessus représente la tendance de la série de données. L'analyse de la tendance peut être vue de deux manières différentes : par le lissage au moyen de moyennes mobiles centrées ou par l'utilisation de moyennes annuelles. Ces deux approches ont pour but d'atténuer les fluctuations saisonnières, en vue d'obtenir une représentation plus stable de l'évolution à long terme de la série.

La méthode par moyennes mobiles centrées consiste à calculer les moyennes des valeurs symétriquement selon un ordre. Ici, nous sommes sur une saisonnalité annuelle donc nous allons prendre un ordre égal à 12 qui va calculer les valeurs selon le symbole de multiplication égal à $[1/24, 11 \cdot (1/12), 1/24]$. La méthode de régression par moyennes annuelles, quant à elle, consiste à calculer les valeurs moyennes annuelles d'une série temporelle, puis à appliquer une analyse de régression sur ces moyennes.

Nous pouvons premièrement remarquer que ces 2 méthodes renvoient des courbes quasi exactes. Ensuite, nous pouvons observer que la tendance

reste presque nulle jusqu'à 2012 et connaît dès lors une augmentation progressive de plus en plus prononcée. En revanche, nous pouvons voir grâce aux courbes que cette augmentation a ralenti entre 2018 et 2019.

➤ Coefficients saisonniers :

