



Description et prévision de l'énergie géothermique

Chen Laurent – Chen Patrick
2AFA VCOD

**SAE « Description et Prévision
de séries temporelles »**



IUT de Paris - Rives de Seine
Université Paris Cité

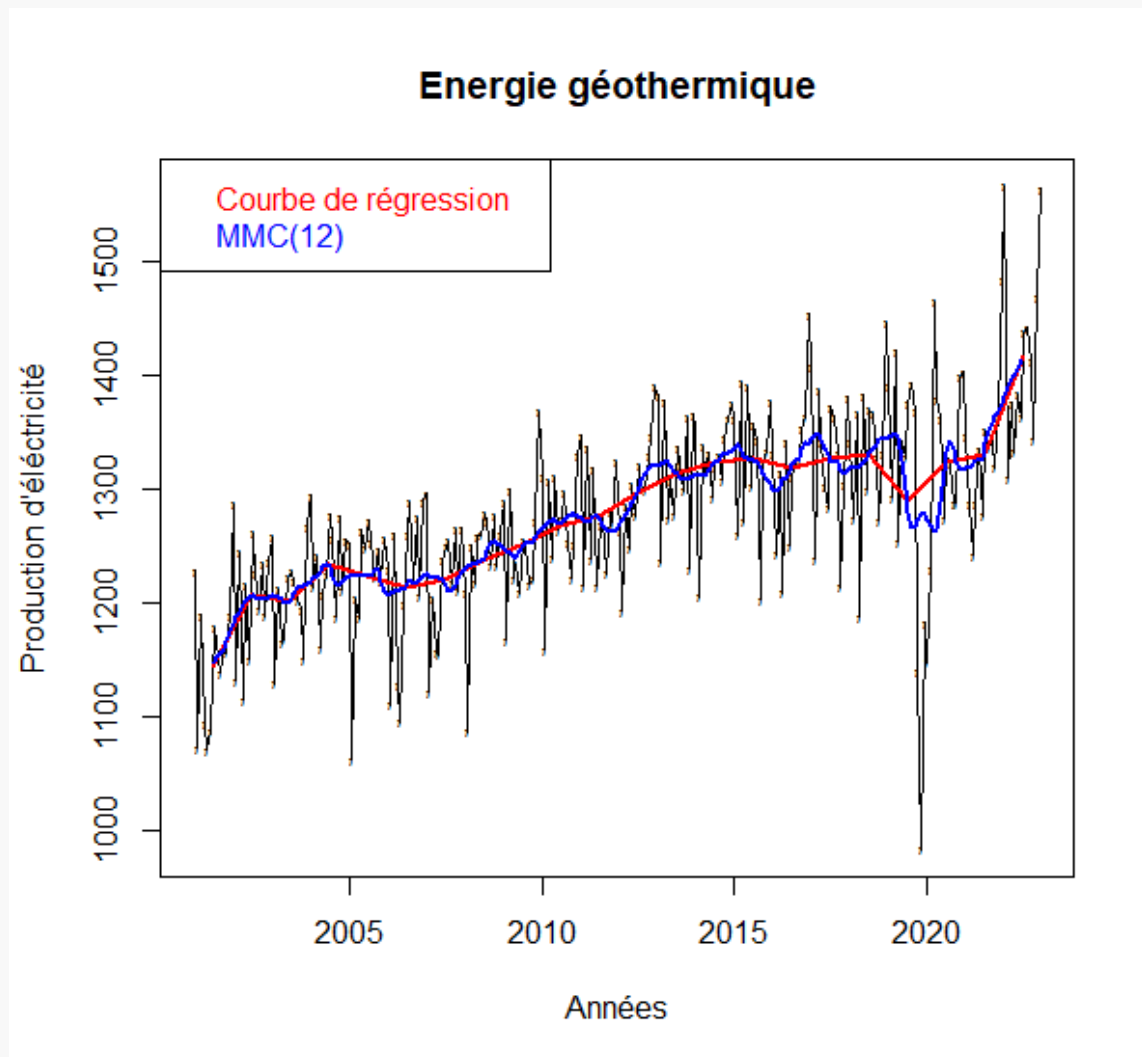
Contexte

L'U.S Energy Information Administration (EIA) est une agence fédérale chargée de la statistique. Cette agence rassemble données sur les réserves d'énergie, la production, la consommation ou encore la distribution. Ces données couvrent plusieurs type de production d'énergie, telles que : l'énergie fossile (charbon, gaz, ...), l'énergie renouvelable ou bien l'énergie nucléaire sur une centaine d'années.

Dans le cadre d'une étude prévisionnelles des données fournies par l'EIA, nos travaux se baseront uniquement sur la production d'énergie géothermique (en mégawatt/heure) au cours des 22 dernières années, soit entre les périodes de 2001 à fin 2022. Nous étudierons les évolutions de la production de l'énergie géothermique, à savoir les variations ainsi que la tendance de la production. Nos analyses nous permettront de réaliser des prédictions sur l'année 2023 aux travers de 3 méthodes prédictives et de vérifier leur fiabilité.

