



PROJET MASTER 1 - SSD  
HAX817X

---

## Changement climatique

Comparaisons statistiques entre les modèles étudiés par le GIEC et  
des données historiques

---

Mario LAPI  
Jeanne VIVIER

Encadré par : David MÉTIVIER

2024

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Données</b>	<b>5</b>
2.1	Données observées . . . . .	5
2.2	Données simulées . . . . .	5
2.3	Stations . . . . .	6
<b>3</b>	<b>Décomposition de séries temporelles</b>	<b>7</b>
3.1	La tendance . . . . .	9
3.1.1	Avec le package Forecast . . . . .	9
3.1.2	Moyenne Mobile . . . . .	9
3.1.3	Moyenne Mobile classique . . . . .	10
3.1.4	Moyenne mobile avec poids gaussiens . . . . .	11
3.1.5	Autres estimations de la tendance . . . . .	12
3.2	La saisonnalité . . . . .	12
3.3	Résidus . . . . .	13
<b>4</b>	<b>Processus Autorégressifs</b>	<b>13</b>
4.1	Autocorrélations . . . . .	13
4.1.1	Autocorrélations simples . . . . .	14
4.1.2	Autocorrélations partielles . . . . .	14
4.2	Sélection de modèle . . . . .	16
4.3	Estimation des coefficients . . . . .	16
4.3.1	Par la méthode du maximum de vraisemblance . . . . .	16
4.3.2	Par la méthode des moindres carrés . . . . .	17
<b>5</b>	<b>Corrélations</b>	<b>18</b>
5.1	Moyennes mensuelles . . . . .	18
5.2	Corrélations spatiales . . . . .	20
<b>6</b>	<b>Précipitations</b>	<b>21</b>
<b>7</b>	<b>Extrêmes</b>	<b>24</b>
<b>8</b>	<b>Conclusion</b>	<b>25</b>
<b>9</b>	<b>Annexes</b>	<b>26</b>
9.1	Bibliographie . . . . .	26
9.2	Code . . . . .	27
9.2.1	Tendance par moyenne mobile . . . . .	27
9.2.2	Tendance par moyenne mobile avec poids gaussiens . . . . .	27
9.2.3	Estimation des coefficients $\phi_1$ , $\phi_2$ et $\phi_3$ par la méthode du maximum de vraisemblance . . . . .	27
9.2.4	Estimation des coefficients $\phi_1$ , $\phi_2$ et $\phi_3$ par les moindres carrés ordinaires . .	28
9.2.5	Vagues de chaleur . . . . .	28

## Résumé

Le changement climatique est l'un des grands défis posés à l'humanité pour le XXI<sup>e</sup> siècle. Les disciplines de l'informatique et des mathématiques jouent un rôle clé dans la résolution de ce problème, notamment grâce à la modélisation. L'objectif de ce rapport est de vérifier que les données de température maximale journalière et pluviométrie issues des simulations conservent les mêmes propriétés statistiques que les données observées. Nous utilisons les modèles ALADIN63 et RACMO22E. Nous analyse comparative portera sur des séries temporelles. Nous commencerons par vérifier la représentation de la tendance et la saisonnalité par les données simulées. Une fois ces paramètres retirés, nous travaillerons avec des processus autorégressifs, que nous définirons et dont nous estimerons les coefficients. Puis, nous étudierons la reproduction des corrélations spatiales et des phénomènes extrêmes. L'une des conclusions de ce rapport est que les propriétés statistiques du jeu de données simulées sont similaires aux propriétés du jeu de données observées pour les températures maximales journalières. Toutefois, cette correspondance est moins évidente lorsqu'il s'agit de la pluviométrie ou des événements extrêmes.

# 1 Introduction

Le changement climatique, phénomène mondial complexe, se réfère aux modifications à long terme des modèles météorologiques de la Terre, résultant principalement de l'augmentation des concentrations de gaz à effet de serre dans l'atmosphère. Ces changements ont des répercussions profondes sur le climat, provoquant des événements météorologiques extrêmes, des perturbations écologiques et des impacts sociaux-économiques à l'échelle mondiale.

Dès la fin du XIX<sup>e</sup> siècle, des scientifiques, comme Svante Arrhenius, ont identifié l'influence du dioxyde de carbone atmosphérique sur le système climatique de la Terre [1].

C'est dans les années 1950 que les avancées technologiques permettent de développer des ordinateurs, ouvrant la voie à la modélisation moderne et conduisent aux premières prévisions météorologiques numériques [2]<sup>1</sup>. Ce nouvel outil vient appuyer les discussions qui fleurissent à partir des années 1970 sur les questions des ressources.

À partir de 1988, la science du climat est institutionnalisée au niveau international, avec la création du Groupe d'Experts Intergouvernemental sur l'évolution du Climat (GIEC ou IPCC en anglais), qui a pour objectif d'évaluer l'état des connaissances sur l'évolution du climat, ses causes, ses impacts, et d'identifier les possibilités de limiter l'ampleur du réchauffement et la gravité de ses impacts et de s'adapter aux changements attendus (d'après le Ministère de la transition écologique et de la cohésion des territoires).

Afin d'essayer d'anticiper les événements futurs, ces experts ont proposé des modèles climatiques, permettant de prédire les températures et précipitations des décennies à venir. Le groupe G1 du GIEC définit les modèles climatiques comme "des programmes informatiques extrêmement sophistiqués qui englobent notre compréhension du système climatique et simulent, avec autant de fidélité que possible, les interactions complexes entre l'atmosphère, l'océan, la surface terrestre, la neige et la glace, l'écosystème mondial et divers processus chimiques et biologiques".

La modélisation permet ainsi de comprendre le caractère historique et inédit du bouleversement climatique actuel, comme l'illustre le papier de Westerhold & al publié dans *Science* en 2020 [3], qui fournit une modélisation du climat des 66 derniers millions d'années. Cette modélisation historique permet de comprendre la rapidité du phénomène de réchauffement actuel, et de fait, les risques que cela induit pour les civilisations modernes. La figure 1 *Past and future trends in global mean temperature and atmospheric CO<sub>2</sub>* issue de l'article susmentionné, avec en abscisse les échelles de temps, témoigne de la vitesse de l'augmentation des températures globales, et de l'urgence climatique à laquelle nous faisons face.

---

1. La publication du rapport du Club de Rome sur les Limites à la croissance est un exemple historique du rôle de la modélisation pour éclairer les défis posés aux sociétés humaines.

Le modèle World 3 n'utilise de variable climat, mais une variable appelée "pollution" qui s'en rapproche

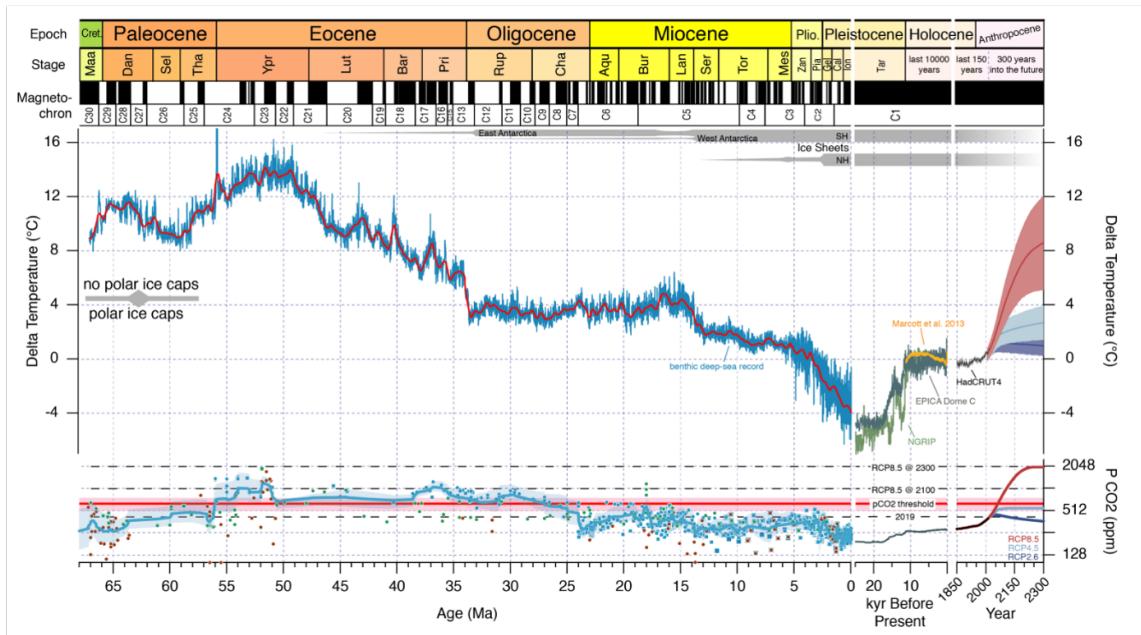


FIGURE 1 – Past and future trends in global mean temperature and atmospheric CO<sub>2</sub>

La figure 2 tirée du rapport AR6 du GIEC, représente la concentration en CO<sub>2</sub>-équivalent dans l'atmosphère selon différents scénarios. Le CO<sub>2</sub> équivalent est une unité de mesure créée par le GIEC, qui sert à quantifier l'impact des différents gaz à effet de serre sur l'environnement, en prenant comme référence le dioxyde de carbone, qui est le principal gaz à effet de serre. Les experts ont défini quatre profils représentatifs d'évolution des concentrations de gaz à effet de serre pour le XXI<sup>e</sup> siècle et au-delà. Ces trajectoires englobent un large éventail de possibilités, correspondant à des efforts plus (en vert) ou moins (en rouge) grands de réduction des émissions de gaz à effet de serre au niveau mondial.

