1

2022-2023

**Khaoula Aroui**

Etudiante en Master statistique pour L’évaluation et la prévision

Une image contenant texte, clipart, carte de visite

Description générée automatiquement



**Modèles SARIMA et Prophet pour la prévision de la température en Irlande**

**Une analyse comparative basée sur des données historiques**

Table des matières

[Données 2](#_Toc128731350)

[Corrigé en variation saisonnière 2](#_Toc128731351)

[Prévision 2](#_Toc128731352)

[1. Approche classique : SARIMA 2](#_Toc128731353)

[2. Prophet 2](#_Toc128731354)

[Conclusion 2](#_Toc128731355)

[Références bibliographiques 3](#_Toc128731356)

**Table des figures**

**Liste des tableaux**

Aucune entrée de table d'illustration n'a été trouvée.

# Introduction

L'analyse de séries chronologiques est une méthode populaire pour étudier les données temporelles. Elle est couramment utilisée pour comprendre les tendances et les modèles saisonniers dans les données, ainsi que pour faire des prévisions. Parmi les différentes méthodes d'analyse de séries chronologiques, SARIMA et Prophet sont deux des plus courantes.

SARIMA (Seasonal Autoregressive Integrated Moving Average) est une méthode statistique qui utilise des modèles autorégressifs intégrés et moyennes mobiles pour modéliser les tendances et les modèles saisonniers dans les données. Il est souvent utilisé pour prévoir des données saisonnières telles que les ventes de détail, les fluctuations du marché boursier et les données météorologiques.

Prophet, quant à lui, est une méthode plus récente développée par Facebook pour modéliser les tendances et les modèles saisonniers dans les données temporelles. Prophet utilise un modèle additif plutôt que multiplicatif, qui permet une plus grande flexibilité dans la modélisation des tendances saisonnières et non saisonnières.

Dans ce travail, nous allons examiner la variation saisonnière de la température en Irlande, en construisant un modèle à l'aide de l'approche statistique SARIMA. Nous allons ensuite comparer les résultats de prévision obtenus avec SARIMA à ceux obtenus avec Prophet.

Nous allons examiner les avantages et les inconvénients de chaque méthode pour comprendre les différences dans les résultats obtenus.

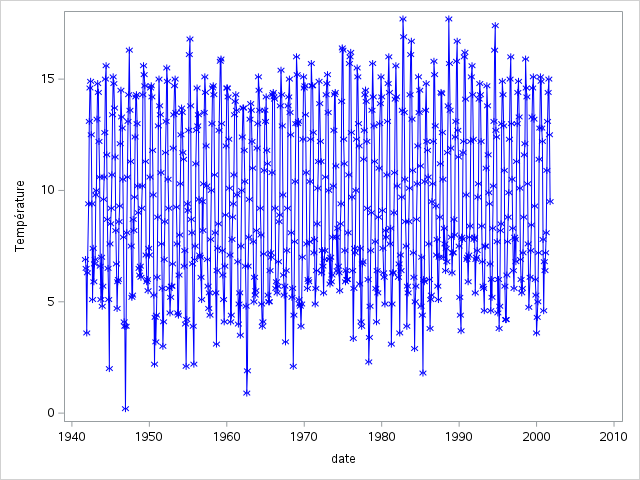
Nous allons également explorer la manière dont les prévisions sont évaluées pour chaque méthode

# Données

Un ensemble de données météorologiques mensuelles de Novembre 1941 à Janvier 2018 pour l'aéroport de Dublin, en Irlande, provenant du diffuseur météorologique irlandais est utilisé.

L'ensemble des données se compose de 915 observations pour 11 variables, dont la première est la date en jours et les autres sont des mesures de température. Dans notre analyse, nous nous intéressons particulièrement à la variable de température moyenne.

# Visualisation des données



# Corrigée en variation saisonnière

La saisonnalité est une caractéristique particulièrement endémique des données météorologiques - ce qui explique pourquoi de nombreuses régions du monde ont quatre saisons.