Tendance de la série

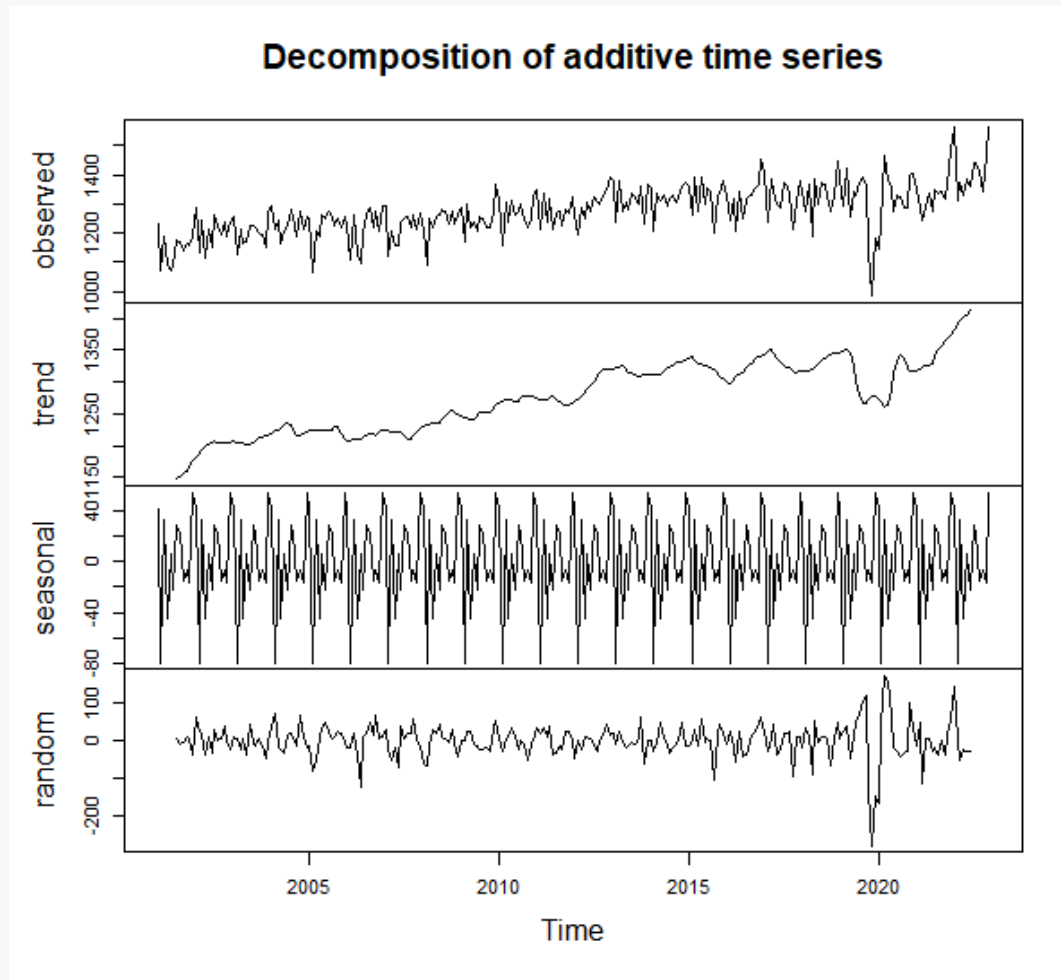
Régression et moyenne mobile centrée



Tout d'abord, on remarque que la série est additive. On observe une tendance croissante qui est plutôt irrégulière. On peut tout de même apercevoir un motif saisonnier sur la courbe MMC(12). Néanmoins, il y a une baisse importante vers 2019-2020, qui peut-être expliqué par la pandémie du Covid-19.

Décomposition de la série

Caractéristiques



Ici, nous avons 4 courbes, la 1ère courbe représente l'observation de nos données, quant aux 3 autres elles représentent la décomposition de la 1ère. Parmi ces 3 courbes nous avons : "trend", "seasonal" et "random" qui désignent respectivement la tendance, la saisonnalité et les résidus.

- La courbe de la tendance représente l'évolution générale de la série. La courbe indique une tendance croissante et les valeurs se concentrent entre 1150 et 1450 MWh.