Ces modèles sont issus de simulations de données, passées et futures, et permettent de comprendre l'impact des émissions anthropiques sur la rapidité du changement climatique. Les modèles du système climatique composent, non pas des prévisions, irréalisables à l'échelle de plusieurs décennies, mais des scénarios d'évolution du climat, supposés couvrir un large éventail d'évolutions possibles. Mais comment juger de la fiabilité de ces modèles ?



FIGURE 2 – Représentation de tous les agents équivalent-CO<sub>2</sub> du forçage atmosphérique, selon quatre scénarios RCP (en ppm)

## 2 Données

### 2.1 Données observées

Dès le XX<sup>e</sup> siècle, la mise en place de stations météorologiques dans les grandes villes françaises, ainsi que les avancées technologiques permettent de recueillir des observations de températures, d'humidité, de précipitations, de couverture nuageuse... Ces observations sont enregistrées et regroupées sous forme de jeux de données. Certains de ces jeux de données sont disponibles en libre accès, comme le propose *European Climate Assessment & Dataset project* [4]. Ce site web présente des informations sur les changements dans les extrêmes météorologiques et climatiques, ainsi que les données quotidiennes nécessaires pour surveiller et analyser ces extrêmes.

Les données observées, représentées en bleu dans la suite de ce rapport, sont issues de ce site web. Nous utiliserons les données journalières de températures maximales (en °C) et de précipitations (en mm/m<sup>2</sup>).

### 2.2 Données simulées

Les modèles climatiques sont produits à partir de données simulées. Ces données sont disponibles sur le site web du Ministère de la transition écologique, dans la section data DRIAS : *DRIAS les futurs du climat*[5], qui a pour vocation de mettre à disposition des projections climatiques régionalisées réalisées dans les laboratoires de recherche sur la modélisation du climat.

**Définition 1.** *La période de référence est la période par rapport à laquelle les évolutions sont calculées. Pour analyser les effets du changement climatique, on parle rarement des valeurs absolues attendues sur une région ou un pays mais plutôt de l'écart avec les valeurs observées sur une période passée. Cette période est longue d'au moins 30 ans. La période de référence du climat passé n'est pas identique pour toutes les modélisations, selon les applications et leur finalité, plusieurs références peuvent cohabiter. Le choix de la période de référence peut modifier la valeur des changements attendus pour différentes variables climatiques, la température notamment.*

Pour ce présent rapport, la période de référence est de 50 ans, du 01 janvier 1955 au 31 décembre 2005.

Nous comparerons deux modèles :

- Le modèle ALADIN63\_CNRM-CM5, sous le scénario d'émission de référence, produit par l'institut Météo France / Centre National de Recherches Météorologiques. Il sera cité dans la suite de ce rapport sous le nom d'ALADIN63, et représenté en orange sur les figures.
- Le modèle RACMO22E\_CNRM-CM5, sous le scénario d'émission de référence, produit par l'institut météorologique royal des Pays-Bas, Royal Netherlands Meteorological Institute, De Bilt, The Netherlands. Il sera cité dans la suite de ce rapport sous de nom de RACMO22E,et représenté en vert sur les figures.

Afin de comparer les simulations aux observations, nous travaillerons également avec les données journalières de températures maximales (en °C) et de précipitations (en mm/m<sup>2</sup>), cette fois issues du site du DRIAS.

*La simulation doit être vue comme une des réalisations possibles du climat. L'enjeu de ce rapport n'est pas de vérifier que les modèles simulés reproduisent les conditions météorologiques exactes au quotidien, mais plutôt qu'ils gardent les mêmes propriétés statistiques que le modèle observé.*

### 2.3 Stations

Afin de quantifier la qualité des modèles, nous proposons une comparaison statistique des trois jeux de données décrits dans la section précédente. Pour cela, nous sélectionnons 15 stations (Figure 3) sur des critères géographiques (sud, nord, littoral...) et de qualité (jeux de données complets).

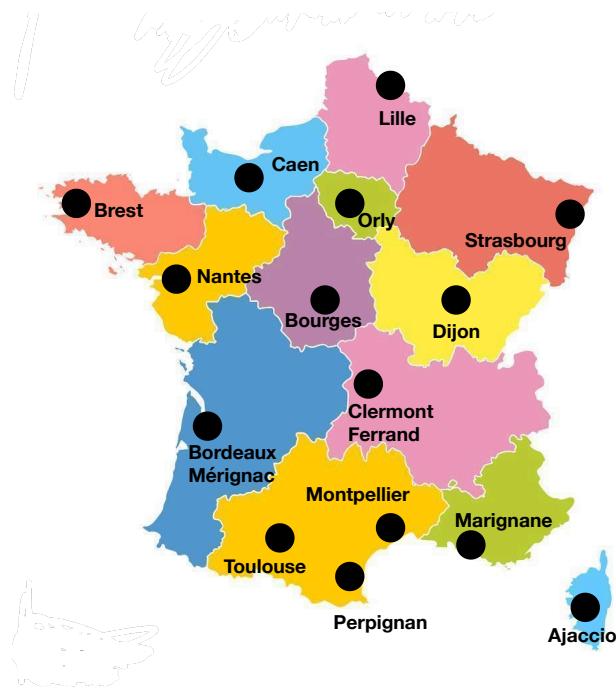


FIGURE 3 – Carte de France avec les stations sélectionnées

Nous obtenons alors un découpage relativement uniforme de la carte de France, avec des stations réparties de manière homogène sur le territoire, permettant d'analyser les corrélations spatiales.

Nous comparerons ainsi les données données simulées aux données historiques, sous différents critères statistiques. D'abord en vérifiant la bonne représentation de la tendance et de la saisonnalité par les données simulées. Une fois ces paramètres retirés, nous travaillerons avec des processus autorégressifs, que nous définirons et dont nous estimerons les coefficients. Puis, nous étudierons la bonne reproduction des corrélations spatio-temporelles, et des extrêmes.

L'intégralité des codes informatiques utiles à ce projet ont été réalisé dans le langage de programmation **Julia**, et sont disponibles sur Github, à l'adresse suivante : [Github](#).

### 3 Décomposition de séries temporelles

Dans cette section, nous traiterons des données quotidiennes de températures maximales. Dans un premier temps, nous ne tiendrons pas compte de la localisation des stations choisies, en faisant une moyenne de toutes les stations sélectionnées.

Nous travaillons avec trois jeux de données quotidiennes, présentés sous forme de séries temporelles, que nous définissons de la manière suivante :

**Définition 2.** *Une série temporelle (ou série chronologique) est une suite finie ( $X_1, \dots, X_n$ ) de données indexées par le temps, où  $n$  représente le nombre d'observations donc la longueur de la série [6].*

Nous représentons graphiquement les trois séries temporelles, relatives aux différents jeux de données.

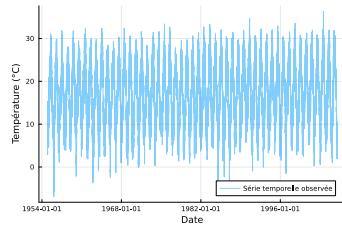


FIGURE 4 – Série temporelle données observée

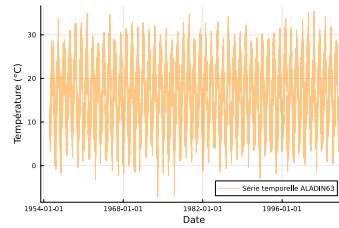


FIGURE 5 – Série temporelle données ALADIN63

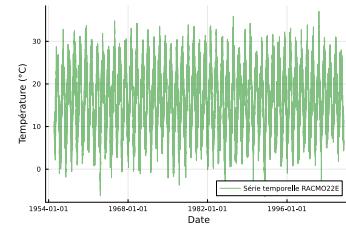


FIGURE 6 – Série temporelle données RACMO22E

À première vue, les trois séries temporelles se ressemblent. Afin de mieux les comparer, nous les affichons toutes sur un même graphique(Figure 7).

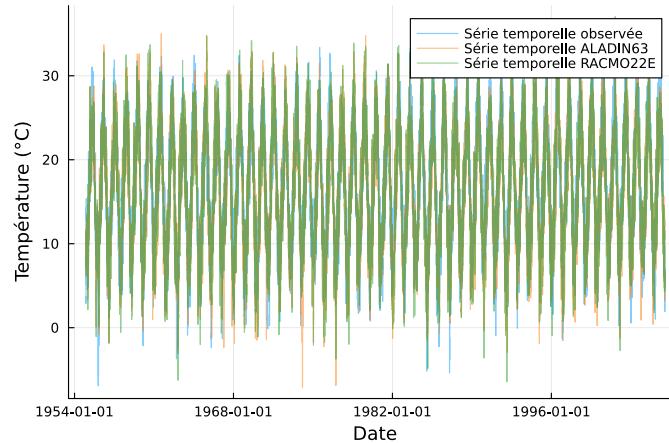


FIGURE 7 – Séries temporelles issues des trois jeux de données

La figure 7 représente les trois séries temporelles étudiées. Elles sont, la plupart du temps, confondues et même si elles diffèrent assez souvent ce n'est pas grave car on s'intéresse plus aux

propriétés statistiques qu'aux réalisations de nos séries. Ainsi, ce qui compte ici, c'est que les séries semblent avoir le même "aspect". En effet, nous pouvons observer un motif presque périodique pour chaque série, il s'agit de la saisonnalité, dont nous discuterons plus tard dans ce rapport.

