

Analyse de séries temporelles

Rasmata Sawadogo - Courtney Saint-Hubert - Sidy B. Diop - Romain Lesueur

BUT Science des données

IUT Grand Ouest Normandie - Site Lisieux

Introduction

Les séries temporelles, constituées de données quantitatives indexées par le temps, jouent un rôle crucial dans divers domaines, tels que la santé, l'environnement, les transports, et surtout la finance. Dans le cadre de ce projet, l'accent est mis sur l'analyse des cotations de matières premières, notamment le café, le cacao, le jus d'orange, le sucre, et le pétrole. Ces données couvrent une période de plus d'une décennie, permettant d'observer les tendances et les variations à long terme.

Les données, contenues dans des fichiers PDF, ont été collectées depuis le site internet



Bibliothèque data.table

Sous R, on utilise la bibliothèque *tabulapdf* pour extraire et traiter les données tabulaires contenues dans des fichiers PDF. Cet outil puissant est spécialement conçu pour gérer des fichiers complexes, permettant une récupération rapide et précise des tables.



Cette bibliothèque est essentielle pour les projets nécessitant une récupération systématique et fiable de données à partir de sources PDF volumineuses.

Visualisation de la structure

La première étape va consister à visualiser la structure des données présentes dans les fichiers PDF provenant d'investing.com afin de mieux comprendre leur structure, de détecter les anomalies et d'adapter efficacement les techniques de préparation aux caractéristiques des données.

Date	Dernier	Ouv.	Plus Haut	Plus Bas
10/10/2024	78,360	76,730	78,640	76,730
09/10/2024	76,780	77,520	78,000	75,190
08/10/2024	77,470	81,010	81,160	76,390
07/10/2024	81,130	77,490	81,170	77,390
06/10/2024	77,520	78,060	78,070	77,270
04/10/2024	78,080	77,800	79,290	77,390
03/10/2024	77,900	74,580	77,990	74,330
02/10/2024	74,670	74,450	76,140	73,600
01/10/2024	74,460	71,860	75,450	69,920

Fig. 1 - Extrait de jeu de données

Préparation des données

Avec quelques lignes d'instructions, on peut récupérer les données présentes dans les fichiers PDF, autrement dit les données de cotation des prix du café, du cacao, du jus d'orange, du sucre et du baril de pétrole.

```
sucre_data <- extract_tables("Futures sucre Londres - Données Historiques.pdf",  
  method = "decide",  
  encoding = "UTF-8",  
  col_names = FALSE,  
  output = "tibble")
```

```
petrol_data <- extract_text("Futures pétrole Brent - Données Historiques.pdf")
```

Fig. 2 - Instruction pour charger les données du baril de pétrole

```
# A tibble: 3,818 × 6  
  Date      Closed_Cotation Opened_Cotation Highest_Cotation Lowest_Cotation  
  <date>      <dbl>          <dbl>          <dbl>          <dbl>  
1 2024-10-10 78.4            76.7            78.6            76.7  
2 2024-10-09 76.8            77.5            78              75.2  
3 2024-10-08 77.5            81.0            81.2            76.4  
4 2024-10-07 81.1            77.5            81.2            77.4  
5 2024-10-06 77.5            78.1            78.1            77.3  
6 2024-10-04 78.1            77.8            79.3            77.4  
7 2024-10-03 77.9            74.6            78.0            74.3  
8 2024-10-02 74.7            74.4            76.1            73.6  
9 2024-10-01 74.5            71.9            75.4            69.9  
10 2024-09-30 71.7            71.9            72.8            71.0  
# i 3,808 more rows  
# i 1 more variable: TTT <chr>  
# i Use `print(n = ...)` to see more rows
```

Fig. 3 - Extrait de tableau des données du baril de pétrole

Analyse Exploratoire

Une fois les données préparées, il est essentiel de les analyser en profondeur à l'aide des outils classiques de statistique exploratoire, tels que des tableaux récapitulatifs et des visualisations graphiques, afin de mettre en évidence les tendances, les relations et les caractéristiques clés du jeu de données.