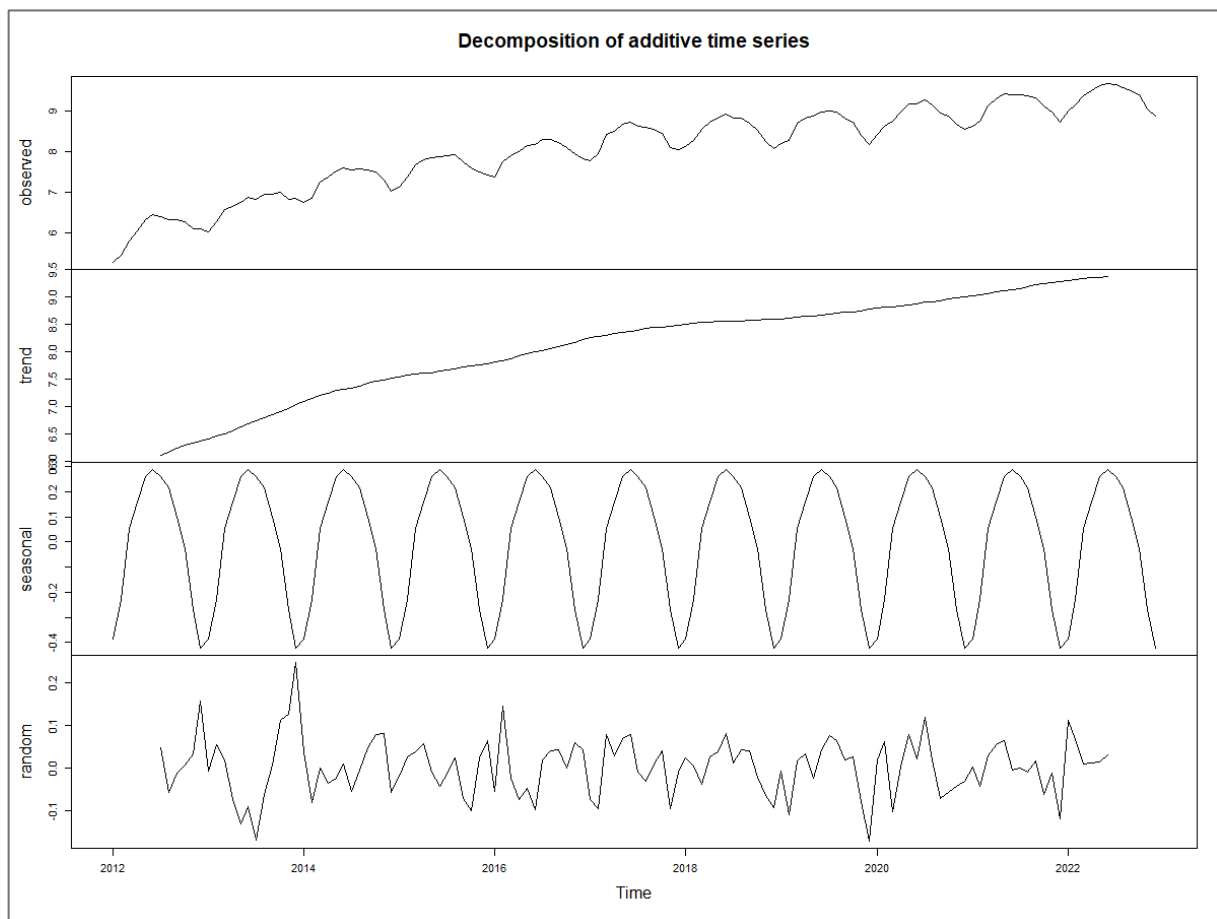


Le graphique ci-dessus représente les coefficients saisonniers sur une année choisie (chacune présente les mêmes coefficients), ce qui va nous permettre d'analyser les variations saisonnières. Nous avons obtenu ces résultats en décomposant la série mise sous forme logarithmique tout en choisissant le modèle saisonnier.

Nous pouvons observer l'augmentation significative de la production d'énergie au cours de la première moitié de l'année, atteignant un pic vers juin, suivie d'une diminution lente puis drastique. Cette variation peut être