Décomposition de la série

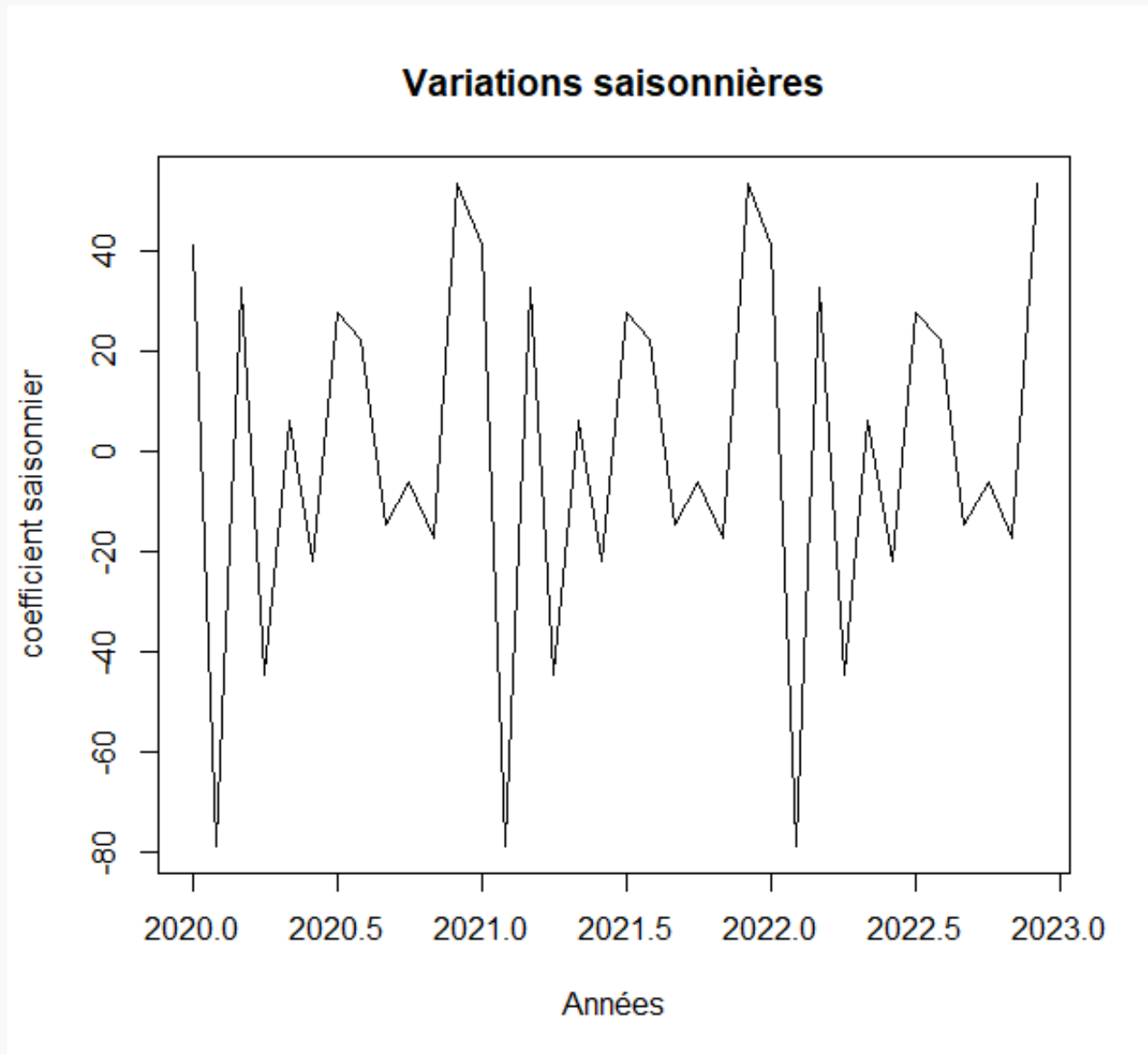
Caractéristiques

- La courbe de la saisonnalité représente les variations des valeurs sur une période de temps définie. Cette courbe aide à distinguer les variations régulières de la tendance générale.
- La courbe des résidus représente les variations inexplicables, soit les fluctuations qu'il reste après avoir isolé la, tendance et la saisonnalité.

Nous avons supposé précédemment une baisse importante dû au Covid-19 entre 2019 et 2020, on remarque qu'une part de cette baisse est expliquée par la tendance mais surtout par les résidus.

Coefficient saisonnier

Ajustement saisonnière



On rappelle que $p=12$ ici.

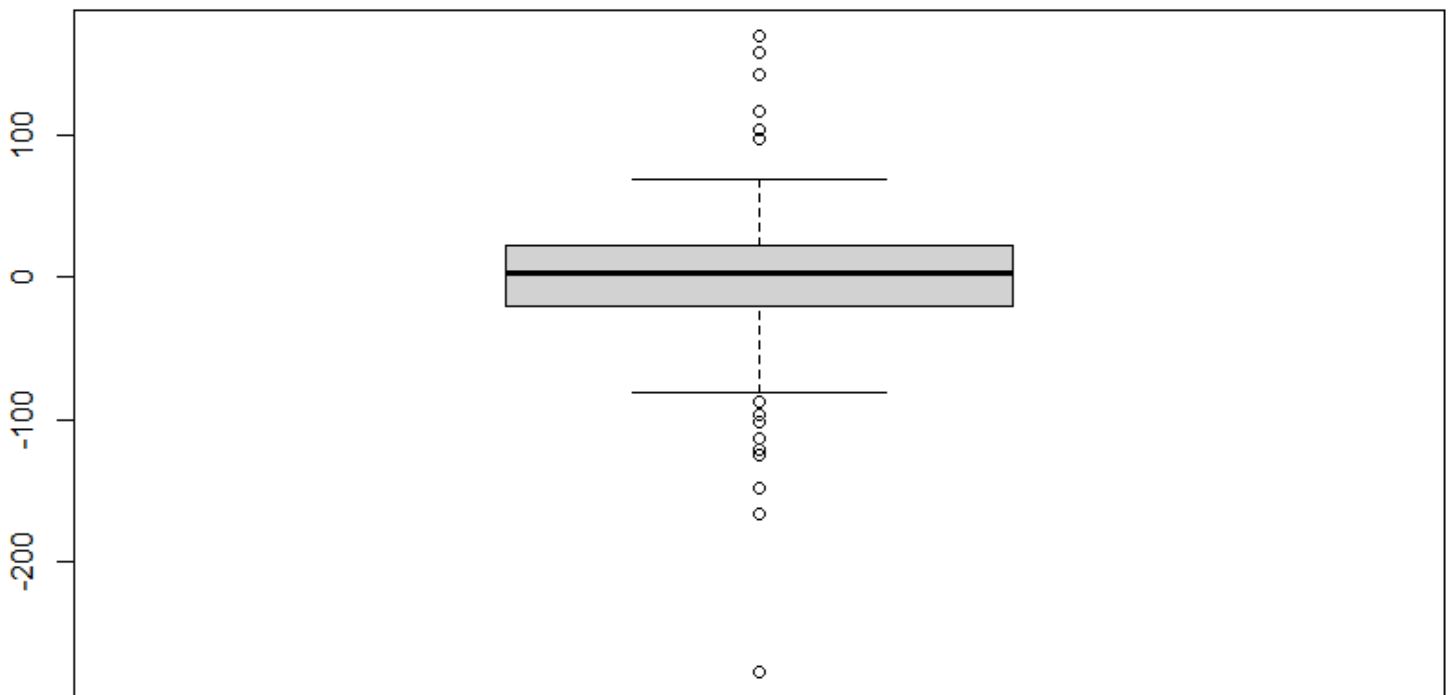
On observe un motif saisonnier qui se répète chaque année.

