09/01/2025

SAE Série temporelle

Etude de production d’énergie sur EIA. Sujet : Nuclear

Matteo Cai Mathias De ridder

BUT 2 VCOD FI

Table des matières

[Introduction 2](#_Toc187274850)

Chargement des données et [Gestion des valeurs manquantes 2](#_Toc187274851)

[Analyse de tendance 3](#_Toc187274852)

[Moyennes mobiles 3](#_Toc187274853)

[Décomposition de la série 4](#_Toc187274854)

[Série désaisonnalisée 5](#_Toc187274855)

[Boxplots des résidus 6](#_Toc187274856)

Prévisions ………………………………………………………………………………………………… 7

[Évaluation des prévisions](#_Toc187274857) 8

[Conclusion](#_Toc187274856) 9

[Summary in English](#_Toc187274856) 10

SAE - Séries Temporelles

Matteo CAI et Mathias DE RIDDER

2025-01-09

L’Agence américaine d’information sur l’énergie (EIA) est une agence principale du Système Statistique Fédéral des États-Unis, responsable de la collecte, de l’analyse et de la diffusion d’informations sur l’énergie. Ses programmes couvrent des données sur le charbon, le pétrole, le gaz naturel, l’électricité, les énergies renouvelables et nucléaires. Nous nous intéressons ici aux séries temporelles de l’analyse et la prévision de la série temporelle de la production nucléaire aux États-Unis sur la période de 2001 à 2023. Nous avons utilisé plusieurs méthodes statistiques pour traiter les données, identifier les tendances et saisons, et effectuer des prévisions pour l’année 2024. Les données sont exprimées en milliers de mégawatt-heures.

**Introduction**

Cette étude se concentre sur l’analyse et la prévision de la production d’énergie nucléaire aux États-Unis, en utilisant des séries temporelles sur la période de 2001 à 2023. L’objectif est de comprendre les tendances et de réaliser des prévisions pour l’année 2024 à l’aide de différentes méthodes statistiques.

**Chargement et préparation des données**

Nous avons chargé le fichier texte nommé "NuclearUSA01-23.txt" et l'avons transformé en série temporelle avec la commande ts.

# Charger les données (supposant que le fichier est déjà téléchargé)  
# Remplacer le chemin par le chemin réel vers votre fichier  
data <- read.table("data\_ratp/NuclearUSA01-23.txt")  
ts\_nuclear <- ts(data, start = c(2001, 1), frequency = 12)

**Gestion des valeurs manquantes**

Nous avons vérifié la présence de valeurs manquantes dans chaque série temporelle :

# 1. Traiter les données : gestion des valeurs manquantes avec justification  
sum(is.na(ts\_nuclear))

## [1] 0

Aucune valeur manquante n’a été détectée dans la série temporelle.

# 

# **Analyse de tendance**

## **Moyennes mobiles**

Un filtre de moyennes mobiles sur 12 mois a été utilisé pour lisser les variations saisonnières et identifier la tendance.

# Tendance : Moyennes mobiles et régression annuelle  
moy\_mobile <- stats::filter(ts\_nuclear, filter = rep(1/12, 12),sides = 2)  
moy\_annuelle <- aggregate(ts\_nuclear, nfrequency = 1, FUN = mean)

Puis nous avons calculé la moyenne annuelle pour chaque année et a tracé la courbe de régression pour visualiser la tendance.

# Tracer la tendance  
plot(ts\_nuclear, main="Série Temporelle de la Production Nucléaire", col="blue")  
lines(moy\_mobile, col="red")  
lines(moy\_annuelle, col="green")  
legend("topright", legend=c("Données", "Moy. Mobile", "Moy. Annuelle"), col=c("blue", "red", "green"), lty=1)

Une image contenant texte, capture d’écran, Police, ligne

Description générée automatiquement

# On observe que les séries de moyennes mobiles (en rouge pour la moyenne mobile simple et en vert pour la moyenne annuelle) ne présentent pas clairement de composante saisonnière. Les moyennes mobiles d’un ordre correspondant à la période de la saisonnalité permettent d’atténuer cette composante saisonnière dans la série initiale. En analysant la série brute (en bleu), on distingue une forte composante saisonnière marquée par des fluctuations régulières. En revanche, les séries lissées (en rouge et en vert) mettent en évidence la tendance générale de la production nucléaire. Cette tendance révèle des phases de croissance et de décroissance, probablement influencées par des facteurs externes tels que les politiques énergétiques, les événements climatiques ou les variations de la demande énergétique.