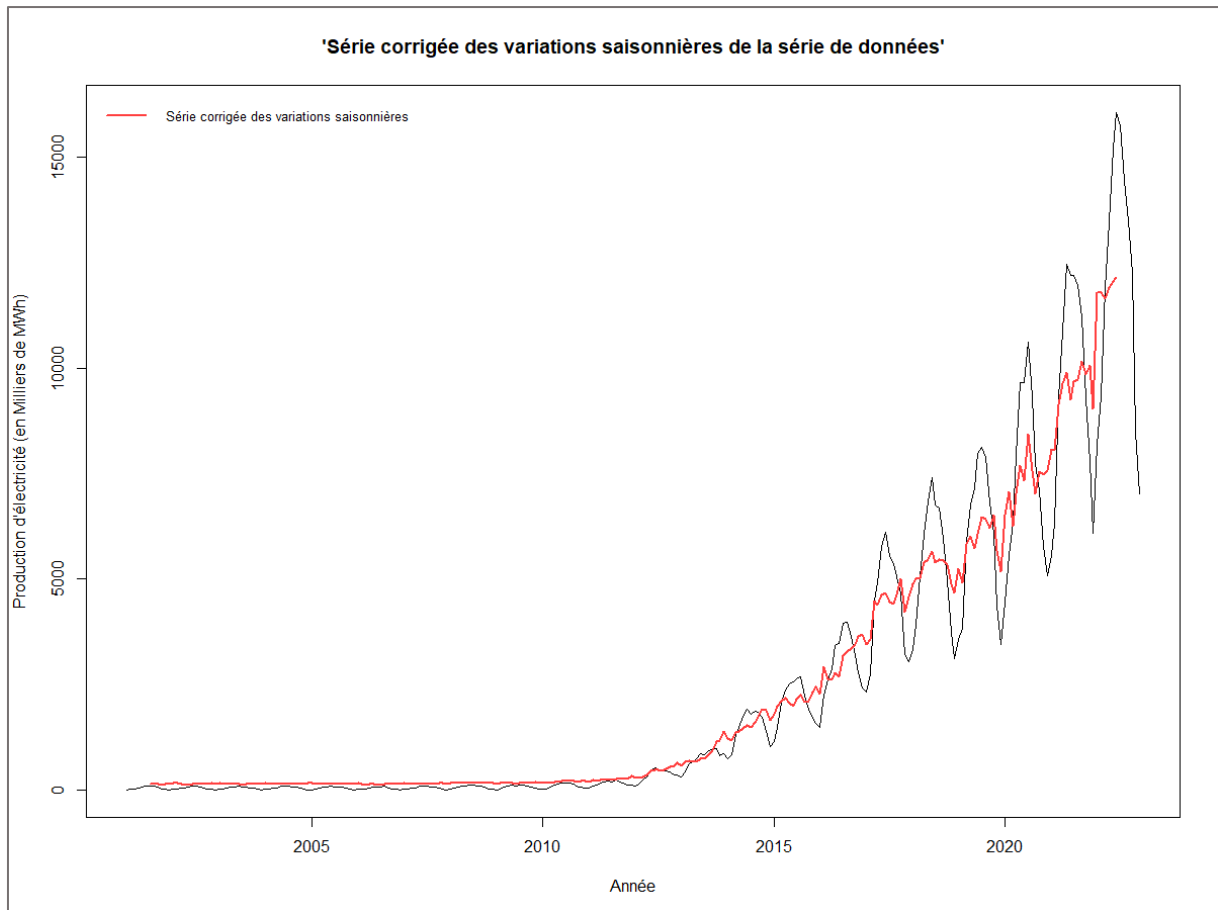
expliquée par la durée croissante de lumière totale du jour entre le solstice d'hiver et le solstice d'été, et inversement, une décroissance de cette lumière totale du solstice d'été jusqu'au solstice d'hiver. Ainsi, au cours du premier semestre de l'année, une augmentation progressive de l'illumination solaire est observée, tandis qu'une diminution correspondante se produit au cours du second semestre.