De plus, on remarque de nombreuses variations d'amplitudes.

Et une très forte en fin et début d'année.

Boîte à moustache des résidus

Boîte à moustache des résidus

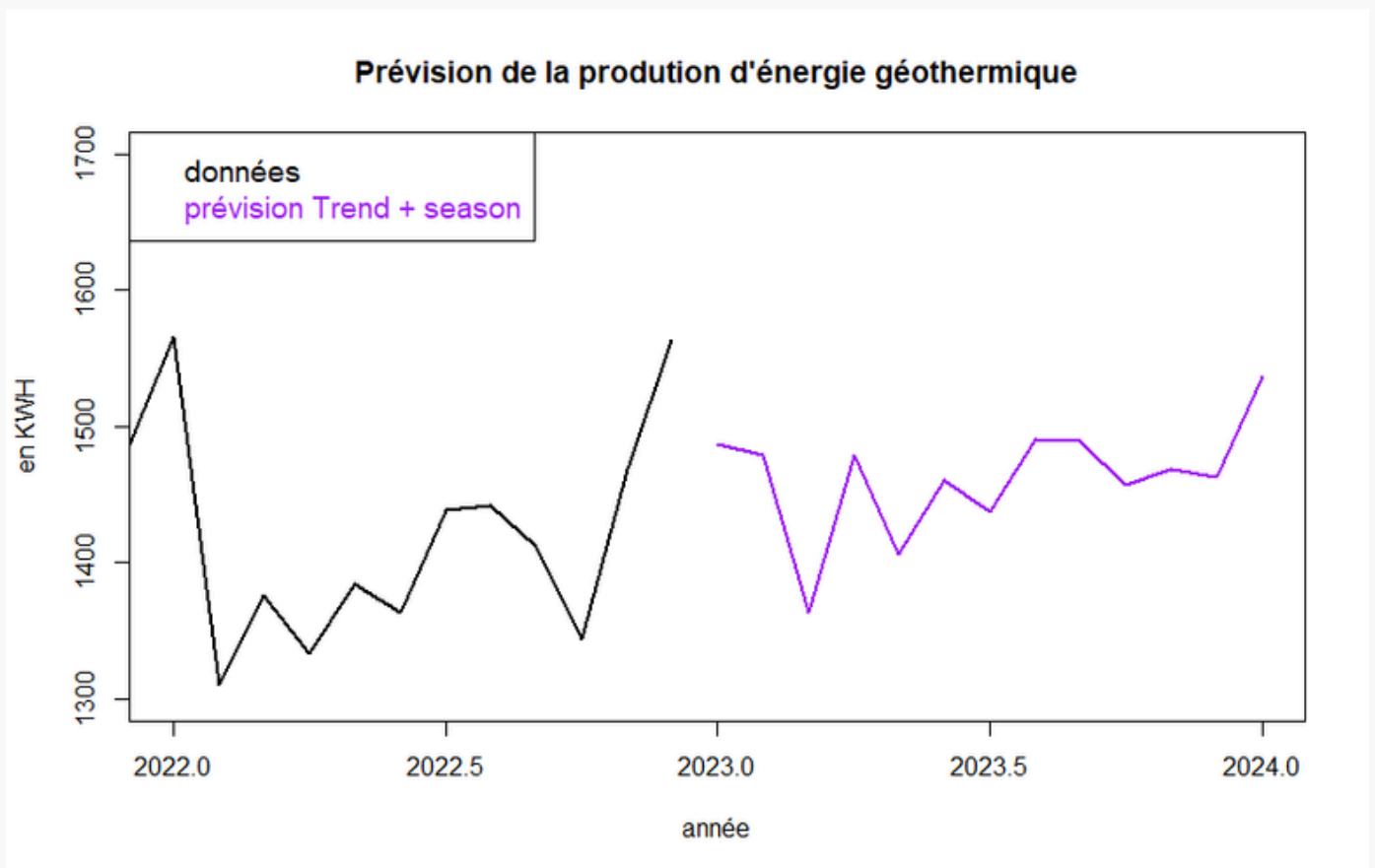


Cette boîte à moustache centrée en 0, illustre la dispersion de nos résidus. On remarque une multitude de points aux extrémités des moustaches, ces valeurs sont dites aberrantes, c'est à dire que ces valeurs sont plus extrêmes que la valeur attendue.

Prédiction de la production d'énergie pour l'année 2023

Tendance + Saisonnalité

La prédiction par Tendance + Saisonnalité consiste à calculer les coefficients saisonniers pour en déduire la série corrigée des variations saisonnières. Puis on ajuste avec la méthode des moindres carrés.