Définissons quelques notions utiles à la suite de ce rapport :

**Définition 3.** *Un processus stochastique est une famille de variables aléatoires  $(X_t)_{t \in T}$  définies sur un espace de probabilisé  $(\Omega, \mathcal{A}, T)$ .*

*Lorsque  $T = \mathbb{Z}$ , le processus  $(X_t)_{t \in T}$  est dit à temps discret.*

*En particulier,*

- $\forall t \in T, \omega \mapsto X_t(\omega)$  est une variable aléatoire
- $\forall \omega \in \Omega, t \mapsto X_t(\omega)$  est une fonction appelée la trajectoire du processus  $(X_T)$

**Définition 4.** *Un processus  $(X_t)_{t \in \mathbb{Z}}$  est dit stationnaire si son espérance et ses autocovariances sont invariables par le temps.*

Étant donné que les séries temporelles que l'on étudie comportent une saisonnalité, cela suffit pour dire qu'elles ne sont pas des réalisations de processus stationnaires. Il est cependant possible de s'y ramener par décomposition de la série temporelle en plusieurs composantes.

- **La tendance :**  $m_t$  modélise l'évolution à long terme de la série.
- **La composante saisonnière :**  $s_t$  représente un comportement qui se répète de façon périodique.
- **La composante aléatoire :**  $e_t$  est aléatoire, c'est-à-dire qu'il s'agit d'un processus stochastique. On la désigne souvent le résidu ou le bruit.

Le modèle de décomposition choisit dans la suite est un modèle additif. Le processus  $X_t$  s'écrit alors de la manière suivante :

$$X_t = m_t + s_t + e_t \quad \text{où} \quad \mathbb{E}[e_t] = 0$$

### 3.1 La tendance

#### 3.1.1 Avec le package Forecast

Afin d'extraire la tendance de nos séries temporelles, nous utilisons la fonction `stl` du package Forecast de Julia [7].

La fonction `stl` (pour Seasonal and Trend decomposition using Loess) est utilisée pour décomposer une série temporelle en ses composantes saisonnières, de tendance et résiduelles. Le calcul de la tendance dans `stl` se fait généralement en utilisant une méthode de lissage non paramétrique appelée Loess (Locally Weighted Scatterplot Smoothing). Cette méthode consiste à ajuster localement des fonctions polynomiales aux données en utilisant une fenêtre de lissage mobile [8].

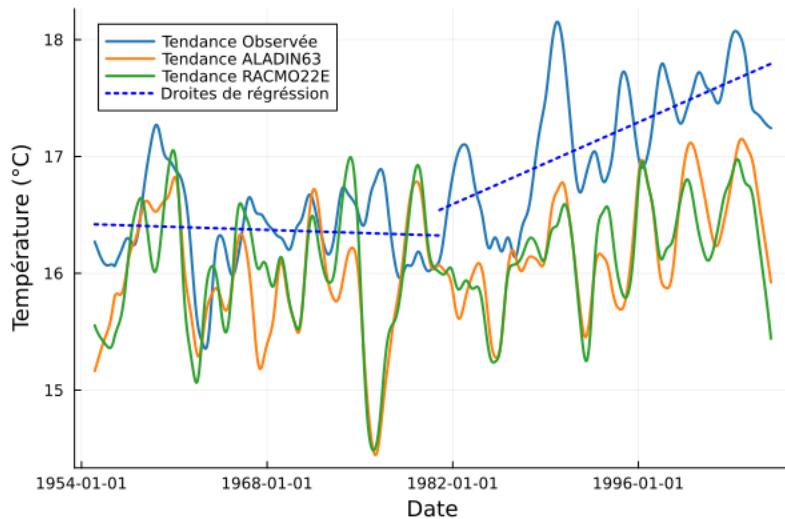


FIGURE 8 – Tendances obtenues à l'aide de la fonction `stl` du package Forecast

La méthode Loess permet d'obtenir des courbes assez lisses. La tendance semble plutôt croissante. Nous remarquons qu'autour de 1981, les modèles du DRIAS, qui ont des tendances assez proches, estimaient une température maximale moyenne en France clairement inférieure aux données observées. À partir de là, les modèles simulés semblent sous-estimer, avec en plus un léger décalage, les températures observées.

Nous avons tracé deux droites de régression linéaire, l'une pour la période 1955 - 1981, et l'autre pour la période 1981 - 2005. Alors qu'une légère baisse de température semblait avoir lieu sur la première période, la seconde droite de régression est clairement croissante. Nous observons une augmentation d'environ 1°C au cours de cette seconde période.

#### 3.1.2 Moyenne Mobile

Il existe d'autres méthodes pour estimer la tendance d'une série temporelle. L'une d'elles est la méthode par moyenne mobile.

La méthode des moyennes mobiles permet d'extraire la tendance d'une série temporelle. C'est une méthode de lissage car elle agit comme un filtre qui élimine le "bruit". La méthode des moyennes mobiles consiste à calculer une moyenne glissante des valeurs de la série que l'on souhaite "lisser". Le calcul de la moyenne mobile dépend d'un paramètre  $q$  qui correspond au nombre d'observations incluses dans le calcul de la moyenne glissante effectuée. Plus  $q$  est grand plus le lissage est important (jusqu'à atteindre la fonction constante égale à la moyenne). La série des moyennes mobiles  $\hat{m}_t$  d'ordre  $2q + 1$ , pour  $q \in \mathbb{N}^*$  se calcule de la manière suivante :

$$\hat{m}_t = \frac{1}{2q+1} \sum_{l=-q}^q X_{t+l}$$

On remarquera quand même qu'on "perd"  $2q$  observations lorsqu'on fait un lissage moyenne mobile d'ordre  $2q + 1$ .

Il existe différentes méthodes de pondérations des valeurs pour la méthode de la moyenne mobile.

### 3.1.3 Moyenne Mobile classique

D'abord, la méthode de pondération définie plus haut, permet d'obtenir les tendances préentes sur la figure 9, qui sont issues des différents jeux de données. Ce résultat est obtenu à partir du code (ANNEXE)

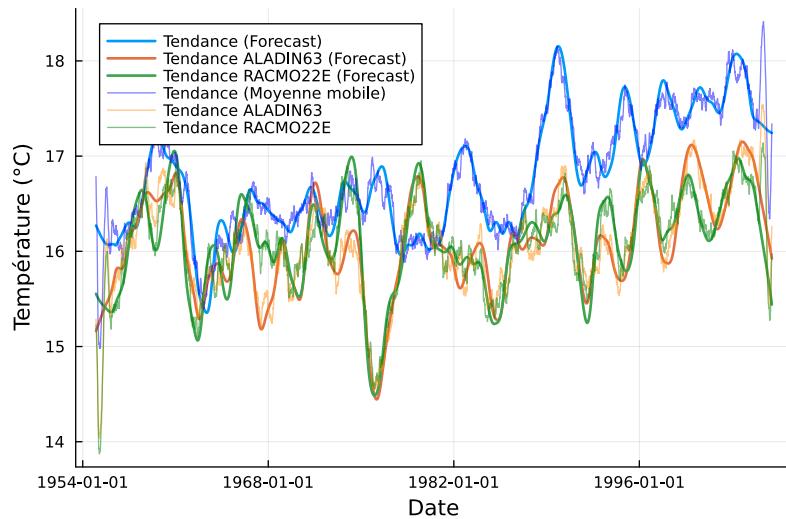


FIGURE 9 – Comparaison des tendances obtenues par Moyenne Mobile et avec le package Forecast

Sur la figure 9, nous avons appliqué la méthode détaillée plus haut avec une fenêtre de taille  $1461 = 2q + 1$  (avec  $q = 730$  soit deux ans). La fenêtre de lissage prend en compte les deux années précédentes et les deux années suivantes. C'est pour cette raison que des valeurs étranges apparaissent au début et à la fin du graphique, et ne doivent pas être prises en compte dans l'interprétation. Nous remarquons que la méthode des moyennes mobiles reproduit plutôt bien l'allure des courbes obtenues avec la fonction `stl` du package Forecast. Toutefois, la méthode des

moyennes mobiles n'offre pas des courbes aussi lisses. Nous préférerons donc continuer à travailler avec le package pour la suite.

Toutefois, cette méthode de lissage par moyennes mobiles n'est pas optimale. En effet, elle donne le même poids à tous les  $X_{t+l}$  pour  $l = -q, \dots, q$  et considère les autres à 0. On perd donc de l'information.

### 3.1.4 Moyenne mobile avec poids gaussiens

En pratique, il semble plus cohérent de donner un poids "plus fort" aux données les plus récentes (lorsque  $l$  est proche de 0) et un poids "plus faible" aux données les plus éloignées.