➤ Décomposition de la série :



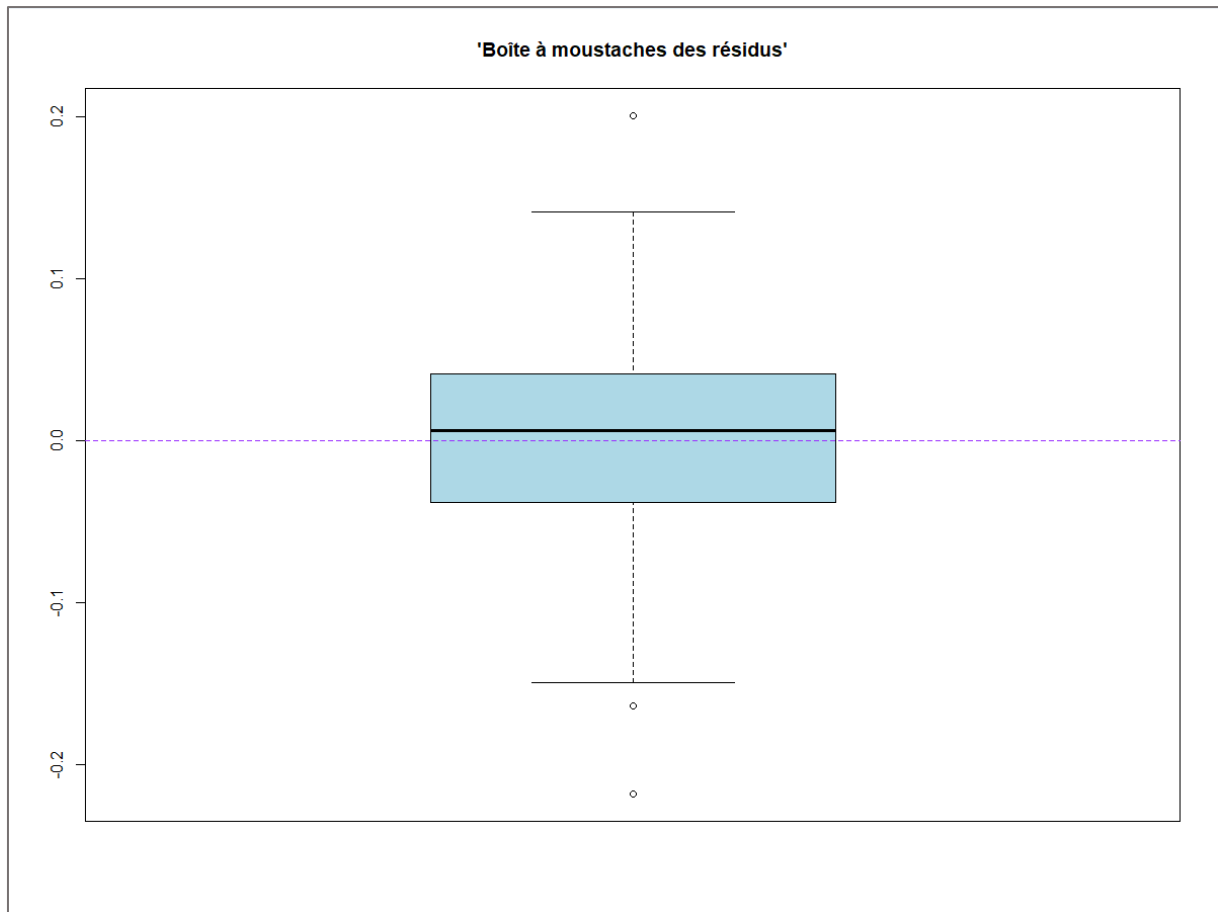
Le graphique ci-dessus représente la décomposition de la série mise sous forme logarithmique donc sous forme additive. Il combine chaque composante d'une série temporelle en affichant la série, la tendance, la saisonnalité et les résidus observés dans un seul et même graphique.