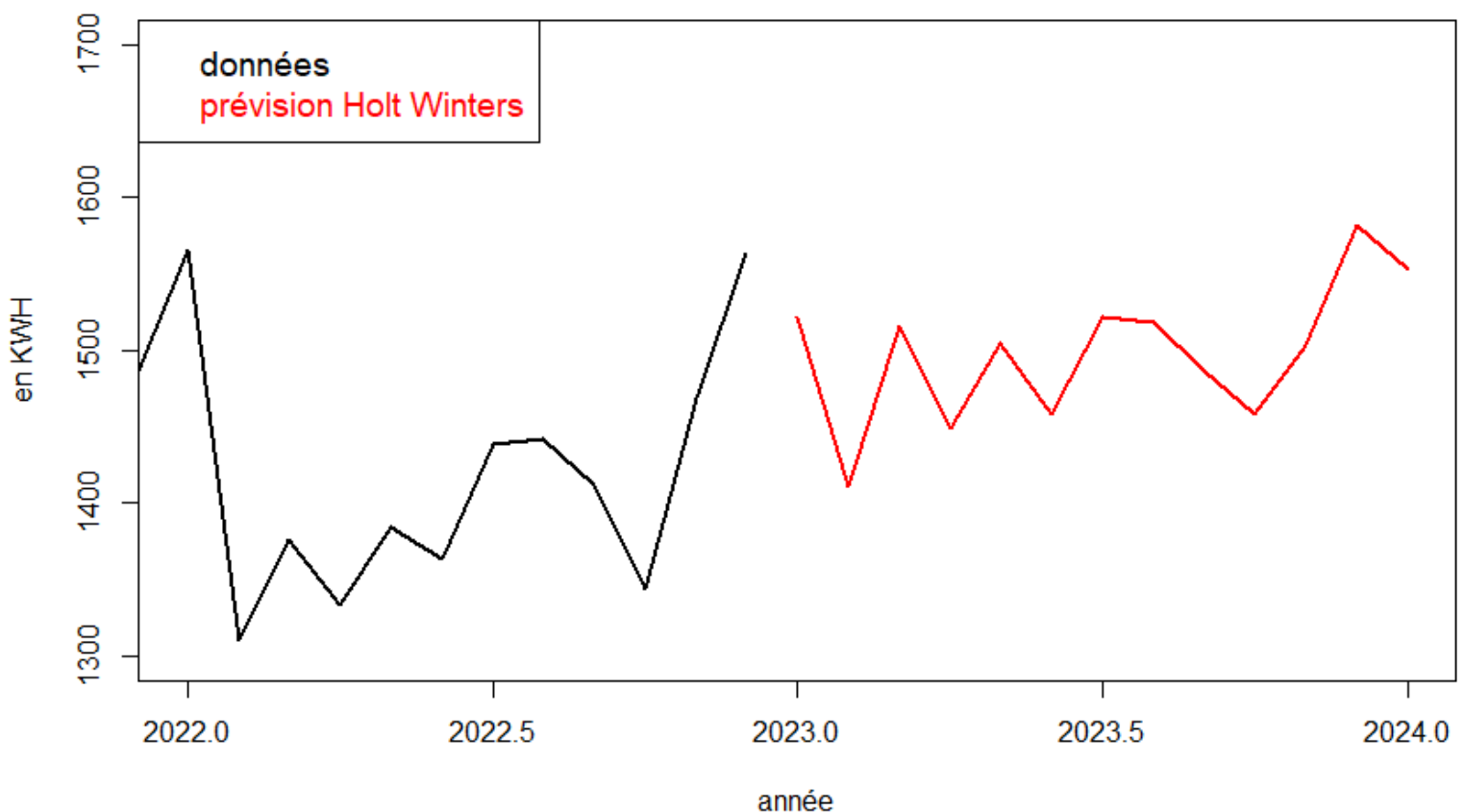


La prédiction a l'air de se rapprocher au niveau du motif saisonnier. Cependant, il n'a pas l'air d'être plus élevé que l'année précédente. Or on avait dit que la tendance était croissante.

Prédiction de la production d'énergie pour l'année 2023

Holt-Winters

Prévision de la production d'énergie géothermique



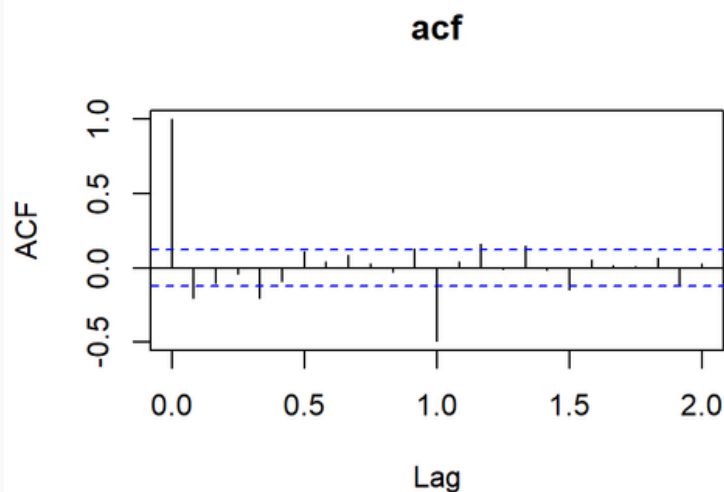
La méthode Holt-Winters est une méthode de prévision utilisée pour l'analyse et prédictions des tendances saisonnières dans les séries temporelles. Elle repose sur 3 composantes principales :

- Lissage de niveau
- Lissage de tendance
- Lissage saisonnier

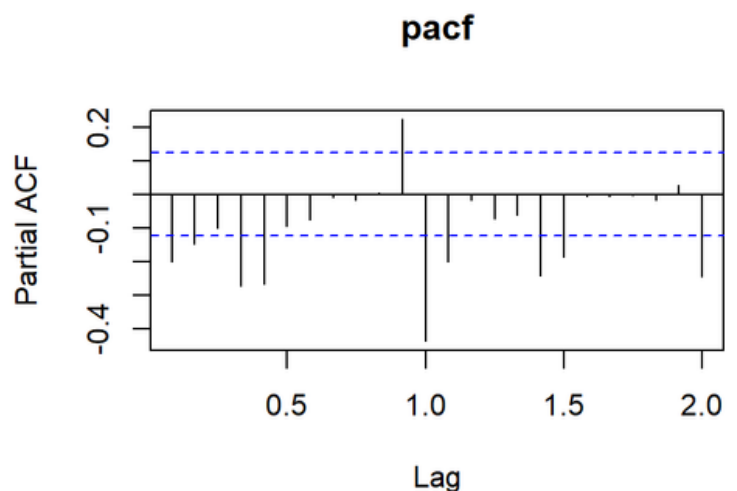
Prédiction de la production d'énergie pour l'année 2023

ARMA

Le modèle ARMA explique la variable aléatoire X_i par le passé proche et le passé à l'ordre p . Sa prédiction consiste à créer la série des différences. On vérifie que la série est une suite sans tendance ni saisonnalité. On définit les paramètres Q, q, P, p .



