

État de l'art des méthodes de prévision

0.1 Introduction

La prédiction de la consommation électrique est cruciale pour les fournisseurs d'électricité afin de garantir un approvisionnement stable et efficace. L'utilisation de modèles de séries temporelles, tels que ARMA, ARIMA, SARIMA, VARIMAX, ainsi que des méthodes avancées telles que Prophet et XGBoost, peut aider à réaliser des prédictions précises. Cette veille explore ces modèles et méthodes pour la prédiction de la consommation électrique.

0.2 Définition

On appelle série chronologique (série temporelle ou encore chronique) une série statistique à deux variables (t, X_t) avec $t \in T$, $T = t_1, t_2, \dots, t_n$ où la première composante du couple t est le temps et la deuxième composante est une variable numérique X_t prenant ses valeurs aux instants t .

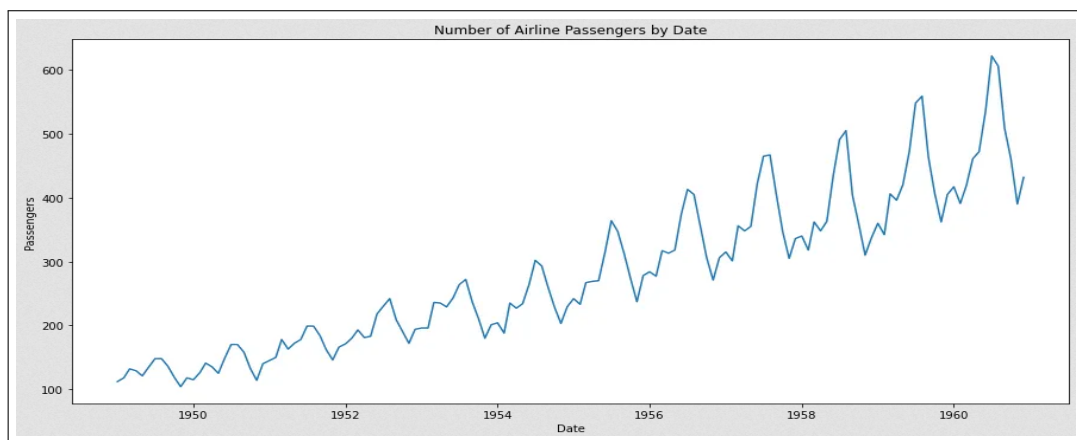


Figure 1: Exemple "Airline Passengers Line Plot — By Author".

0.3 Décomposition d'une série chronologique

Il en existe essentiellement deux grands types :

Schéma additif

Dans un modèle additif, on suppose que les 3 composantes : tendance, variations saisonnières et variations accidentelles sont indépendantes les unes des autres.

On considère que la série X_t s'écrit comme la somme de ces 3 composantes :

$$X_t = f_t + S_t + \epsilon_t$$

Graphiquement, l'amplitude des variations est constante autour de la tendance

Schéma multiplicatif

Graphiquement, l'amplitude des variations (saisonniers) varie.

On suppose que les variations saisonnières et les variations accidentelles dépendent de la tendance et on considère que X_t s'écrit de la manière suivante :

$$X_t = f_t * S_t * \epsilon_t$$

Composantes d'une série temporelle

Une série temporelle peut être décomposée en 3 composantes :

- **Tendance:** Mouvement de hausse et de baisse des données au cours du temps sur une longue période (par exemple le prix de l'immobilier).
- **Saisonnalité:** Variance saisonnière (par exemple une augmentation de la demande de glace pendant l'été).
- **Bruit:** Des pics et des creux à intervalles aléatoires.