Dans ce cas, on peut améliorer l'expression de la série des moyennes mobiles en attribuant un poids gaussien aux valeurs  $X_t$ . On définit alors  $\hat{m}_t$  de la manière suivante :

$$\hat{m}_t = \frac{1}{S} \sum_{l=-q}^q X_{t+l} \cdot \exp\left(-\frac{l^2}{2q^2}\right)$$

Ici, on normalise en multipliant par  $\frac{1}{S}$  avec :  $S = \sum_{l=-q}^q \exp\left(-\frac{l^2}{2q^2}\right)$

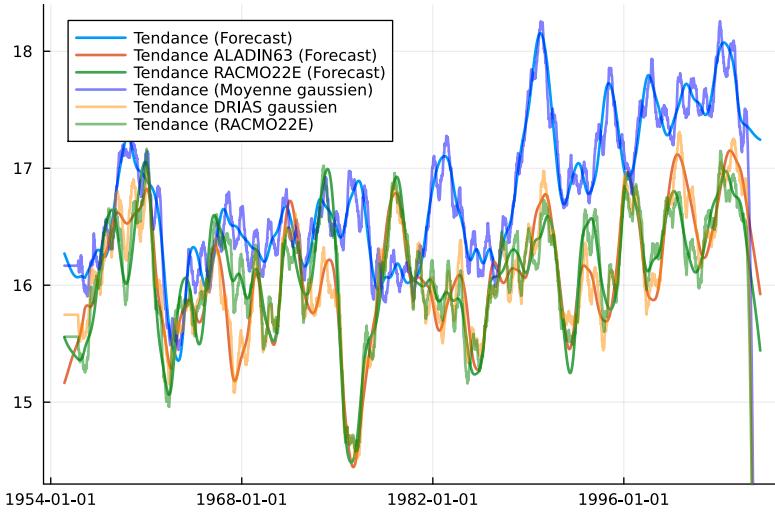


FIGURE 10 – Comparaison des tendances obtenues par Moyenne Mobile avec poids Gaussiens et Forecast

La figure 10 ressemble beaucoup à la figure 9. En effet, ces deux méthodes d'estimation de la tendance sont très proches. La méthode des moyennes mobiles semble légèrement meilleure que la méthode classique. La tendance par moyenne mobile gaussienne colle un peu plus à la tendance obtenue avec le package Forecast, mais reste toujours moins lisse.

C'est pourquoi nous préférerons dans la suite, utiliser la fonction `stl`. Il semble que cette méthode

non paramétrique d'estimation de la tendance soit la meilleure des trois méthodes détaillées dans ce rapport.

### 3.1.5 Autres estimations de la tendance

Il existe d'autres méthodes pour estimer la tendance d'une série temporelle, qui ne seront pas traitées dans ce projet, mais qui méritent d'être mentionnées :

- Estimation de la tendance par les moindres carrés [9]
- Estimation de la tendance par différenciation [10]
- Estimation de la tendance par des polynômes locaux (non-paramétrique)
- Estimation de la tendance par la méthode des noyaux (non-paramétrique)

## 3.2 La saisonnalité

La saisonnalité reflète la présence d'un phénomène périodique qui se répète au long de la série temporelle. Sur la figure 11, nous observons clairement cette périodicité. Chaque période (chaque vague) représente une année, soit 365 jours. Cela n'a rien de surprenant, la température varie tout au long de l'année, de froid en hiver à chaud en été. Et ce phénomène se reproduit chaque année, ce qui donne lieu à une telle sinusoïde.

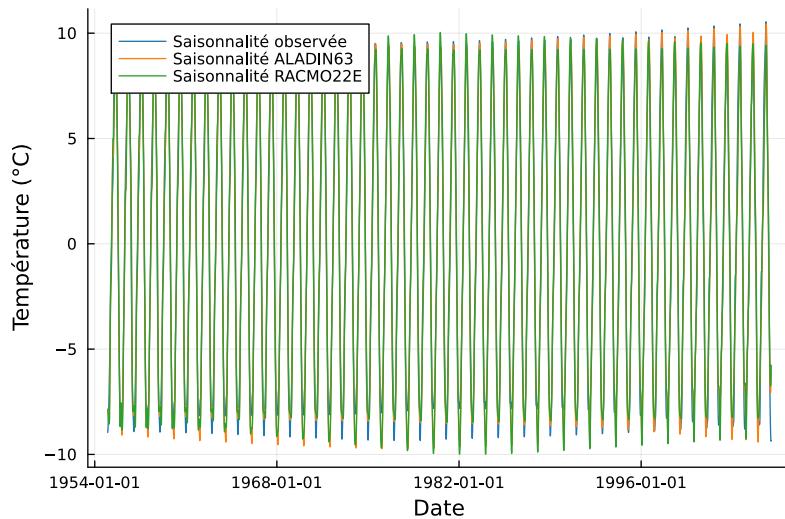


FIGURE 11 – Saisonnalités obtenues à l'aide de la fonction `stl` du package Forecast

Sur la figure 11, la courbe des saisonnalités varie environ entre  $-10$  et  $10$  degrés sur l'axe des ordonnées. Lorsqu'on retire la tendance (figure 8) pour n'obtenir que la saisonnalité, nous observons alors des valeurs de températures plus basses que dans la série temporelle initiale. La tendance prend en compte des constantes, faisant "monter" la saisonnalité dans nos séries temporelles non-décomposées (figure 7).

De plus, les trois saisonnalités représentées sur la figure 11 semblent confondues. À l'exception de

quelques valeurs atypiques, les modèles ALADIN63 et RACMO22E semblent bien reproduire la saisonnalité de la série temporelle des données observées.

### 3.3 Résidus

Dans la suite, nous retirons la tendance et la saisonnalité des séries temporelles étudiées. On obtient alors la série résiduelle (ou serie des résidus). Il s'agit de toutes les informations qui ne font ni partie de la tendance, ni de la saisonnalité.

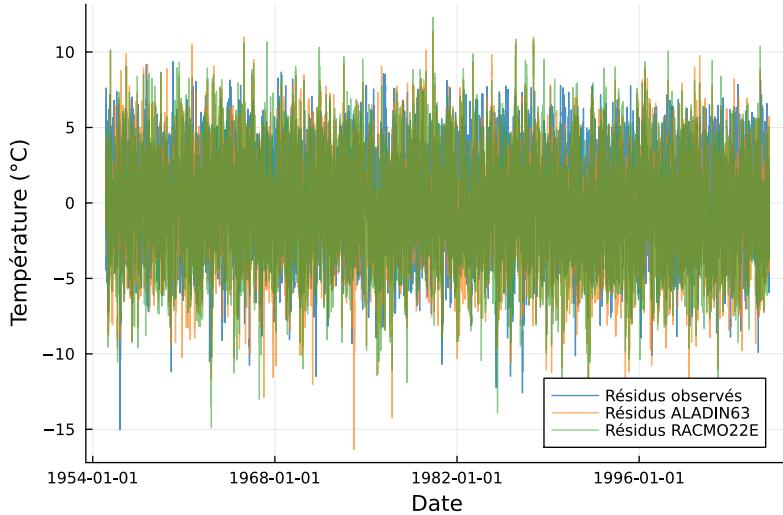


FIGURE 12 – Résidus obtenus à l'aide de la fonction stl du package Forecast

La figure 12 représente les résidus obtenus pour les trois modèles étudiés. En calculant numériquement leur espérance, nous pouvons dire que les résidus sont centrés (avec une espérance presque nulle pour les données observées, et inférieure à 0.4 pour les deux jeux de données simulées ALADIN63 et RACMO22E).

## 4 Processus Autorégressifs

**Définition 5.** Soit  $p$  un entier non nul. On dit que  $(X_t)$  est un processus autorégressif d'ordre  $p$ , noté  $AR(p)$ , si  $(X_t)$  est un processus stationnaire au second ordre solution de l'équation :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t$$

avec  $Z_t \sim \mathcal{N}(0, \sigma^2)$  et  $\phi_1, \dots, \phi_p \in \mathbb{R}$  et  $\phi_p \neq 0$ .

### 4.1 Autocorrélations

L'autocorrélation est une mesure de l'association entre des valeurs de séries actuelles et passées. Autrement dit, elle fait référence au fait que dans une série temporelle, la mesure d'un phénomène

à un instant  $t$  peut être corrélée aux mesures précédentes ou aux mesures suivantes. Une série autocorrélée est ainsi corrélée à elle-même, avec un décalage (lag) donné. L'autocorrélation indique les valeurs de séries passées les plus utiles à la prévision de valeurs futures.

#### 4.1.1 Autocorrélations simples

**Définition 6.** Fonction d'autocorrélation (ACF). Au décalage  $k$ , il s'agit de la corrélation entre les valeurs de séries séparées par  $k$  intervalles.

Soit notre processus  $(X_t)$  AR( $p$ ). Alors, sa fonction d'autocorrélation est donnée par

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\mathbb{E}[X_t X_{t-h}]}{\mathbb{E}[X_t^2]}$$

où  $\gamma$  est la fonction d'autocovariance.

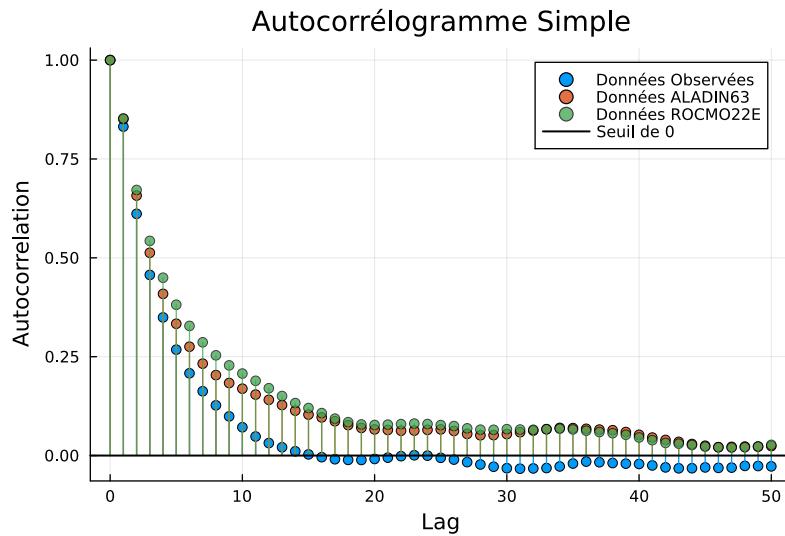


FIGURE 13 – Autocorrélogramme Empirique

Sur l'autocorrélogramme simple (Figure 13) nous observons une décroissance exponentielle.