$Q=1, q=1$ ou 4



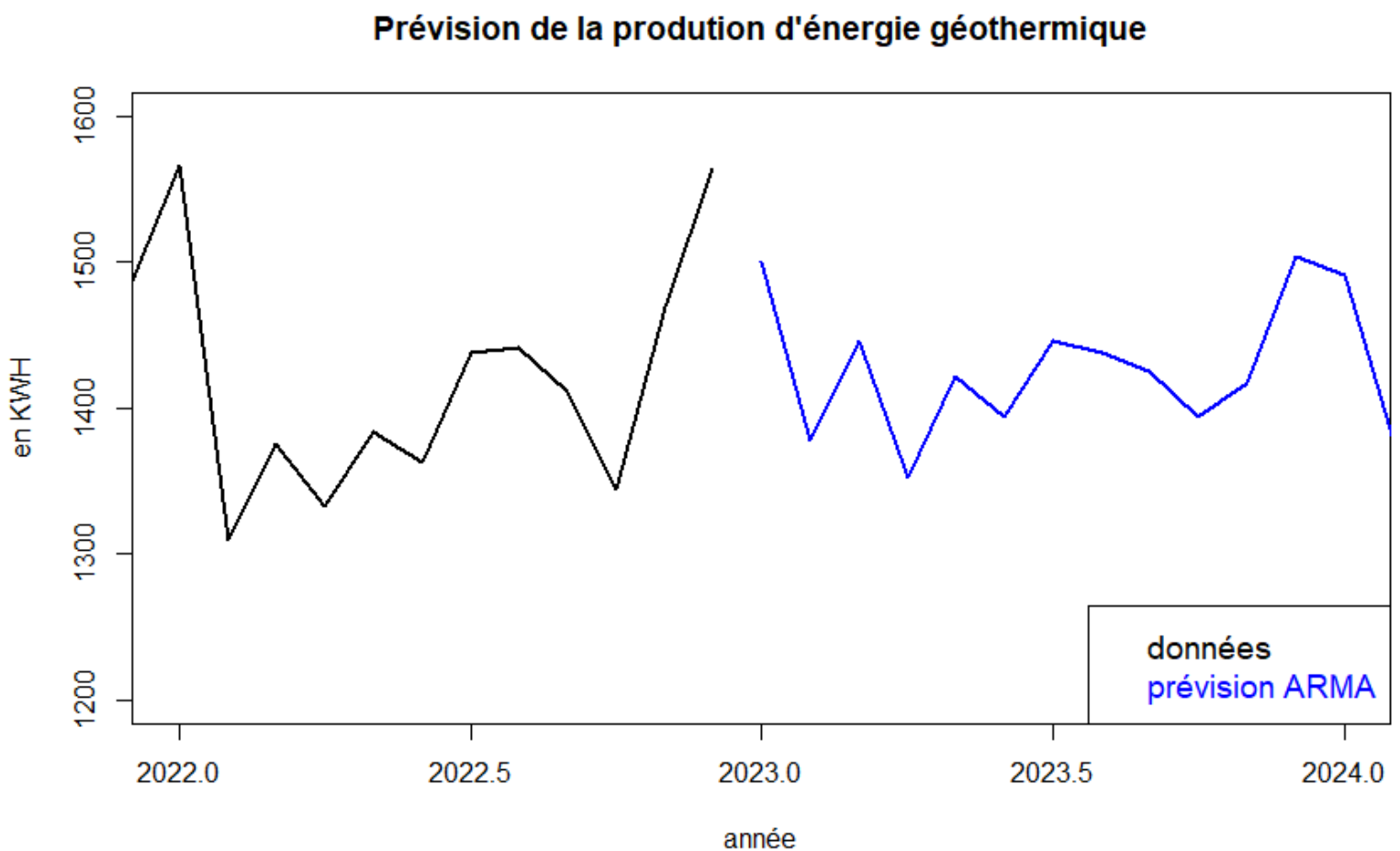
$P=2, p=1$ ou 2 ou 4 ou 5

Après plusieurs différenciations, l'AIC le plus faible est 2676.38 avec :

```
arima(data, order=c(4,1,4), seasonal=list(order=c(0,1,1), period=12))
```