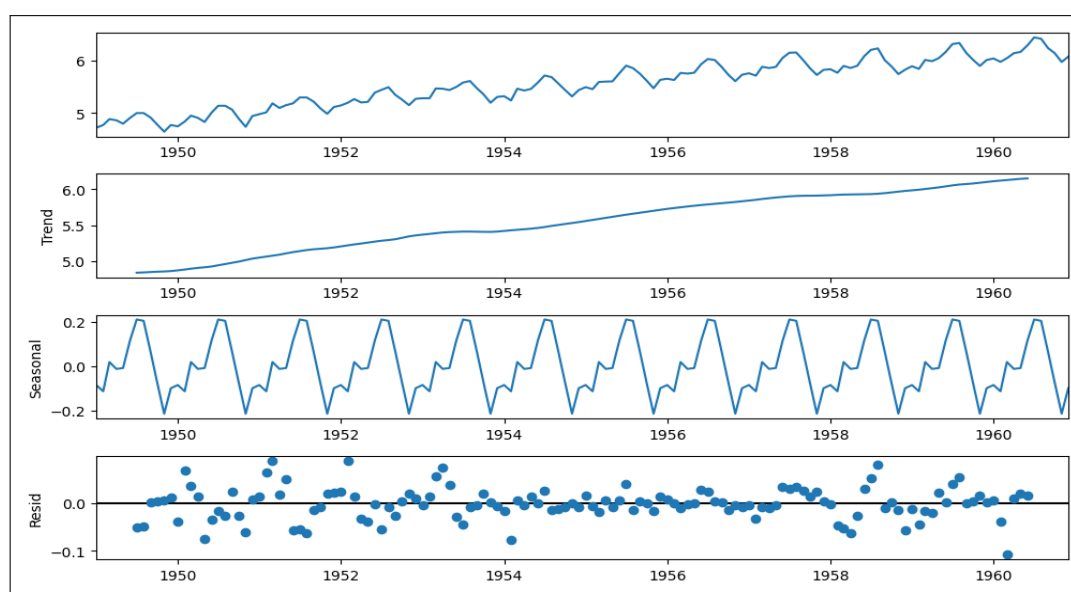


Figure 2: Décomposition de la série temporelle.

0.4 Processus stationnaire

Une série temporelle stationnaire est caractérisée par sa stabilité statistique au fil du temps. Ses propriétés, telles que la moyenne et la variance, restent constantes, ce qui simplifie l'analyse et la modélisation. Les relations entre les données passées et futures demeurent invariables.

0.5 Les modèles de séries temporelles

• Processus Auto Régressif d'ordre p , AR(p)

Dans le processus autorégressif d'ordre p , l'observation présente X_t est générée par une moyenne pondérée des observations passées jusqu'à la p -ième période sous la forme:

$$X_t = \mu + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t.$$

où $\varphi_1, \dots, \varphi_p$ sont les paramètres du modèle, μ est une constante et ε_t un bruit blanc.

p : ordre de la partie autorégressive (AR), l'observation courante est la combinaison linéaire de ses instants passés (1 à p).

- **Processus Moyen mobile d'ordre q, MA(q)**

Dans le processus moyenne mobile d'ordre q, chaque observation X_t est générée par une moyenne pondérée d'aléas jusqu'à la q-ième période sous la forme :

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^q \vartheta_i \varepsilon_{t-i}$$

où $\vartheta_1, \dots, \vartheta_q$ sont les paramètres du modèle et $\varepsilon_t, \varepsilon_{t-1}, \dots$ sont encore une fois des termes d'erreur.
q : ordre de la moyenne mobile(MA)

- **Processus ARMA (p, q)**

Les processus ARMA sont donc représentatifs d'un processus généré par une combinaison des valeurs passées et des erreurs passées. Le modèle est composé donc de deux parties : une part autorégressive (AR) et une part moyenne-mobile (MA), il s'écrit sous la forme:

$$X_t = \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

⇒ Il est utile pour modéliser des séries temporelles stationnaires avec des tendances linéaires. ARMA nécessite que les données soient stationnaires.