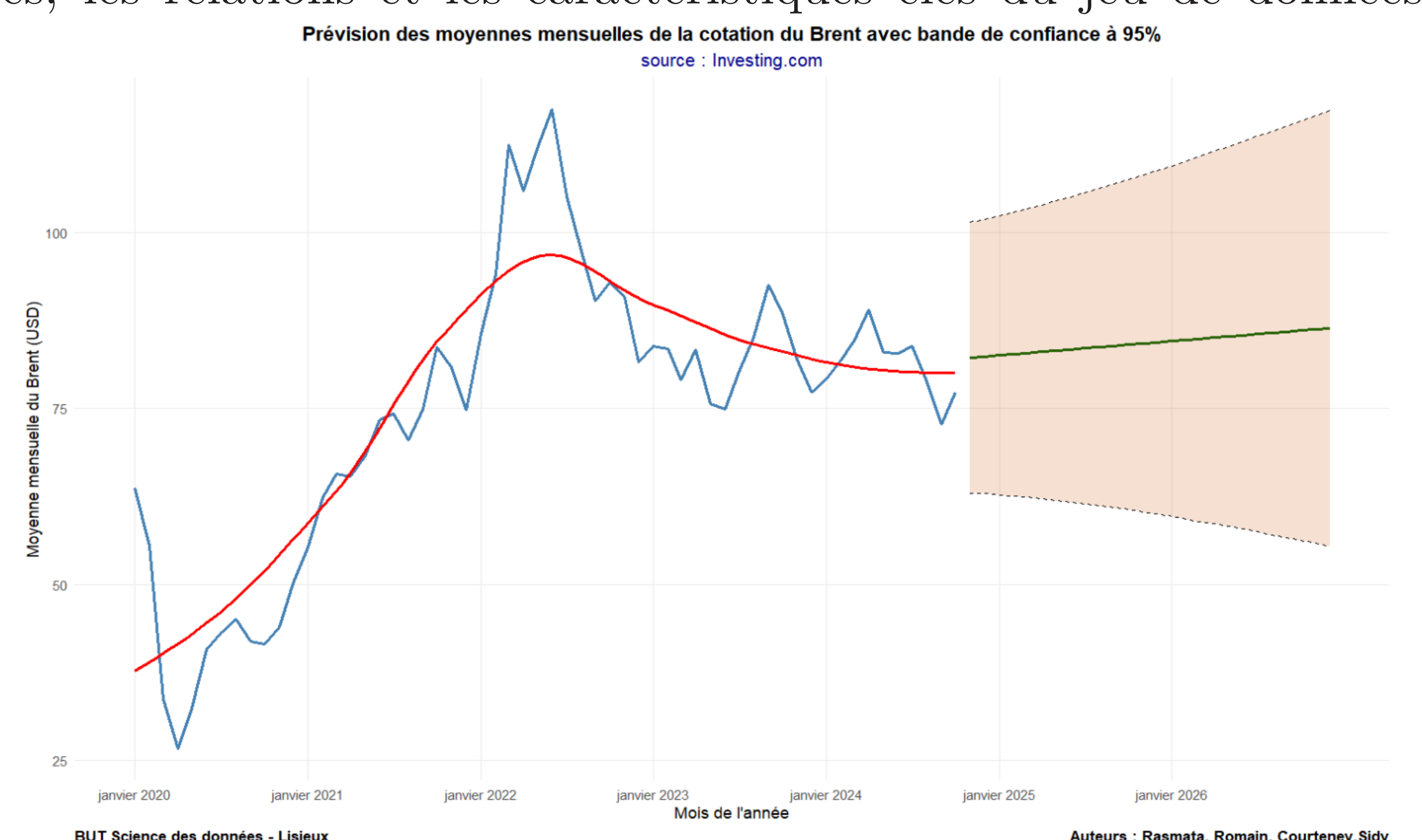
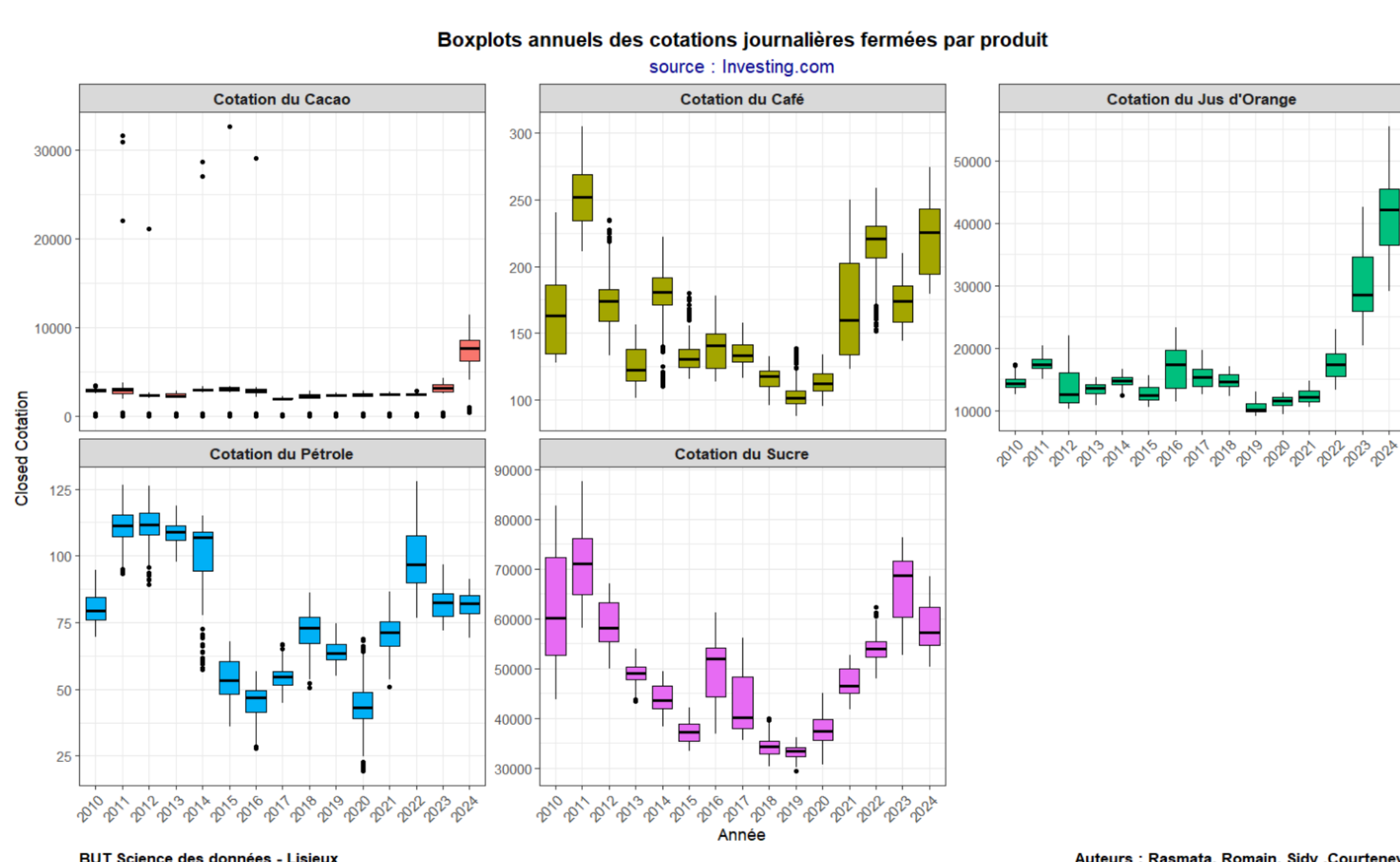


Fig. 4 - Boxplots annuels des cotations journalières fermées par produit

Fig. 5 - Prévision des moyennes mensuelles de la cotation du Brent