#### 4.1.2 Autocorrélations partielles

**Définition 7.** Fonction d'autocorrélation partielle (PACF). Au décalage  $k$ , il s'agit de la corrélation entre les valeurs de séries séparées par  $k$  intervalles, compte tenu des valeurs des intervalles intermédiaires.

Soit notre processus autorégressif  $(X_t)_{t \in \mathbb{Z}}$ . En considérant la projection orthogonale de  $X_t$  sur le sous espace vectoriel engendré par  $X_{t-1}, \dots, X_{t-h}$ , notée  $X_{t,h}^*$ , de  $L^2$  muni du produit scalaire  $\langle X, Y \rangle = \mathbb{E}[X, Y]$ , nous obtenons :

$$X_{t,h}^* = a_{1,h}X_{t-1} + \dots + a_{h,h}X_{t-h} \quad \text{et}$$

$$(X_t - X_{t,h}^*) \perp X_{t,h}^*$$

Nous définissons alors l'autocorrélation partielle d'ordre  $h$  comme le coefficient  $a_{h,h}$  dans la régression de  $X_t$  sur  $X_{t-1}, \dots, X_{t-h}$ , pour  $h \neq 0$ .

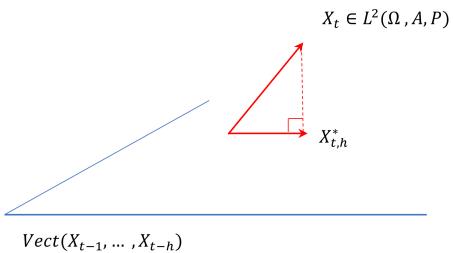


FIGURE 14 – Illustration de la projection

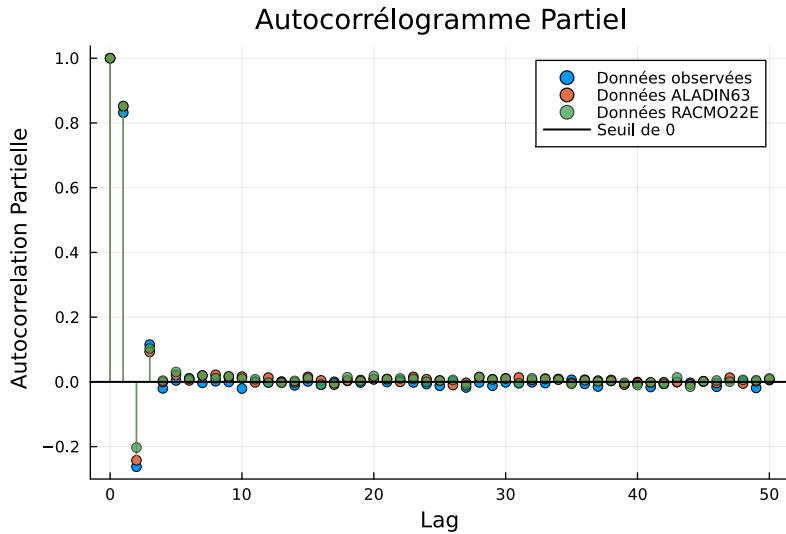


FIGURE 15 – Autocorrélogramme Partiel

L'autocorrélogramme partiel d'un processus AR( $p$ ) permet d'identifier des ordres candidats de  $p$ . Pour cela, il suffit de regarder la longueur des pics ; plus le pic est haut, plus il doit compter dans notre jugement. Le premier pic, au pas 0, ne doit pas être pris en compte pour déterminer l'ordre. Nous remarquons donc trois autres pics, aux pas 1, 2 et 3 qui se distinguent des autres.

Ainsi, nous sélectionnons 3 modèles autorégressifs candidats : AR(1), AR(2) ou AR(3).

## 4.2 Sélection de modèle

Afin de choisir lequel de ces trois modèles retenir, nous utilisons un critère de comparaison de modèles.

**Définition 8.** *Le Critère d'Information Bayésien (BIC), est une mesure de la qualité d'un modèle statistique. Lorsque l'on estime un modèle statistique, il est possible d'augmenter la vraisemblance du modèle en ajoutant un paramètre. Le BIC permet de pénaliser les modèles en fonction du nombre de paramètres, et de la taille de l'échantillon. Il s'écrit comme suit :*

$$BIC = -2\ln(L) + k * \ln(N)$$

avec  $L$  la vraisemblance du modèle estimée,  $N$  le nombre d'observations dans l'échantillon et  $k$  le nombre de paramètres libres du modèle.

Le modèle qui sera sélectionné est celui qui minimise le critère BIC, soit :

$$M_{BIC} = \arg \min_M BIC(M)$$

Nous obtenons alors les résultats suivants :

	AR (1)	AR(2)	AR(3)	AR(4)
Observées	24 162	23 217	23 051	23 055
ALADIN63	26 328	25 492	25 383	25 392
RACMO22E	26 407	25 829	25 688	25 697

TABLE 1 – Comparaison des BIC pour les différents modèles

En comparant les valeurs obtenues pour les BIC en fonction des ordres des différents processus autorégressifs, nous choisissons de sélectionner l'AR(3). En effet, c'est le modèle qui présente le plus petit BIC, et ce, pour les trois jeux de données étudiés.

## 4.3 Estimation des coefficients

### 4.3.1 Par la méthode du maximum de vraisemblance

Dans le cas où  $p = 3$ , nous parlons de processus autorégressif d'ordre 3. Le processus  $X_t$  s'exprime alors de la façon suivante :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + Z_t$$

où  $X_t$  est la valeur de la série temporelle à l'instant  $t$ ,  $\phi_1$ ,  $\phi_2$  et  $\phi_3$  sont les coefficients du modèle, et  $Z_t$  est un terme d'erreur supposé suivre une distribution normale  $N(0, \sigma^2)$ .

Nous allons chercher à estimer les coefficients  $\phi_1$ ,  $\phi_2$  et  $\phi_3$ , à l'aide de la méthode du maximum de vraisemblance, dont voici le principe :

Pour un échantillon de taille  $n$ , la fonction de vraisemblance  $L$  pour les coefficients  $\phi_1, \phi_2, \phi_3$  et  $\sigma^2$  est donnée par :

$$L(\phi_1, \phi_2, \phi_3, \sigma^2) = \prod_{t=4}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \phi_3 X_{t-3})^2}{2\sigma^2}\right)$$

Il est souvent plus pratique de maximiser la log-vraisemblance  $\ln L$  plutôt que la vraisemblance elle-même. La fonction de log-vraisemblance est :

$$\ln L(\phi_1, \phi_2, \phi_3, \sigma^2) = -\frac{(n-2)}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=4}^n (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \phi_3 X_{t-3})^2$$

Pour estimer les paramètres, nous maximisons la log-vraisemblance par rapport à  $\phi_1, \phi_2, \phi_3$ , et  $\sigma^2$ . Cette maximisation peut être réalisée à l'aide d'algorithmes d'optimisation numérique.

#### 4.3.2 Par la méthode des moindres carrés

Pour estimer les coefficients  $\phi_1, \phi_2$  et  $\phi_3$  grâce à la méthode des moindres carrés, nous reformulons le modèle en termes de matrices et vecteurs.

Pour  $t = 4, \dots, n$  (où  $n$  est la longueur des résidus), nous pouvons écrire :

$$X_t = \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ X_{t-3} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix} + Z_t$$

En notation matricielle, cela devient  $X = A\phi + Z$  où  $X$  est un vecteur  $(n-3) \times 1$  des valeurs cibles  $X_4, X_5, \dots, X_n$ ,  $A$  est une matrice  $(n-3) \times 3$  contenant les valeurs retardées :

$$A = \begin{pmatrix} X_3 & X_2 & X_1 \\ X_4 & X_3 & X_2 \\ \vdots & \vdots & \vdots \\ X_{n-1} & X_{n-2} & X_{n-3} \end{pmatrix}$$

$\phi$  est le vecteur des coefficients :

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix}$$

et  $Z$  est le vecteur des termes d'erreur.

On estime alors les coefficients  $\phi_1, \phi_2$  et  $\phi_3$  par les moindres carrés ordinaires, avec pour objectif de minimiser la somme des carrés des résidus  $Z$ . En termes de matrice, cela revient à minimiser :

$$S(\phi) = (X - A\phi)^\top (X - A\phi)$$

La solution de ce problème de minimisation est donnée par la formule des moindres carrés :

$$\hat{\phi} = (A^\top A)^{-1} A^\top X$$

Nous obtenons finalement numériquement les coefficients estimés pour les deux méthodes dans le Tableau 2, que nous comparons à ceux obtenus avec la fonction `ar` du package Forecast.

	Données observées			Données ALADIN63			Données RACMO22E		
	$\phi_1$	$\phi_2$	$\phi_3$	$\phi_1$	$\phi_2$	$\phi_3$	$\phi_1$	$\phi_2$	$\phi_3$
Méthode max de vraisemblance	1.080	-0.383	0.115	1.085	-0.342	0.097	1.051	-0.309	0.107
Méthode moindres carrés	1.080	-0.383	0.115	1.084	-0.342	0.097	1.051	-0.309	0.107
Fonction <code>ar</code> Forecast	1.080	-0.383	0.115	1.080	-0.340	0.093	1.047	-0.308	0.103

TABLE 2 – Valeurs estimées pour les coefficients  $\phi_1$ ,  $\phi_2$  et  $\phi_3$

Par défaut la fonction `ar` du package Forecast estime les coefficients d'un processus AR( $p$ ) par la méthode des moindres carrés. Mais elle peut aussi faire appel à d'autres méthodes, comme celle du maximum de vraisemblance ou à celle des équations de Yule-Walker (que nous ne détaillerons pas dans ce rapport), dans des situations spécifiques où ces méthodes offrent des avantages en termes de rapidité ou de précision des estimations.