- **Processus ARIMA (p, d, q)**

ARIMA est également connu sous le nom de moyenne mobile autorégressive intégrée. Il permet en effet de déterminer les valeurs intégrées aux séries temporelles en s'appuyant sur les précédentes valeurs observées. Avec ce système autorégressif, les fonctions prédictives peuvent être appliquées à un cas isolé. Ce modèle statistique permet surtout de suivre et d'anticiper l'évolution d'un phénomène. Le modèle Arima peut être utilisé pour les séries de nombres non saisonniers qui ne présentent pas de série d'événements aléatoires.

d : ordre de la différence ou dérivation du processus. Cela permet de rendre stationnaire la série au cas où elle ne l'était pas. Ainsi, d=0 si la série est au préalable stationnaire sinon $d \geq 1$.

⇒ ARIMA est une extension d'ARMA qui prend en compte la différenciation pour rendre les données stationnaires. Il est largement utilisé pour modéliser des séries temporelles non stationnaires en ajustant les composantes AR et MA après différenciation.

De ce fait nous avons 2 solutions possibles pour traiter cette limitation : Appliquer un ajustement saisonnier sur la série temporelle afin d'éliminer la composante saisonnière ou utiliser le modèle SARIMA.

- **Processus SARIMA(p,d,q)(P,D,Q)s**

SARIMA: Seasonal ARIMA ou ARIMA saisonnier est une extension du modèle ARIMA. Il permet de modéliser les séries temporelles comportant une composante saisonnière et désigné par 7 paramètres :

p, d, q : les même que ceux de ARIMA.

P : ordre de la partie autorégressive saisonnière.

D : ordre de la différence saisonnière.

Q : ordre de la moyenne mobile saisonnière.

s : la période de la composante saisonnière.

⇒ SARIMA étend ARIMA pour prendre en compte la saisonnalité dans les données temporelles. Il est efficace pour la prédiction de séries temporelles avec des modèles saisonniers.

- **Processus VARIMAX(Vector AutoRegressive Moving Average with Exogenous Variables)**

Le modèle VARIMAX permet de modéliser plusieurs séries temporelles en même temps, en les reliant les unes aux autres. Il peut également inclure des variables exogènes, c'est-à-dire des variables extérieures au système qui influencent les séries temporelles.

Il permet aussi d'analyser les relations de rétroaction et de causalité entre les différentes séries, ce qui est utile pour comprendre comment elles s'influencent mutuellement.

⇒ VARIMAX est une extension du modèle VAR (Vector AutoRegressive) qui inclut des variables exogènes pour améliorer la prédiction. Il est utile lorsque la consommation électrique est influencée par des facteurs externes, tels que la température. Il suppose que les séries sont stationnaires après différenciation, ce qui signifie que vous pouvez avoir besoin de différencier vos données avant de les utiliser.

- **Processus XGBoost**

XGBoost signifie eXtreme Gradient Boosting. Comme son nom l'indique, c'est un algorithme de Gradient Boosting qui peut être utilisé pour la prédiction de séries temporelles. Il est robuste, performant et peut gérer des données complexes avec des variables exogènes.

⇒ Dans le contexte de la prédiction de la consommation électrique pour la région des Hauts-de-France, XGBoost pourrait être appliqué pour améliorer la précision des prédictions en utilisant des données historiques de consommation électrique, de température et d'autres variables exogènes pertinentes. Il est particulièrement utile lorsque les relations entre les variables sont complexes et non linéaires. XGBoost est capable de modéliser ces relations et d'ajuster automatiquement les poids des différentes variables pour optimiser les prédictions.

- **Processus Prophet**

Prophet est un modèle de prévision de séries temporelles développé par Facebook. C'est une librairie open source (R et Python) de prévision des données de séries temporelles basée sur un modèle additif. Cette librairie permet d'analyser aisément des séries temporelles même pour des personnes n'ayant pas une grande expertise dans ce domaine. ⇒ Le modèle Prophet est spécialement conçu pour la prévision de séries temporelles avec des tendances saisonnières. Dans ce contexte, Prophet pourrait être utilisé pour capturer les motifs saisonniers dans la consommation électrique, par exemple, en prenant en compte la saisonnalité quotidienne, hebdomadaire et annuelle. Prophet est également capable de gérer les données manquantes et les vacances, ce qui le rend adapté à la prédiction de la consommation électrique dans une région spécifique comme les Hauts-de-France.