➤ Série corrigée des variations saisonnières :



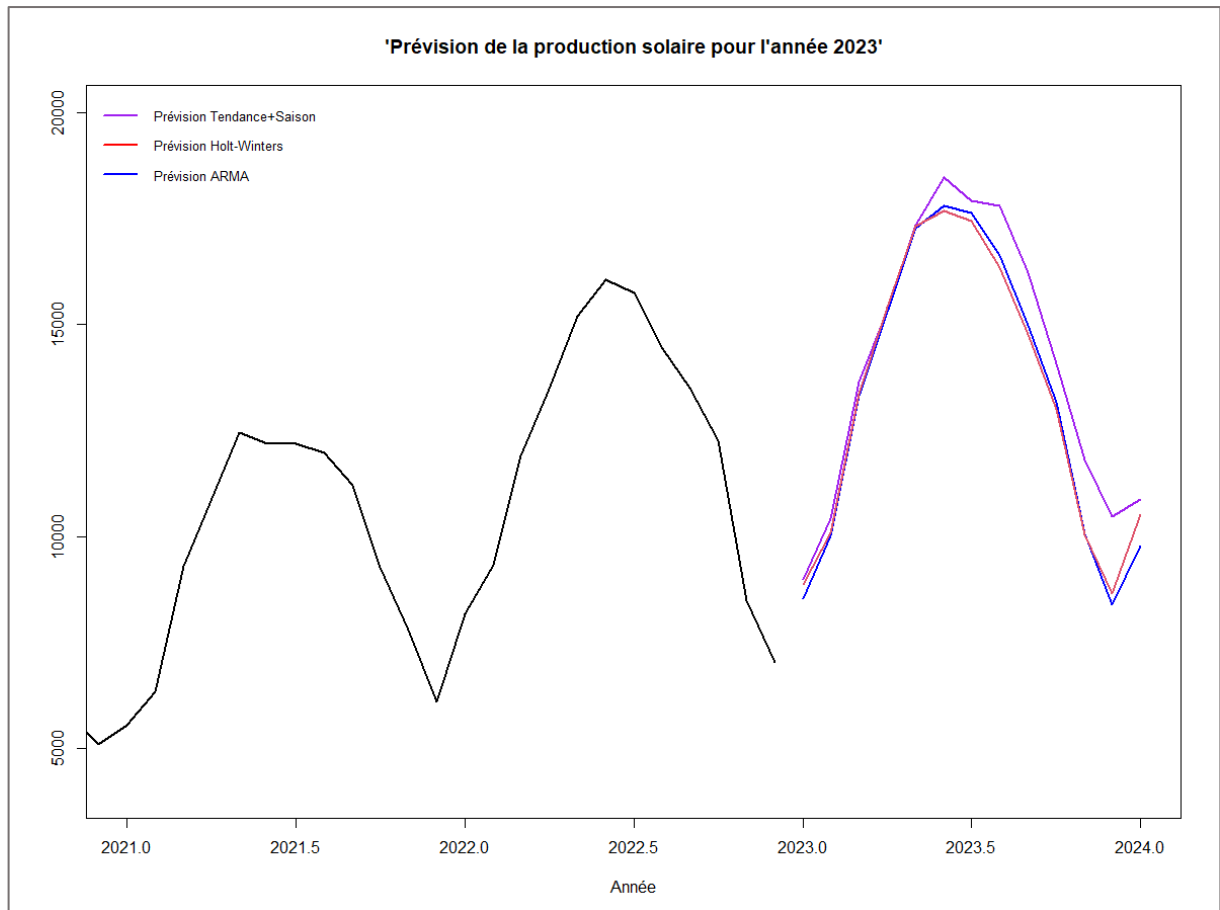
Le graphique ci-dessus représente la série corrigée des variations saisonnières de la série de données. Le graphique de la série corrigée des variations saisonnières met en évidence la tendance à long terme de la série temporelle en éliminant les effets saisonniers. Nous pouvons observer que la CVS ne parvient pas à clarifier la tendance, laissant la lisibilité floue. Les fluctuations saisonnières supprimées ne révèlent pas de manière évidente les tendances à long terme, ce qui rend difficile l'interprétation des mouvements de notre série. En effet, nous pouvons encore remarquer des diminutions sur cette courbe dans les périodes où la production d'énergie baisse, comme durant la période d'hiver 2019-2020.