Les trois méthodes utilisées pour l'estimation des coefficients  $\phi_1$ ,  $\phi_2$ , et  $\phi_3$  de nos AR(3) offrent des résultats plutôt proches. Ce qui est plutôt bon signe car les résultats ne risquent pas de beaucoup varier en fonction de la méthode utilisée.

Dans le tableau 2 nous remarquons que les coefficients estimés pour les données observées, le modèle ALADIN63 et RACMO22E sont assez proches (à  $10^{-3}$  près). Les coefficients  $\phi_1$ ,  $\phi_2$ , et  $\phi_3$  du modèle ALADIN63 sont plus proches de ceux des données observées, que pour le modèle RACMO22E.

Pour conclure cette analyse des résidus, nous remarquons que les processus autorégressifs issus des modèles simulés sont très proches du processus issu des données observées. Ainsi les résidus des données simulées ont un comportement très similaire à celui du jeu de données observé.

## 5 Corrélation

### 5.1 Moyennes mensuelles

Une fois l'analyse des séries temporelles sur la moyenne de toutes les stations effectuées, nous nous sommes intéressés à la bonne représentation des données historiques sur certaines stations en particulier.

Pour amorcer notre analyse par une étape simple, nous commençons par étudier la bonne reproduction de la moyenne mensuelle par les données simulées (DRIAS). Cette première étape nous permettra d'obtenir une vue d'ensemble initiale avant d'approfondir notre analyse comparative entre les deux ensembles de données. Nous choisissons donc de comparer les moyennes mensuelles obtenues pour deux villes françaises, Montpellier et Strasbourg. Nous avons choisi de regarder ces deux stations puisqu'elles sont géographiquement éloignées. De cette façon, nous espérons pouvoir tirer des informations sur la reproduction des corrélations spatiales par les modèles du DRIAS.

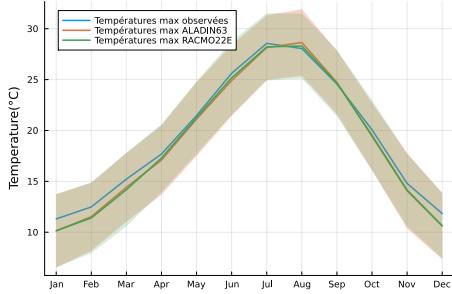


FIGURE 16 – Moyenne mensuelle des températures maximales à Montpellier

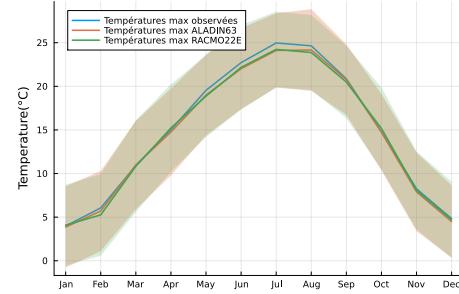


FIGURE 17 – Moyenne mensuelle des températures maximales à Strasbourg

Nous avons donc affiché la moyenne mensuelle des températures maximales à Montpellier et à Strasbourg, respectivement sur les figures 16 et 17. À première vue, les données simulées reproduisent bien les moyennes mensuelles des températures observées. Néanmoins nous pouvons noter que les estimations par les données simulées sont différentes entre les deux villes. En effet, il semble que les données simulées collent d'avantage aux données observées en été qu'en hiver à Montpellier, et que ce soit l'inverse à Strasbourg (c'est-à-dire que les données simulées semblent mieux représenter les données observées en hiver qu'en été). Cette première remarque soulève la question de bonne représentation spatio-temporelle des données observées.

Afin d'approfondir notre réflexion, nous avons distingué plusieurs zones géographiques sur la carte 3 : les villes Nord, du Sud, en bord de mer (littoral), et à l'intérieur du territoires (non littoral). Ces zones géographiques regroupent les stations suivantes :

- **Nord** : Lille, Caen, Orly, Strasbourg
- **Sud** : Ajaccio, Marignane, Montpellier, Perpignan, Toulouse
- **Littoral** : Brest, Nantes, Bordeaux, Perpignan, Montpellier, Marignane, Ajaccio
- **Non-littoral** : Lille, Caen, Orly, Bourges, Strasbourg, Dijon, Clermont-Ferrand, Toulouse

Nous les comparons graphiquement en représentant les moyennes mesuelles.

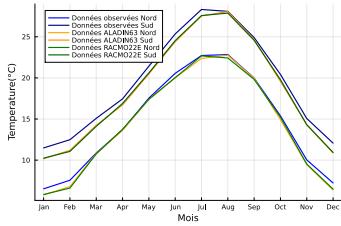


FIGURE 18 – Moyennes mensuelles Sud / Nord

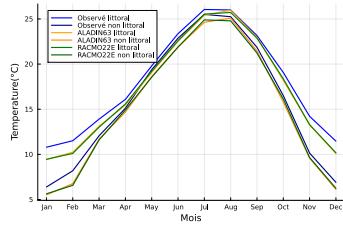


FIGURE 19 – Moyennes mensuelles littoral / non-littoral

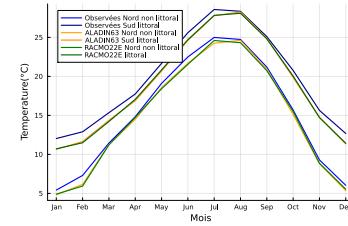


FIGURE 20 – Moyennes mensuelles Sud-littoral / Nord-non-littoral

Sur la figure 18, nous comparons les moyennes mensuelles des villes du Nord et du Sud. Les données simulées reproduisent plutôt bien la différence de température observée entre ces deux

groupes de stations. Nous remarquons tout de même que les données simulées semblent légèrement sous estimer les températures maximales observées pour les villes du Sud de la France.

Nous comparons ensuite les moyennes mensuelles pour les villes du littoral et les villes qui ne sont pas sur le littoral sur la figure 19. Premièrement, nous ne remarquons pas de différence de température significative entre les villes du littoral et les villes continentales. Ensuite, les modèles du DRIAS semblent de nouveau bien reproduire les températures observées.

Finalement, nous distinguons (sur la figure 20) les villes Sud, sur le littoral, et les villes du Nord à l'intérieur des terres. Comme précédemment, les données des modèles ALADIN63 et RACMO22E semble plutôt bien reproduire les corrélations spatiales, mais sous-estiment légèrement les températures moyennes durant l'hiver et le printemps.

## 5.2 Corrélations spatiales

Afin de vérifier la bonne reproduction des corrélations spatiales par les modèles simulés issus du DRIAS, nous proposons de représenter la matrice des différences des corrélations observées et simulées. Pour cela, nous représentons sur une grille, toutes nos stations. Nous calculons ensuite les corrélations (avec la fonction `cor()` de Julia) des stations deux à deux, et on fait de même pour chacun des modèles. Nous obtenons alors trois matrices de corrélations. En faisant la différence entre la matrice des corrélations pour le modèle observé, et la matrice des corrélations pour le modèle simulé, nous obtenons une matrice indiquant la bonne ou mauvaise reproduction des corrélations spatiales observées, par le modèle. Plus la case est claire, plus la reproduction de la corrélation spatiale est bonne.

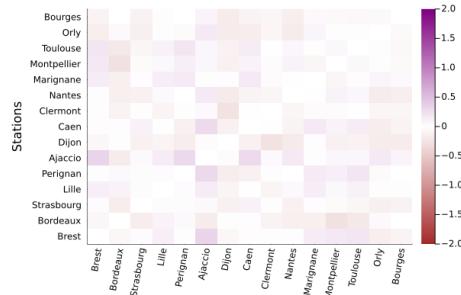


FIGURE 21 – Matrice des corrélations spatiales (températures maximales) - modèle ALADIN63

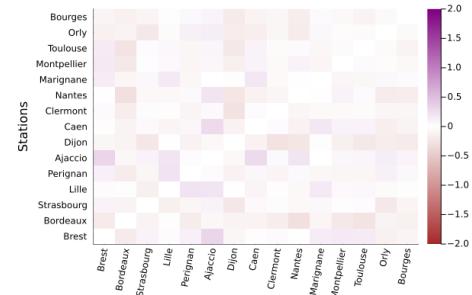


FIGURE 22 – Matrice des corrélations spatiales (températures maximales) - modèle RACMO22E

À la vue de ces deux graphiques, le résultat paraît satisfaisant. En effet, les couleurs sont globalement très claires ce qui indique de petites différences ( $<0.4$ ) entre les corrélations spatiales du jeu de données observé et simulé.

## 6 Précipitations

Les précipitations désignent toutes les formes de l'eau à l'état liquide ou solide provenant de l'atmosphère. Elles se caractérisent par trois principaux paramètres : leur volume, leur intensité et leur fréquence qui varient selon les lieux et les périodes.

Dans cette section, nous nous intéresserons seulement à une forme de précipitation : les pluies, mesurées en  $mm/m^2$ .