Prédiction de la production d'énergie pour l'année 2023

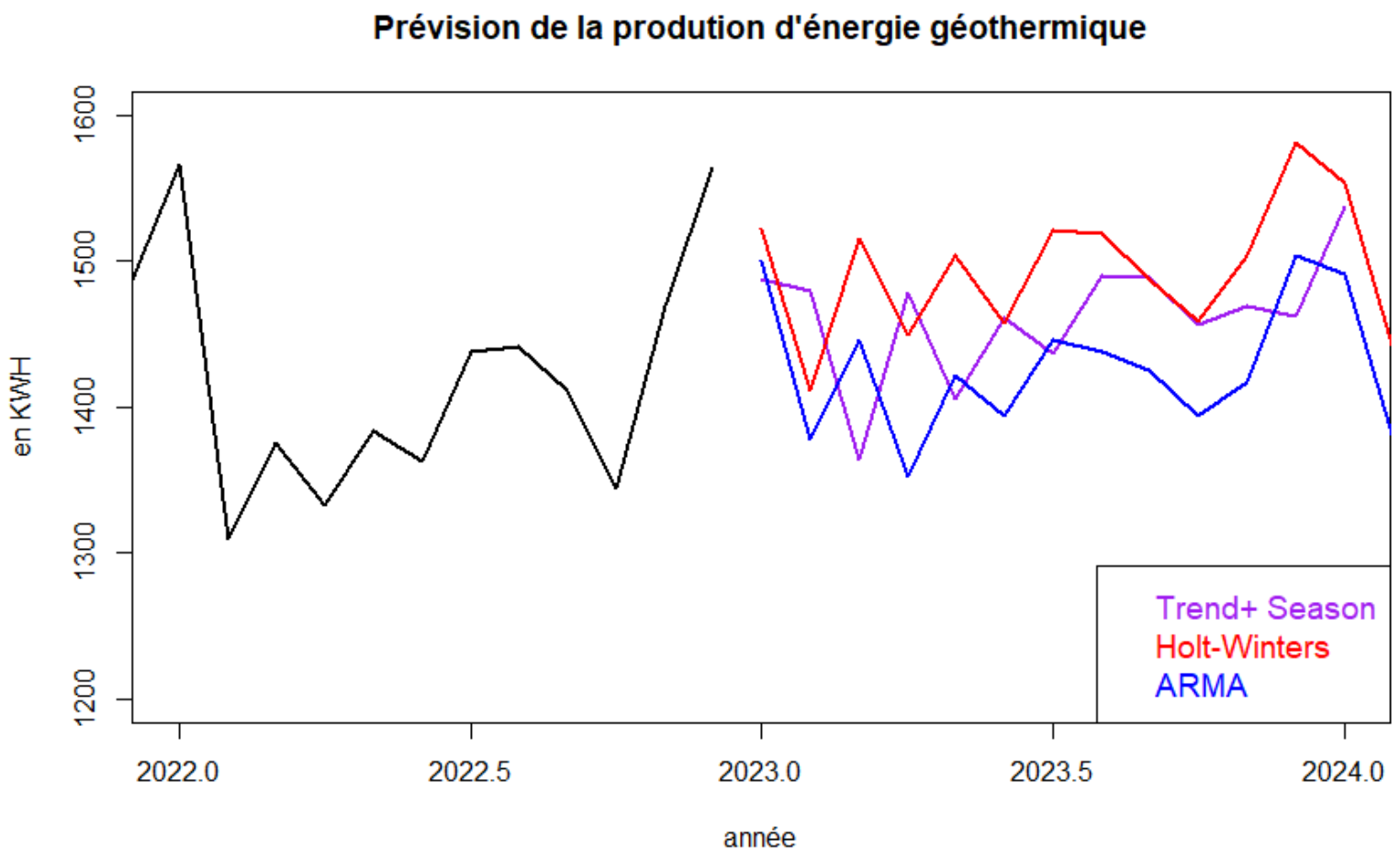
ARMA



La prédiction ARMA a l'air de se rapprocher au niveau du motif saisonnier. Cependant, comme pour la prédiction par Tendence + Saisonnalité, il n'a pas l'air d'être plus élevé que l'année précédente.