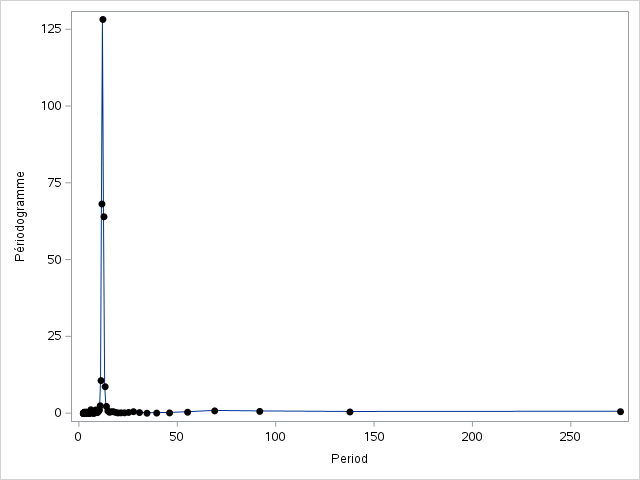
Lorsque la saisonnalité n'est pas prise en compte, on risque de faire des prévisions erronées des données. Alors que l'on peut prévoir une valeur moyenne pour une série temporelle particulière, les pics et les creux autour de cette moyenne affectent considérablement les prévisions pour cette série temporelle.

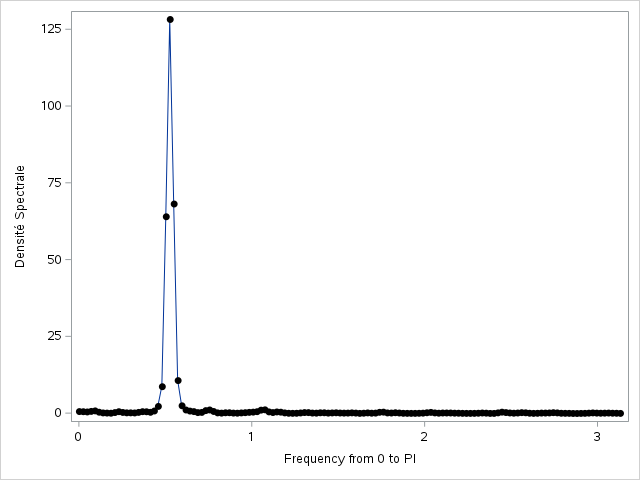
La saisonnalité est une préoccupation importante lorsqu'il s'agit de modéliser des séries temporelles.

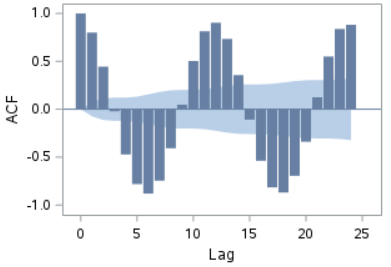
En premier lieu, on voulait investiguer la saisonnalité de nos données et puis après corriger cette composante saisonnière.

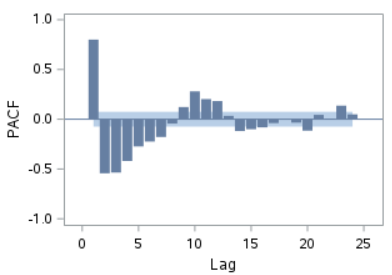
Comme première étape on voulait confirmer l’existence d’une saisonnalité et la définir.

Une première méthode à adopter est de faire sortir le périodogramme de la densité spectrale.









# Corrigé en variation saisonnière

Dans l'estimation de modèles ARIMA, il est courant de commencer par considérer des modèles de base tels que les modèles ARIMA (p,0,0), (0,d,q), ou (p,d,q) avec p = 1 et/ou q = 1.

Le modèle (p,0,0) est un modèle autorégressif d'ordre p, qui ne contient pas de termes de moyenne mobile. Le modèle (0,d,q) est un modèle de moyenne mobile d'ordre q, qui ne contient pas de termes autorégressifs. Le modèle (p,d,q) est un modèle ARIMA qui contient des termes autorégressifs et de moyenne mobile.

Ces modèles de base sont souvent utilisés comme point de départ dans la recherche d'un modèle ARIMA optimal, car ils sont simples et faciles à interpréter. En ajustant ces modèles de base, on peut évaluer l'adéquation de ces modèles aux données et identifier les retards significatifs dans l'ACF et la PACF.