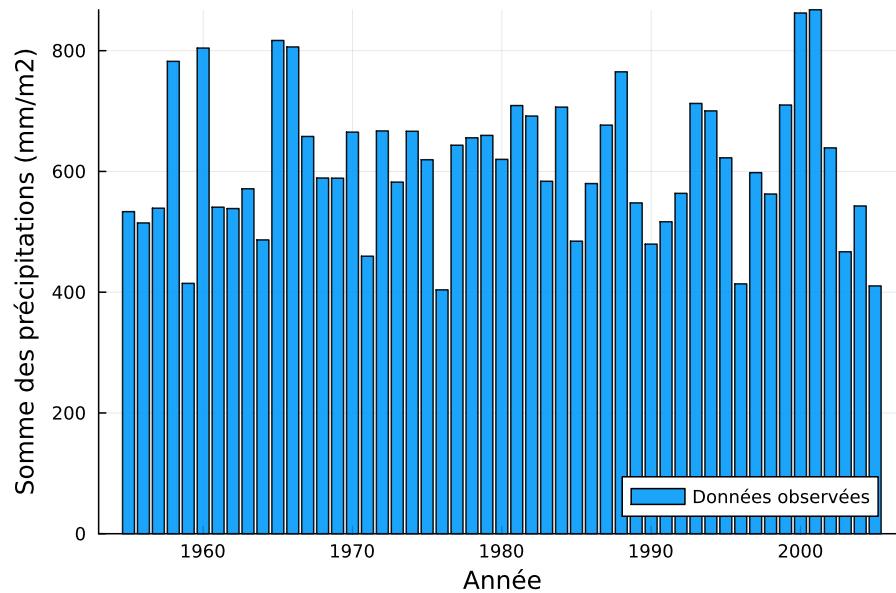


FIGURE 23 – Sommes des précipitations sur toutes les stations par année

L'histogramme 23 représente la somme moyenne des précipitations observées sur toutes les stations étudiées. Autrement dit, nous sommes la quantité de pluie par station, et en faisons la moyenne avec toutes les stations. Nous n'avons représenté que les précipitations observées, mais cela est suffisant pour se rendre compte de l'irrégularité du phénomène pluvieux. La figure 23 met ainsi en évidence la forte variabilité interannuelle des précipitations, à l'échelle du territoire Français.

L'approche utilisée pour les températures quotidiennes maximales ne semble alors pas très adaptée aux précipitations. En effet, la décomposition de la série temporelle de ce phénomène en tendance et saisonnalité n'offre pas des résultats très informatifs :

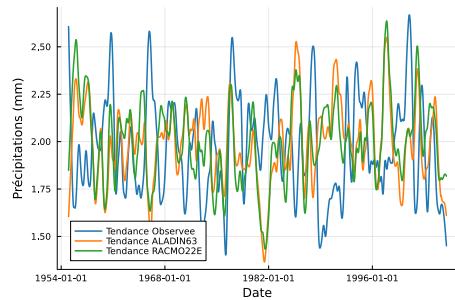


FIGURE 24 – Tendances des précipitations obtenues à l'aide de la fonction `stl`

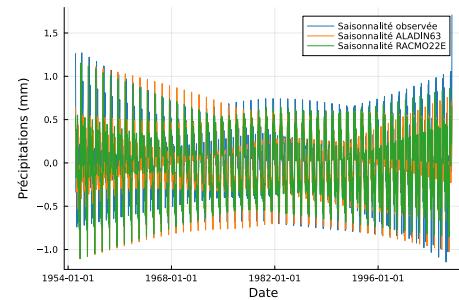


FIGURE 25 – Saisonnalités des précipitations obtenues à l'aide de la fonction `stl`

Les figures 24 et 25 représentent respectivement les tendances et les saisonsnalités pour les précipitations des trois modèles étudiés. Il ne semble pas y avoir de tendance pour les précipitations. En ce qui concerne la saisonnalité, nous remarquons un motif, qui n'est pas sinusoïdal.

Afin d'étudier la bonne reproduction des précipitations observées par les modèles simulés, nous affichons leur différence. C'est-à-dire que nous affichons *Données simulées – Données observées* pour les modèles ALADIN63 et RACMO22E.

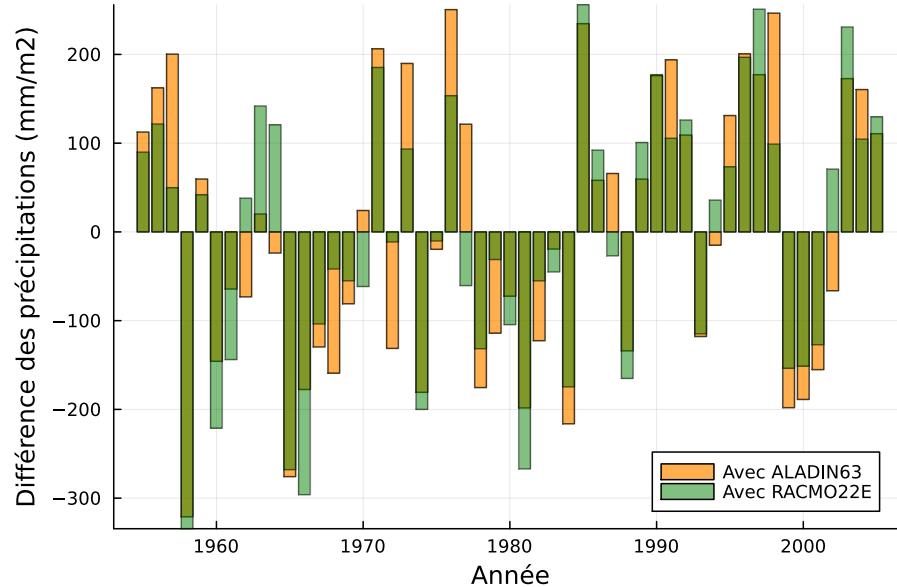


FIGURE 26 – Différence des précipitations simulées et observées

Ainsi lorsque, sur la figure 26, les barres sont au dessus de l'axe 0, les modèles surestiment la quantité moyenne de pluie par an, tandis que lorsque la barre est en dessous de l'axe 0, les modèles la sous-estime.

Pour étudier la représentation des corrélations spatiales par les modèles ALADIN63 et RACMO22E, nous reprenons la même méthode utilisée pour les températures maximales. Nous rappelons qu'il s'agit de matrices de différence des corrélations entre les corrélations des stations pour les données observées et les corrélations des stations pour les données. De nouveau, plus la case est foncée, moins la reproduction de la corrélation spatiale est bonne, et plus la case est claire, plus la reproduction de la corrélation spatiale est bonne.

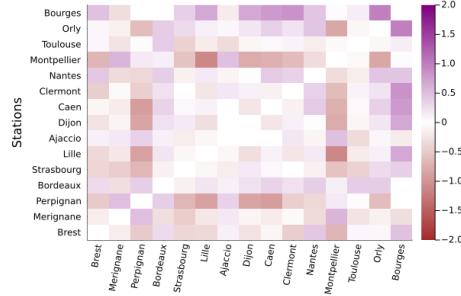


FIGURE 27 – Matrice des corrélations spatiales (pluies) - modèle ALADIN63

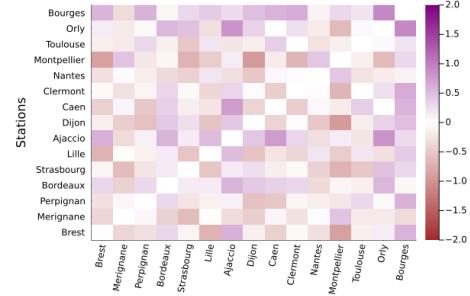


FIGURE 28 – Matrice des corrélations spatiales (pluies) - modèle RACMO22E

En comparant les figures 27 et 28, nous remarquons que le modèle ALADIN63 semble mieux reproduire les corrélations spatiales pour les précipitations. Entre une même ville pour le modèle observé et le modèle simulé, la corrélation est très bonne (une différence nulle ou presque).

Les corrélations spatiales pour les précipitations sont moins bien reproduites que pour les températures maximales pour les deux modèles issus des données simulées. Les épisodes pluvieux sont plus difficiles à prédire, et souvent très localisés.

Les modèles simulés ne reproduisent que partiellement les corrélations spatiales.

