

Séries Temporelles 1 (2A) - ENSAI

TP 3 : Mise en œuvre des outils du cours

Novembre 2023

Exercice 1 : Prédiction de la qualité de l'air

Le but de cet exercice est d'étudier la qualité de l'air dans les stations de RER parisiennes. En particulier, on va se concentrer sur la station Auber. Les données sont disponibles sur moodle sous le nom `qualite-de-lair-auber.csv`. Vous pouvez aussi télécharger les données actualisées sur le site <https://air-interieur.ratp.fr/details/AUBEA>.

Le tableau de données contient plusieurs variables comme la date/heure (aaaa-mm-jj-hh), le volume de monoxyde d'azote (CO), le volume de dioxyde d'azote (CO₂), dioxyde de carbone (CO₂), etc. Pour les deux premières parties, nous nous concentrerons sur les mesures de volume de dioxyde d'azote (NO₂) jusqu'au **31 décembre 2017** car entre les 01/01/2017 et 23/12/2021 il n'y a pas de registres.

Première partie : modélisation et prévision de la moyenne journalière de NO₂

- 1.1.- Construire un tableau de données contenant les mesures moyennes journalières de NO₂.
- 1.2.- Appliquer la méthode de Box-Jenkins pour proposer un modèle (S)ARIMA pour cette série. N'oubliez pas de choisir un échantillon de test pour vérifier la qualité de prédiction.
- 1.3.- Réaliser une prédiction des mesures moyennes journalières de NO₂ pour janvier 2018.

Deuxième partie : modélisation et prévision des données horaires de NO₂

- 2.1.- Appliquer la méthode de Box-Jenkins pour proposer un modèle (S)ARIMA pour les mesures horaires de NO₂. N'oubliez pas de choisir un échantillon de test pour vérifier la qualité de prédiction.

Remarque : les mesures horaires présentent une variance assez irrégulière, vous pouvez donc travailler avec l'échelle logarithmique afin d'obtenir une variance plus régulière.

- 2.2.- Réaliser une prédiction des mesures horaires de NO₂ pour la première semaine de 2018.

Bonus : Questions pratiques

B.1.- Répéter les points 1.2 et 2.1 en utilisant les données fournies après le 23 décembre 2021. Les modèles ont-ils changé ? Comparez ces deux périodes et posez-vous des questions sur l'évolution de NO₂ dans cette station. Est-ce que le projet de *désenfumage et de renouvellement de l'air des espaces souterrains parisiens* (2021-2024) a donné des résultats significativement positifs à la station Auber ?

B.2.- On a constaté la présence de valeurs manquantes dans le jeu de données : (i) données manquantes isolées, (ii) données manquantes sur courte période et (iii) données manquantes sur longues périodes. Pour les cas (i) et (ii) on a réussi à estimer plusieurs paramètres à l'aide des options `omit.na` et `'na.action' = na.pass`.

B.2.1.- Expliquer brièvement comment les estimateurs empiriques de l'ACF, la moyenne, etc fonctionnent sous la présence de données manquantes.

B.2.2.- Imaginons que la RATP a caché quelques données et qu'un organisme d'arbitrage vous demande d'imputer (i.e. compléter) les données manquantes. Proposer une démarche pour imputer les données manquantes. Pour cela, vous devez :

- Choisir un échantillon d'entraînement et de test
- Utiliser tous les modèles que vous considérez pertinent pour imputer ces données. Vous pouvez aussi utiliser les modèles du cours d'Apprentissage Supervisé (Quels seraient les modèles adaptés dans ce cadre ?).
- Comparer les modèles dans ce cadre en commentant leurs faiblesses et leur robustesse.

Exercice 2 : Prix du Bitcoin

On sait que le prix du Bitcoin a des comportements assez complexes. Cette complexité est encore plus marquée lorsque les registres sont donnés à des intervalles de temps plus courts, en particulier à cause de la volatilité du prix. Dans cet exercice, on va essayer de modéliser le prix de clôture (en \$) en utilisant les modèles vus en cours et en allant plus loin avec des modèles que vous pourriez trouver dans la littérature et/ou en utilisant ChatGPT.

Le jeu de données se trouve sur la page moodle du cours sous le nom `BTC-USD.csv`. Vous pouvez charger les données via la commande suivante :

```
library(readr)
bitcoin <- read_csv("BTC-USD.csv")
```

Première partie : Analyse des prix moyens mensuels de clôture.

- 1.1.- Construire un tableau de données contenant les prix moyens mensuels de clôture. Dans ce tableau, inclure aussi une colonne contenant le logarithme des prix moyens mensuels de clôture.
- 1.2.- Appliquer la méthode de Box-Jenkins pour proposer un modèle (S)ARIMA pour la série *log des prix moyens mensuels de clôture*. Pourquoi on ne travaille pas directement avec *les prix moyens mensuels de clôture* ?
- 1.3.- Donner une prédiction du prix moyen du Bitcoin pour décembre 2023.

Seconde partie : Analyse des prix de clôture.

Dans cette partie vous avez la liberté d'utiliser les modèles (vus en cours, supports supplémentaires du cours, littérature, ChatGPT, etc) que vous voulez afin de modéliser le prix de clôture du Bitcoin. Ne vous limitez pas. Une fois validé votre modèle, réalisez une prédiction pour la semaine du 27 novembre et 1er décembre 2023. Au moment de ce TP, vous aurez déjà les prix et on pourra les comparer avec votre prédiction.