Prédiction de la production d'énergie pour l'année 2023

Régression, Holt–Winters, ARMA

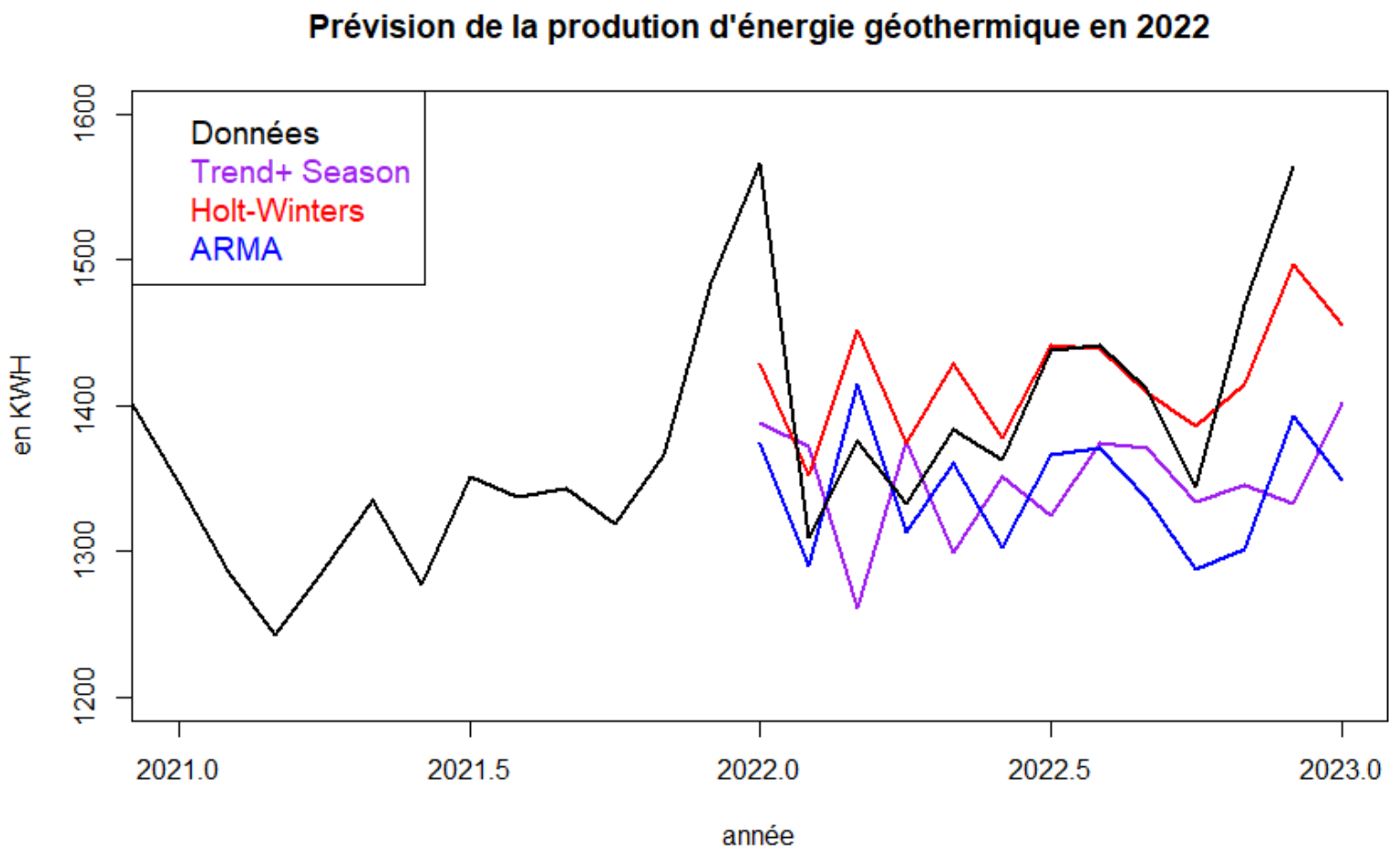


Ces trois prédictions sont différentes les unes des autres. Néanmoins, elles ont tous la même allure. Maintenant, il faudrait les comparer avec les vraies données.

Prédiction 2022 et Erreur

Quadratique Moyenne

Tendance+Saisonnalité, Holt-Winters, ARMA



On souhaite déterminer la meilleure méthode de prédiction associée à nos données. Pour cela, nous avons calculé les prédictions de l'année 2022 que l'on connaît déjà. La comparaison des prédictions par rapport aux données réelles nous permettra de définir la meilleure prédiction de nos données par l'une de ces 3 méthodes.

Prédiction 2022 et Erreur

Quadratique Moyenne

Régression, Holt–Winters, ARMA

On constate graphiquement que la méthode “Tendance+Saisonnalité” et ARMA ont tendance à sous-estimer les données réelles tout au long de l’année 2022. Contrairement aux 2 méthodes citées précédemment, la méthode Holt-Winters surestime légèrement les données réelle et épouse la courbe des données réelle en milieu d’années 2022. Pour savoir laquelle de ces 3 méthodes est susceptible de convenir au mieux à nos données, il nous faut calculer l’erreur quadratique moyenne (EQM). Plus le résultat du calcul de l’EQM est faible plus la méthode est appropriée.

Formule :

$$EQM = 1/12 \sum (y_i - \hat{y}_i)^2$$

Résultats :

```
> EQM_TrendSeason <- (sum((VraiValeurs[1:12]-ValPredits[1:12])**2))/12
> EQM_TrendSeason
[1] 12184.62
>
> EQM_HoltWinter <- (sum((VraiValeurs[1:12]-Val_HoltWinter[1:12])**2))/12
> EQM_HoltWinter
[1] 3295.28
>
> EQM_ARMA <- (sum((VraiValeurs[1:12]-Val_ARMA[1:12])**2))/12
> EQM_ARMA
[1] 9984.744
```

Prédiction 2022 et Erreur

Quadratique Moyenne

Régression, Holt–Winters, ARMA

La méthode de Holt Winters a une erreur quadratique moyenne la plus faible.

Donc ici, c'est cette méthode qui prédit le mieux. En deuxième, on a la prédiction ARMA. Et la prédiction la moins précise est Trend + Season.

Conclusion

Nos travaux consistaient à comprendre les données mis à disposition par l'EIA et fournir une prédiction sur la production d'électricité pour une année à venir. Nous avons vu que la série est additive, qu'il y a une tendance croissante plutôt irrégulière. Ces variations irrégulières causées par le Covid-19 sont expliquées par la courbe de la tendance et surtout par celle des résidus. On a remarqué, d'après la boîte à moustache qu'il y avait beaucoup de valeurs aberrantes parmi les résidus, ce qui marque des variations inattendues. Pour $p = 12$, on a observé un motif saisonnier qui se répète chaque année. Enfin, après avoir ajusté nos prédictions, nous avons vu que la méthode Holt-Winters est la meilleure, car sa courbe de prédiction épouse mieux les données réelles que les autres méthodes, mais aussi que son erreur quadratique moyenne est la plus faible.

Conclusion

Our work consisted in understanding the data made available by the EIA and providing a prediction of electricity production for a year ahead. We have seen that the series is additive, with a rather irregular upward trend. These irregular variations caused by Covid-19 are explained by the trend curve and, above all, by the residual curve. We noticed, from the boxplot, that there were many outliers among the residuals, marking unexpected variations. For $p = 12$, we observed a seasonal pattern that repeats itself every year. Finally, after adjusting our predictions, we saw that the Holt-Winters method is the best, not only because its prediction curve fits the real data better than the other methods, but also because its Mean-Squared Error is the lowest.