# **Décomposition de la série**

Nous avons décomposé la série temporelle en trois composantes distinctes : la tendance, la saisonnalité et les résidus.

# 3. Coefficients saisonniers et décomposition de la série  
dec <- decompose(ts\_nuclear)  
plot(dec)

Une image contenant texte, Police, capture d’écran, ligne

Description générée automatiquement

Le graphique illustre une décomposition additive de la série temporelle en quatre composantes : la série observée, la tendance, la saisonnalité et l’aléatoire. La série observée présente des variations saisonnières régulières. La tendance, représentée par la moyenne mobile centrée, augmente jusqu’en 2010 avant de diminuer progressivement. La composante saisonnière reste stable et se répète de manière similaire chaque année.

# **Série désaisonnalisée**

serie\_desais <- ts\_nuclear - dec$seasonal  
plot(serie\_desais, main="Série Désaisonnalisée")

Une image contenant texte, Police, capture d’écran, diagramme

Description générée automatiquement

Une fois la composante saisonnière retirée, on observe une hausse progressive de la production jusqu’en 2010, suivie d’une phase de déclin et d’instabilité, marquée par des pics prononcés entre 2015 et 2020.

# 

# **Boxplots des résidus**

residus <- ts\_nuclear - (dec$trend+dec$seasonal)  
boxplot(residus, main="Boxplot des Résidus")

Une image contenant texte, capture d’écran, diagramme, Rectangle

Description générée automatiquement

# Commentaire : Les résidus semblent relativement centrés autour de zéro avec quelques valeurs extrêmes

La répartition des résidus est centrée autour de zéro, avec la majorité des valeurs situées à l’intérieur de la boîte et entre les moustaches, ce qui indique que la plupart des prédictions sont proches des valeurs observées. Toutefois, certaines valeurs en dehors des moustaches, appelées outliers, reflètent des prédictions moins précises. Ces outliers se concentrent principalement au-dessus de 2000 et en dessous de -6000.

# **Prévisions**

Pour les prévisions de 2024, trois méthodes seront utilisées : ARIMA, Holt-Winters et la méthode linéaire.

train\_data <- window(ts\_nuclear,end=c(2022, 12))  
test\_data <- window(ts\_nuclear,end=c(2023, 1))  
  
# Méthode 1 : Modèle ARIMA  
arima\_model <- auto.arima(train\_data)  
arima\_forecast <- forecast(arima\_model, h=12)  
  
# Méthode 2 : Modèle exponentiel Holt-Winters  
hw\_model <- HoltWinters(train\_data)  
hw\_forecast <- forecast(hw\_model, h=12)  
  
# Méthode 3 : Modèle linéaire  
seq\_2024 <- seq(2023,2024, by = 1/12)  
trend <- time(train\_data)  
ML <- lm(train\_data~trend)  
predict\_lineaire <- predict(ML, newdata = data.frame(trend = seq\_2024))

Afin de mieux visualiser et comparer les différentes méthodes de prévision, elles seront représentées sur le même graphique.

# Tracer les prévisions avec commentaires  
plot(ts\_nuclear,xlim = c(2022,2025), main="Prévisions")  
lines(arima\_forecast$mean, col="red")  
lines(hw\_forecast$mean, col="green")  
lines(seq\_2024, predict\_lineaire, col="blue")  
legend("topright", legend = c("Données", "ARIMA", "Holt-Winters", "Linéaire"), col = c("black", "red", "green", "blue"), lty = 1)

Une image contenant texte, diagramme, capture d’écran, ligne

Description générée automatiquement

Les prévisions des modèles ARIMA (en rouge), Holt-Winters (en vert) et linéaire (en bleu) sont tracées sur le même graphique. On observe que les modèles ARIMA et Holt-Winters suivent assez bien les fluctuations des données, tandis que le modèle linéaire offre une prévision constante. Ce dernier semble être le moins adapté aux variations de la courbe. En revanche, les deux autres modèles sont plus cohérents. Toutefois, il est important d'examiner leurs valeurs d'EQM pour une évaluation plus précise.