## 7 Extrêmes

**Définition 9.** *Un événement climatique extrême est caractérisé par sa rareté, son intensité ou les dégâts qu'il provoque. Les canicules, vagues de froid, cyclones tropicaux, sécheresses en sont des exemples. La fréquence et l'intensité de certains événements climatiques extrêmes augmentent en conséquence du dérèglement climatique d'origine anthropique.*

Dans cette partie, nous nous intéresserons aux vagues de chaleur. Il s'agit de températures maximales anormalement élevées sur une période de plusieurs jours (15 dans notre exemple). Ces événements sont rares, et sûrement difficiles à prédire. C'est pourquoi nous nous interrogeons sur la bonne reproductibilité des vagues de chaleurs par les modèles ALADIN63 et RACMO22E. Nous travaillons de nouveau sur la moyenne des températures maximales de toutes les stations étudiées. Pour cela, nous avons, pour chaque année, effectué une moyenne glissante sur une période de 15 jours entre le 1 juin et le 31 août. Autrement dit, chaque point par année est obtenu de la manière suivante :

$$\max(\text{moyenne } (1 - 15 \text{ juin}), \text{moyenne } (2 - 16 \text{ juin}), \dots, \text{moyenne } (16 - 31 \text{ août}))$$

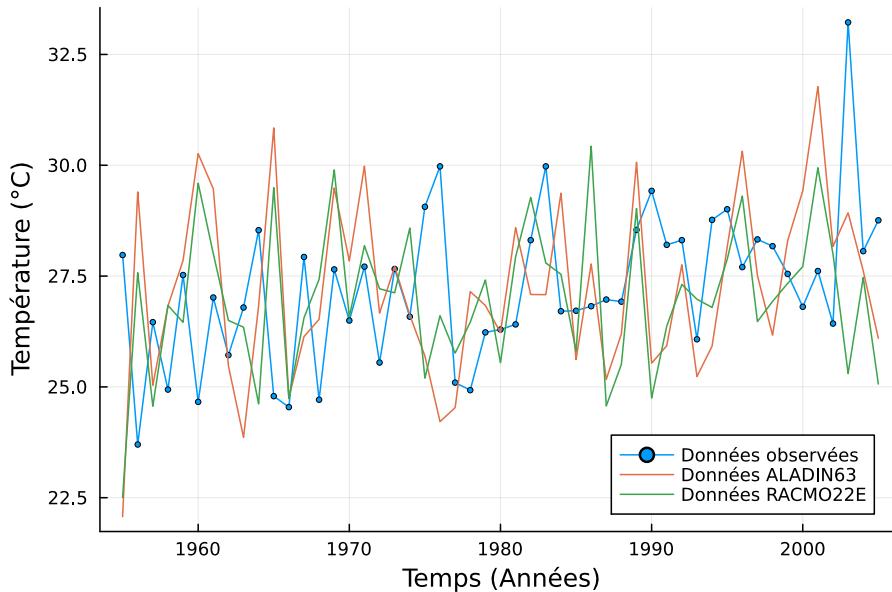


FIGURE 29 – Vagues de chaleurs observées et simulées

La figure 29 représente les vagues de chaleurs observées et simulées entre 1955 et 2005, sur la moyenne des stations étudiées. La vague de chaleur de 2003, d'une intensité exceptionnelle, survenue durant la première quinzaine d'août avec des températures maximales et minimales significativement au-dessus des normales saisonnières<sup>2</sup>, est clairement remarquable sur ce graphique. D'autant plus

2. D'après *Santé publique France*, "La vague de chaleur d'août 2003 : que s'est-il passé?", 1 juin 2004 et mis à jour le 6 septembre 2019

qu'aucun des deux modèles RACMO22E et ALADIN63 ne semblent avoir pu anticiper cet évènement extrême.

Nous remarquons en effet, que les modèles issus de données simulées ont du mal à reproduire les vagues de chaleurs observées. Sur la figure 29, les courbes orange et verte ont une allure très différente de la courbe des données observées. Ainsi, il semble que les modèles issus du DRIAS aient du mal à reproduire des évènements extrêmes comme des vagues de chaleur.

## 8 Conclusion

L'objectif de ce rapport était de comparer les données observées et simulées de température maximale quotidienne et de pluviométrie.

Nous avons utilisé les modèles ALADIN63 et RACMO22E. Nous avons commencé par vérifier la bonne représentation de la tendance et la saisonnalité par les données simulées, pour les températures maximales journalières. Une fois ces paramètres retirés, nous avons défini et travaillé avec des processus autorégressifs. Les processus issus des données simulées se rapprochaient significativement du processus issu des données simulées. L'étude de la reproduction des corrélations spatiales pour les températures maximales nous permet d'affirmer que les modèles du DRIAS reproduisent correctement ces corrélations entre stations. Toutefois, la représentation des corrélations spatiales pour les données de pluviométrie est plus difficile. De même, les évènements extrêmes sont difficiles à reproduire par les données simulées.

Ainsi, les propriétés statistiques du jeu de données simulées semblent similaires aux propriétés du jeu de données observées pour les températures maximales journalières. Toutefois, cette correspondance est moins évidente lorsqu'il s'agit de la pluviométrie ou des évènements extrêmes.

## 9 Annexes

### 9.1 Bibliographie

La partie théorique de ce rapport s'appuie largement sur le cours de Séries Temporelles dispensé par Madame Brunel, durant le semestre.

## Références

- [1] Arrhenius, Svante. « On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground ». Philosophical Magazine and Journal of Science 41 (avril 1896) : 237 76.  
[https://www.rsc.org/images/Arrhenius1896\\_tcm18-173546.pdf](https://www.rsc.org/images/Arrhenius1896_tcm18-173546.pdf)
- [2] Meadows D. H., D. L. Meadows, J. Randers, et W. W. Behrens. The Limits to Growth : A Report for the Club of Rome's Project on the Predicament of Mankind. Universe Press. New York, 1972
- [3] Westerhold, Thomas, et al. « An Astronomically Dated Record of Earth's Climate and Its Predictability over the Last 66 Million Years ». Science 369, no 6509 (11 septembre 2020) : 1383 87.  
<https://doi.org/10.1126/science.aba6853>
- [4] Site web ECA Blend  
<https://www.ecad.eu/dailydata/predefinedseries.php>
- [5] Site web DRIAS, les futurs du climat - Page des commandes  
<https://www.driias-climat.fr/commande>
- [6] M. -C. Viano, " Cours de Séries Temporelles ", *Université des Sciences et Technologies de Lille* (page 9)  
<https://math.univ-lille1.fr/~viano/economcours.pdf>
- [7] Documentation en ligne sur le package Forecast de Julia  
<https://docs.juliahub.com/Forecast/AiZvp/0.1.0/man/quickstart/>  
<https://github.com/viraltux/Forecast.jl/blob/main/src/AR.jl>
- [8] Lise Vaudor. « Régression loess », ENS Lyon. Juin 2015  
<http://perso.ens-lyon.fr/lise.vaudor/regression-loess/>
- [9] Angelina Roche, "Introduction au cours de séries temporelles", *Executive Master Statistique et Big Data*, 2018/2019  
[https://www.ceremade.dauphine.fr/~roche/SeriesTemp\\_Cours1](https://www.ceremade.dauphine.fr/~roche/SeriesTemp_Cours1)
- [10] Jean-Yves Dauxois, "Introduction à l'Étude des Séries Temporelles", *INSA Toulouse*, 2016/2017 (page 9)  
[https://perso.math.univ-toulouse.fr/jydauxoi/files/2017/04/poly\\_eleves.pdf](https://perso.math.univ-toulouse.fr/jydauxoi/files/2017/04/poly_eleves.pdf)

## 9.2 Code

Tous les programmes informatiques utilisés dans ce projet sont disponibles sur notre page Github, à l'adresse suivante : [https://github.com/jeannevivierum/Changement\\_climatique](https://github.com/jeannevivierum/Changement_climatique).

### 9.2.1 Tendance par moyenne mobile

```
function moyenne_mobile(data::Vector{T}, window_size::Int) where T
    n = length(data)
    trend = similar(data, T)
    for i in 1:n
        start_idx = max(1, i - window_size ÷ 2)
        end_idx = min(n, i + window_size ÷ 2)
        trend[i] = mean(@view(data[start_idx:end_idx]))
    end
    return trend
end
```

### 9.2.2 Tendance par moyenne mobile avec poids gaussiens

```
function moyenne_gauss(data::Vector{T}, window_size::Int) where T
    n = length(data)
    trend = similar(data, T)
    for i in 1:n
        weights = exp.(-(collect(1:window_size) .- (window_size÷2)).^2/(2 * (window_size/2)^2))
        weights/=sum(weights)
        start_idx = max(1, i - window_size ÷ 2)
        end_idx = min(n, start_idx + window_size - 1)
        data_slice = @view(data[start_idx:end_idx])
        trend[i] = sum(data_slice .* weights[1:length(data_slice)])
    end
    return trend
end
```

### 9.2.3 Estimation des coefficients $\phi_1$ , $\phi_2$ et $\phi_3$ par la méthode du maximum de vraisemblance

```
function log_likelihood(), data)
    n = length(data)
    residuals = zeros(n)
    for t in 4:n
        residuals[t] = data[t] - ([1]*data[t-1] + [2]*data[t-2] + [3]*data[t-3])
    end
    2 = mean(residuals[4:end].^2)
    logL = -n/2 * log(2*2) - sum(residuals[4:end].^2) / (2*2)
    return -logL
end
```

#### 9.2.4 Estimation des coefficients $\phi_1$ , $\phi_2$ et $\phi_3$ par les moindres carrés ordinaires

```
n = length(residus)
X = hcat(residus[3:end-1], residus[2:end-2], residus[1:end-3])
X_drias = hcat(residus_drias[3:end-1], residus_drias[2:end-2], residus_drias[1:end-3])
X_mod2 = hcat(residus_mod2[3:end-1], residus_mod2[2:end-2], residus_mod2[1:end-3])
y_target = residus[4:end]
y_target_drias = residus_drias[4:end]
y_target_mod2 = residus_mod2[4:end]

coefficients = X \ y_target
coefficients_drias = X_drias \ y_target_drias
coefficients_mod2 = X_mod2 \ y_target_mod2
```

#### 9.2.5 Vagues de chaleur

```
function vagues_de_chaleur(df)
    moyennes_par_an = Float64[]
    for year in start_year:end_year
        moyennes_juin_aout = Float64[]
        for day_offset in 0:15:75
            start_date = Date(year, 6, 1) + Day(day_offset)
            end_date = Date(year, 8, 31) - Day(75 - day_offset)
            push!(moyennes_juin_aout, moyenne_valeurs(df, start_date, end_date))
        end
        push!(moyennes_par_an, maximum(moyennes_juin_aout))
    end
    return moyennes_par_an
end
```