Conclusion générale

La prédiction de la consommation électrique dans la région des Hauts de France est essentielle pour optimiser la gestion de l'énergie. Les modèles de séries temporelles classiques tels que ARMA, ARIMA, SARIMA, VARIMAX, ainsi que les méthodes plus avancées comme Prophet et XGBoost, offrent différentes approches pour la modélisation et la prédiction des séries temporelles. Le choix du modèle dépendra de la nature des données et des caractéristiques spécifiques du problème de prédiction. Compte tenu de la saisonnalité mensuelle dans la série temporelle de consommation électrique, SARIMA et Prophet semblent être les choix les plus adaptés. SARIMA est spécifiquement conçu pour les séries temporelles saisonnières, tandis que Prophet peut également gérer des motifs saisonniers complexes. Le choix final dépendra de la performance des modèles sur les données.

Webographie

- Site Web 1 : <https://moncoachdata.com/blog/modele-arima-avec-python/>
- Site Web 2 : <https://www.visual-design.net/post/time-series-analysis-arma-arima-sarima>
- Site Web 3 : <https://www.jedha.co/formation-ia/algorithme-arima>
- Site Web 4 : <https://labs.ivoiretalents.com/news/series-temporelles-sarima-pour-initiation-aux-predictions-temporelles-2>
- Site Web 5 : https://www.imo.universite-paris-saclay.fr/~yannig.goude/Materials/ProjetMLF/time_series.html
- Site Web 6 : <https://datascientest.com/arima-series-temporelles>
- Site Web 7 : <https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6>
- Site Web 8 : <https://www.kaggle.com/code/aminesnoussi/time-series-forecasting>
- Site Web 9 : <https://datascientest.com/xgboost-grand-gagnant-des-competitions-machine-learning-algorithme>
- Site Web 10 : <https://ledatascientist.com/facebook-prophet-la-prevision-a-grande-echelle/>

Analyse de fiabilité des sources

- Site Web 1 (moncoachdata.com) : un site web avec 61.000+ Participants, 100.000+ Leçons visionnées / an et 2.000.000+ Minutes consommées / an
- Site Web 2 (visual-design.net) : Il s'agit d'une ressource qui aborde l'analyse de séries temporelles ARMA, ARIMA, et SARIMA, cet article a eu 7,183 views.
- Site Web 3 (jedha.co) : Ce site propose des formations sur l'IA, ce qui est une source fiable d'informations.
- Site Web 4 (ivoiretalents.com) : Il s'agit d'une ressource qui parle de séries temporelles SARIMA. Son auteur est *Arthus ANIN* est un expert dans le domaine.
- Site Web 5 (universite-paris-saclay.fr) : Ce site est une ressource académique, ce qui lui confère généralement une grande fiabilité.
- Site Web 6 (datascientest.com) : est un leader des formations en Data Science, en France. Et plus de 30 groupes du CAC40 leur font confiance (TotalEnergies, Axa, Crédit Agricole, Michelin, Allianz...), et ils ont formé plus de 7 000,00 apprenants !
- Site Web 7 (towardsdatascience.com) : Il s'agit d'une plateforme connue pour des articles de qualité sur la science des données. L'auteur de cet article a 271 Followers sur la plateforme Medium.
- Site Web 8 (kaggle.com) : Kaggle est une plateforme réputée dans le domaine de l'apprentissage automatique. Les compétitions et les ressources partagées sont souvent fiables et l'article a eu 2 988 views.
- Site Web 10 (ledatascientist.com) : c'est un blog qui pour but de démocratiser la data science, l'apprentissage automatique, l'IA, et bien d'autres thématiques autour de la donnée.