➤ Boîte à moustaches des résidus :



Le graphique ci-dessus représente la boîte à moustache des résidus de la série de données. Étudier la boîte à moustaches des résidus est essentiel pour évaluer la qualité d'ajustement d'un modèle ou pour identifier des tendances non capturées. Nous pouvons observer que la boîte à moustaches indique une médiane très proche de 0, ce qui montre un ajustement assez équilibré du modèle où les erreurs de prédiction ne présentent pas de biais. La symétrie de la boîte à moustache des deux côtés de la médiane confirme aussi le bon équilibre de la série de données.

➤ Prédiction des valeurs pour 2023 :



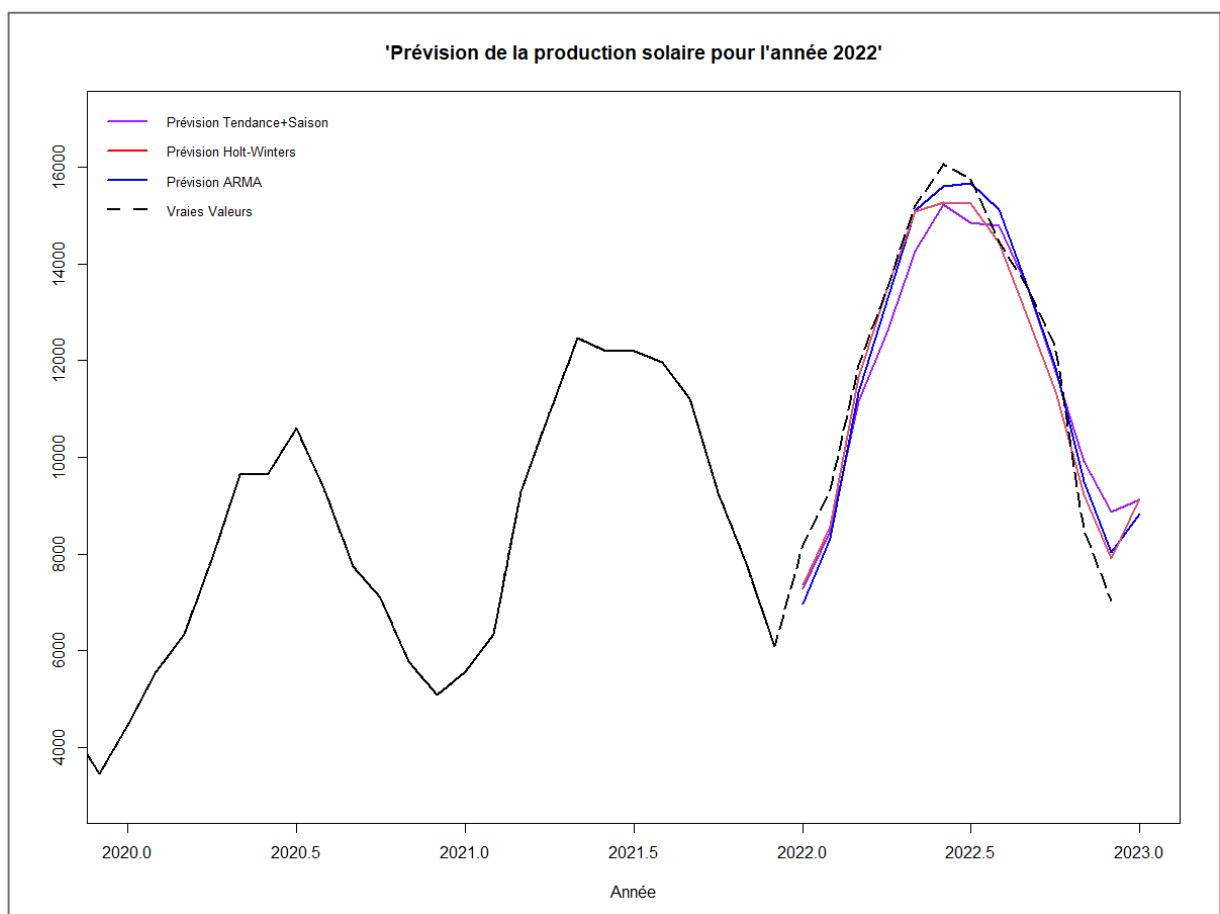
Le graphique ci-dessus représente la prévision de la production d'énergie pour l'année 2023. La prévision, en l'occurrence sur des séries temporelles, permet par exemple d'anticiper les futures tendances et ainsi de faciliter la prise de décision. Pour avoir ces résultats, nous avons utilisé 3 méthodes : une méthode « Tendence + Saison », une méthode « Holt-Winters », et une méthode « ARMA ».

La première méthode consiste à faire un ajustement polynomial, où la composante saisonnière est supprimée en calculant la CVS à partir des données log-transformées. Ensuite, une tendance paramétrique est ajustée à la série CVS en utilisant deux vecteurs, dont l'un est fonction du temps. La tendance polynomiale résultante est obtenue en combinant les coefficients saisonniers avec la tendance paramétrique ajustée. Finalement, la courbe obtenue est retransformée de manière exponentielle afin de revenir à la série de départ. La deuxième méthode consiste à ajuster les données selon la fonction « HoltWinters » de sorte à pouvoir prédire les futures valeurs. La troisième méthode, quant à elle, consiste à ajuster les données selon la fonction « arima ». Au préalable, 5 modèles ARIMA différents ont été évalués

afin de choisir le meilleur pour prédire les futures valeurs. Pour cela, le modèle présentant le plus faible AIC a été gardé. Le reste de la méthode est similaire à celle « Holt-Winters ».

Nous pouvons tout de même remarquer que la deuxième et troisième méthode montrent à peu près les mêmes prédictions de valeurs tandis que la première méthode génère des valeurs légèrement supérieures.

➤ Prédiction et comparaison des valeurs pour 2022 :



Le graphique ci-dessus représente la prévision de la production d'énergie pour l'année 2022. L'objectif de cette prédiction est d'évaluer l'exactitude des prévisions générées par rapport aux données réelles par trois méthodes distinctes, utilisées préalablement pour prédire les valeurs de l'année 2023. Pour ce faire, nous avons exclu les données de l'année 2022 afin de créer les courbes correspondantes. Nous avons ensuite superposé la courbe des valeurs réelles de 2022 à chaque méthode pour établir des comparaisons.