# 

# **Évaluation des prévisions**

Les données de 2001 à 2022 sont utilisées pour l'entraînement, tandis que l'année 2023 est utilisée pour les tests.

# Calcul des EQM  
arima\_test\_pred <- forecast(arima\_model, h = length(test\_data))  
hw\_test\_pred <- forecast(hw\_model, h = length(test\_data))  
linear\_test\_pred <- predict(ML, newdata = data.frame(trend = time(test\_data)))  
  
mse\_sarima <- mean((test\_data - arima\_test\_pred$mean)^2)  
mse\_hw <- mean((test\_data - hw\_test\_pred$mean)^2)  
mse\_linear <- mean((test\_data - linear\_test\_pred)^2)  
  
  
# Afficher les erreurs sous forme de tableau comparatif  
mse\_table <- data.frame(  
 Méthode = c("ARIMA", "Holt-Winters", "Naïve Saisonnière"),  
 ErreurQuadratiqueMoyenne = c(mse\_sarima, mse\_hw, mse\_linear)  
)  
print(mse\_table)

## Méthode ErreurQuadratiqueMoyenne  
## 1 ARIMA 3039.642  
## 2 Holt-Winters 59291.508  
## 3 Naïve Saisonnière 22953781.406

Le modèle ARIMA semble offrir la meilleure performance sur l'année de test, mais une validation croisée serait nécessaire pour confirmer cette conclusion.

**Conclusion**

Cette étude a permis d’analyser et de prévoir la production d’énergie nucléaire aux États-Unis sur la période 2001-2023. Grâce à différentes méthodes statistiques (moyennes mobiles, décomposition des séries temporelles, désaisonnalisation et prévisions via ARIMA, Holt-Winters et régressions linéaires), nous avons identifié des tendances, des composantes saisonnières et évalué la précision des prévisions.

Les résultats montrent que :

* La tendance de production a fluctué avec une augmentation jusqu’en 2010, suivie d’un déclin progressif.
* Le modèle ARIMA s’est révélé le plus performant, avec l’erreur quadratique moyenne la plus faible sur les données de test de 2023.
* Les prévisions pour 2024 sont cohérentes avec les données historiques, bien que les variations saisonnières doivent être surveillées de près.

Ces résultats mettent en évidence l’importance d’un modèle adapté pour capturer les fluctuations complexes de la production d’énergie nucléaire.

**Répartition des tâches**

**Mathias De Ridder :**

* Écriture du premier code pour l’analyse des séries temporelles, incluant le traitement des données, l’implémentation des modèles de prévision (ARIMA, Holt-Winters, linéaire) et les visualisations.
* Réécriture et correction du rapport, en s'assurant de la cohérence et de la clarté des explications techniques.

**Matteo Cai :**

* Vérification du 1er code et correction de celui-ci, en testant les fonctions et en validant les résultats obtenus.
* Rédaction d’un premier rapport qui a servi de base pour le document final.

Cette collaboration a permis de combiner des compétences complémentaires pour produire un travail rigoureux et précis.

**Summary in English**

This study aimed to analyze and forecast nuclear energy production in the United States using time series data from 2001 to 2023. The work involved several key steps: preparing and cleaning the data, identifying trends and seasonal components, and applying various forecasting methods to predict production for 2024. After ensuring that the dataset contained no missing values, moving averages and annual regressions were used to observe trends. These analyses revealed an increase in production until 2010, followed by a gradual decline influenced by external factors such as energy policies and demand variations.

The time series decomposition highlighted three main components: trend, seasonality, and residuals. The deseasonalized series provided a clearer view of production changes over time. Forecasting models, including ARIMA, Holt-Winters, and linear regression, were implemented and evaluated. Among these, ARIMA emerged as the most accurate model, achieving the lowest mean squared error on test data. The forecasted values for 2024 align closely with historical patterns, demonstrating the reliability of the chosen methods.

This project also involved a collaborative approach. Mathias De Ridder initially developed the code for analysis and model implementation, and later revised and refined the final report. Matteo Cai verified the corrected code for consistency and accuracy while preparing a preliminary version of the report. This teamwork ensured a comprehensive and methodologically sound exploration of nuclear energy trends in the United States.