Ensuite, en fonction des résultats de l'identification initiale, on peut explorer d'autres modèles ARIMA plus complexes (par exemple, en augmentant l'ordre de p et/ou q) ou des modèles ARIMA saisonniers pour améliorer l'ajustement et la précision des prévisions. Cela permet de déterminer le meilleur modèle ARIMA pour les données analysées.

# Prévision

## Approche classique : SARIMA

## Prophet

Prophet FB was developed by Facebook as an algorithm for the in-house prediction of time series values for different business applications. Therefore, it is specifically designed for the prediction of business time series.

It is an additive model consisting of four components:



Let us discuss the meaning of each component:

1. **g(t):** It represents the *trend* and the objective is to capture the general trend of the series. For example, the number of advertisements views on Facebook is likely to increase over time as more people join the network. But what would be the exact function of increase?
2. **s(t):**It is the*Seasonality*component. The number of advertisement views might also depend on the season. For example, in the Northern hemisphere during the summer months, people are likely to spend more time outdoors and less time in from of their computers. Such seasonal fluctuations can be very different for different business time series. The second component is thus a function that models seasonal trends.
3. **h(t):** The *Holidays* component. We use the information for holidays which have a clear impact on most business time series. Note that holidays vary between years, countries, etc. and therefore the information needs to be explicitly provided to the model.
4. The **error term** εt stands for random fluctuations that cannot be explained by the model. As usual, it is assumed that εt follows a normal distribution *N*(0, σ2) with zero mean and unknown variance σ that has to be derived from the data.

# Conclusion

La pandémie de COVID-19 a été un événement mondial qui a mis à l'épreuve l'état de préparation du gouvernement face à un virus hautement infectieux. Nous avons étudié, dans le cadre de ce projet, les principaux facteurs ayant contribué à la propagation rapide au niveau régional et départemental, ainsi que l'efficacité des procédures gouvernementales pour faire face à la menace du COVID-19.

Une étude sur l’année 2021 sur les indicateurs de propagation a montré un comportement similaire entre les régions d'un point de vue de l’évolution dans le temps, avec des pics de taux d'incidence pour la plupart de la région au printemps, un second pic à la fin de l'été, et un pic énorme l'hiver suivant avec l'apparition des nouvelles variantes du virus. D'un autre côté, nous avons remarqué des différences entre les régions, et donc entre les différents départements, tel que certaines régions

# Références bibliographiques

1. "Info Coronavirus COVID-19 - Les actions du Gouvernement." Gouvernement.fr. [En ligne]. Disponible à l'adresse : <https://www.gouvernement.fr/info-coronavirus>
2. "Population des départements français." Insee. [En ligne]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/>
3. "Population par département − France, portrait social." Insee. [En ligne]. Disponible à l'adresse : <https://www.insee.fr/fr/statistiques/>
4. "Département français : liste, carte, région, préfecture." regions-departements-france.fr. [En ligne]. Disponible à l'adresse : <https://regions-departements-france.fr/departement/>
5. "Indicateurs de suivi de l’épidémie de COVID-19." data.gouv.fr. [En ligne]. Disponible à l'adresse : <https://www.data.gouv.fr/fr/datasets/indicateurs-de-suivi-de-lepidemie-de-covid-19/>
6. "Données relatives aux personnes vaccinées contre la COVID-19 (VAC-SI)." data.gouv.fr. [En ligne]. Disponible à l'adresse : <https://www.data.gouv.fr/fr/datasets/donnees-relatives-aux-personnes-vaccinees-contre-la-covid-19-vac-si/>
7. "A Complete Guide to Violin Plots." Chartio. [En ligne]. Disponible à l'adresse : <https://chartio.com/learn/charts/violin-plots/>

**Enoncé**

je vais donc vous demander de me rendre  
un petit série à analyser après les vacances, comme on a fait en cours,  
avec saisonnalité, de me donner la corrigée en variation saisonnière.  
J'aimerais bien également que vous me compariez des prévision avec un  
algo de machine Learning comme Prophet (si j'ai bien compris :-)) pour  
voir ainsi ce qui donne les meilleurs résultats. Avec votre série, vous  
coupez 10% des données, vous faîtes de la prévision sur les 90% puis  
comparez les résultats avec les données réelles, modèle SARIMA versus  
Prophet.