De manière générale, nous observons que les courbes générées par les méthodes de « Holt-Winters » et « ARMA » présentent une plus grande continuité avec les données réelles par rapport à la méthode « Tendance + Saison ». Cependant, des conclusions plus précises demeurent difficiles à faire. Pour déterminer la méthode la plus performante en termes de prédictions pour l'année 2022, nous prévoyons de calculer l'erreur quadratique moyenne respective à chaque méthode. La méthode qui affichera la plus petite erreur quadratique moyenne sera considérée comme la plus efficace.

➤ Calcul des Erreurs Quadratiques Moyennes :

```
> #### ERREUR QUADRATIQUE
>
> vrai_valeur = tail(cs3,12)
> vrai_valeur
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
2022  8171  9333 11898 13476 15198 16058 15749 14442 13504 12252  8484  7033
>
> Prediction_arma= head(exp(prev$pred),12)
> Prediction_arma
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
2022  6969.747  8319.214 11343.609 13212.272 15100.608 15597.542 15652.628 15111.640 13538.907 11855.255  9493.066  8026.477
>
> prediction_hw=head(exp(predict(cs2.hw, n.ahead=13)),12)
> prediction_hw
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
2022  7354.265  8571.540 11649.397 13464.109 15075.130 15261.319 15235.282 14435.119 12905.437 11363.663  9220.222  7913.464
>
> prediction_alternative = head(exp(z),12)
> prediction_alternative
[1] 7283.691  8475.517 11144.684 12572.557 14261.394 15226.628 14837.632 14787.573 13514.147 11778.159  9911.370  8873.410
>
> EQM=mean((vrai_valeur-Prediction_arma)^2)
> sqrt(EQM)
[1] 688.6412
>
> EQM1=mean((vrai_valeur-prediction_hw)^2)
> sqrt(EQM1)
[1] 624.9823
>
> EQM2=mean((vrai_valeur-prediction_alternative)^2)
> sqrt(EQM2)
[1] 960.6349
```

Pour calculer chaque EQM, nous avons moyenné les valeurs réelles de chaque mois de 2022 moins la valeur de leur prédiction, le tout au carré. La méthode « Holt-Winters » a affiché la plus petite erreur quadratique moyenne, indiquant ainsi son aptitude à fournir les prédictions les plus précises en termes de valeurs.

➤ Conclusion :

En conclusion, notre analyse d'une série portant sur l'évolution de la production d'énergie solaire entre 2001 et 2022 a montré une nette croissance exponentielle à partir de 2012. De manière plus détaillée, la saisonnalité a mis en évidence l'augmentation de la production pendant le premier semestre de chaque année et inversement pour le deuxième semestre.

L'étude de la série corrigée des variations saisonnières n'a globalement pas permis d'avoir une meilleure visibilité de la tendance. En revanche, la boîte à moustaches des résidus a montré une bonne symétrie et une absence de biais, donc un ajustement équilibré des erreurs.

La méthode « Tendance + Saison », « ARMA » et « Holt-Winters » ont été utilisées pour prévoir les valeurs de l'année 2023. Elles ont été réutilisées pour reprédire les valeurs de 2022 afin de s'assurer de leur fiabilité tout en les comparant. La méthode « Holt-Winters » était la plus efficace dans nos prévisions grâce à sa plus petite erreur quadratique moyenne.

➤ Summary :

In conclusion, our analysis of a series depicting the evolution of solar energy production between 2001 and 2022 revealed a clear exponential growth starting from 2012. In more detail, seasonality highlighted an increase in production during the first half of each year and vice versa for the second half.

The study of the seasonally adjusted series did not significantly improve the visibility of the trend overall. However, the boxplot of the residuals exhibited good symmetry and an absence of bias, indicating a balanced fit of the errors.

The "Trend + Season," "ARMA," and "Holt-Winters" methods were employed to forecast values for the year 2023. They were then reused to predict the values for 2022 to ensure their reliability while making comparisons. The "Holt-Winters" method proved to be the most effective in our forecasts due to its smaller mean squared error.