

# DESCRIPTION ET PRÉVISION DE SÉRIES TEMPORELLES

## LA QUALITÉ DE L'AIR DANS LES STATIONS DE RER PARISIENNES



# SOMMAIRE

01

Introduction

02

Tendance de la série

05

Les moyennes mobiles

06

Décomposition de la série temporelle

07

Les coefficients saisonniers

08

La série désaisonnalisée

09

Les résidus

10

Prévisions

11

Conclusion

12

Résumé

13

Tâches réalisées

# INTRODUCTION

La qualité de l'air dans les stations de métro parisiennes est une préoccupation majeure, en raison de la concentration élevée de polluants dans ces espaces confinés. La station Châtelet, au cœur du réseau parisien, représente un cas d'étude emblématique en raison de son trafic intense et de sa situation centrale. Depuis 1997, la RATP mesure en continu divers paramètres de qualité de l'air pour évaluer l'impact des activités humaines et des conditions environnementales sur ces espaces.

Dans le cadre de cette mission, nous analysons deux bases de données distinctes concernant la station Châtelet :

- Une première base qui couvre la période 2013 à 2020, comprenant **70 077 observations**,
- Une seconde base qui couvre la période 2021 à 2024, comprenant **33 836 observations**.

Ces bases contiennent des mesures quantitatives sur sept variables environnementales :

- Les volumes de monoxyde d'azote (**NO**)
- de dioxyde d'azote (**NO<sub>2</sub>**)
- de dioxyde de carbone (**CO<sub>2</sub>**),
- La concentration de particules fines (**PM10**),
- Les mesures de température (**TEMP**)
- humidité relative (**HUMI**),
- La variable temporelle date/heure.

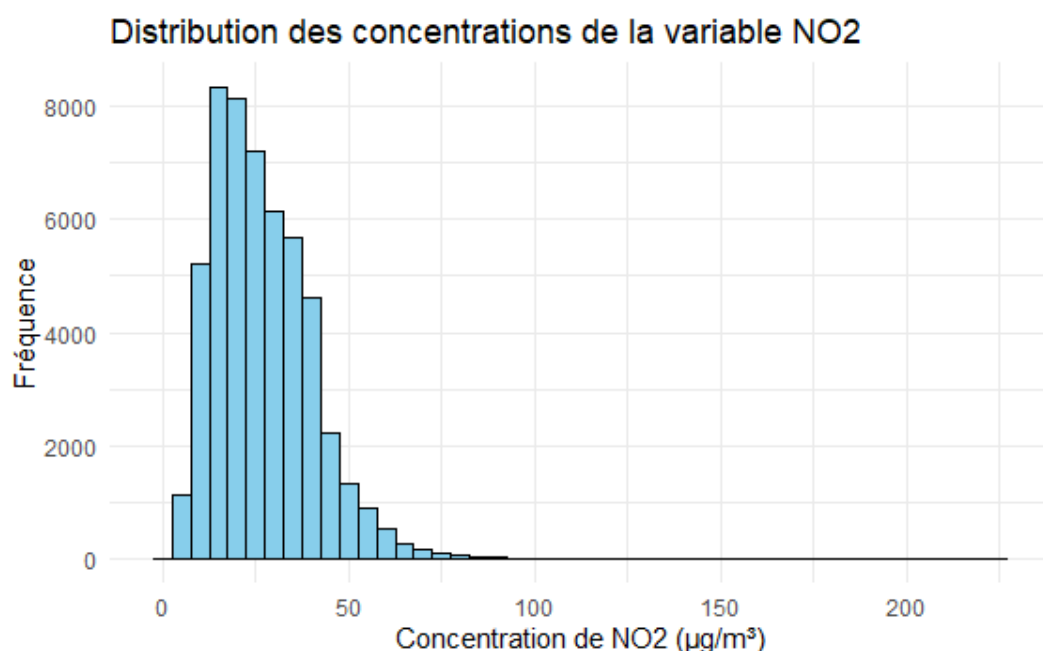
Notre analyse se concentre sur les données collectées à partir du 22 novembre 2018, en raison de données manquantes avant cette date. L'objectif principal est de comprendre l'évolution de la qualité de l'air à Châtelet sur les périodes étudiées, en identifiant les variations saisonnières, les pics de pollution, et les éventuelles tendances à long terme.

## Problématique

Comment la qualité de l'air a-t-elle évolué dans la station Châtelet entre 2018 et 2024, et quelles sont les principales variations observées?

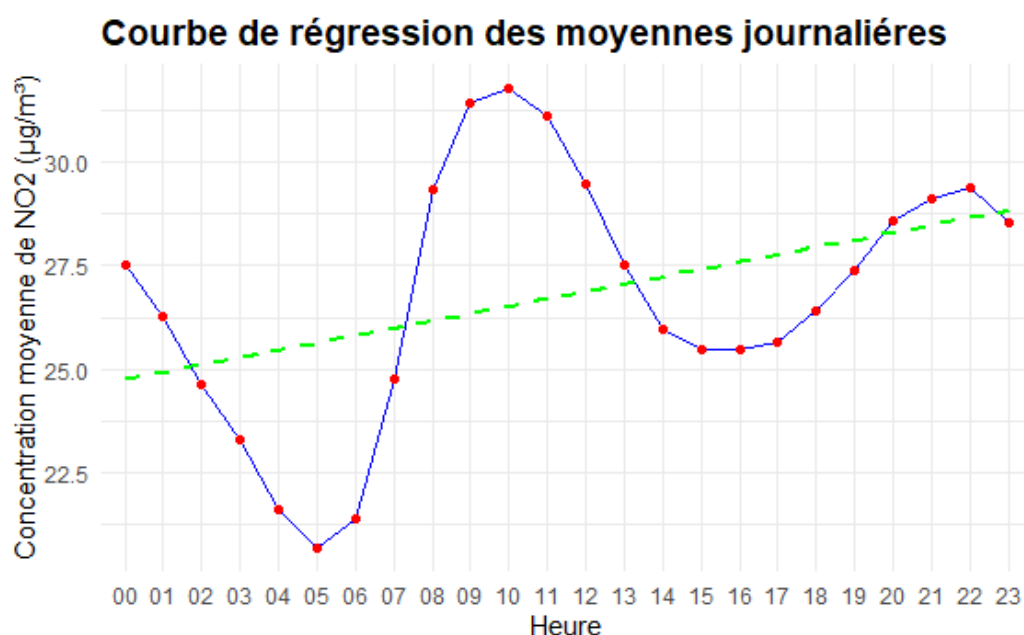
Pour répondre à cette problématique, nous allons étudier la variable **NO<sub>2</sub> (dioxyde d'azote)**, qui représente un indicateur important de la pollution de l'air.

## PRÉSENTATION DE LA VARIABLE



La majorité des concentrations de NO<sub>2</sub> se situent entre 0 et 50 µg/m<sup>3</sup>, comme le montre l'histogramme. Cela suggère que la pollution de l'air par le NO<sub>2</sub> est généralement à un niveau modéré, dans une plage où la majorité des observations se regroupent. mais il existe des valeurs extrêmes qui s'étendent vers des concentrations plus élevées.

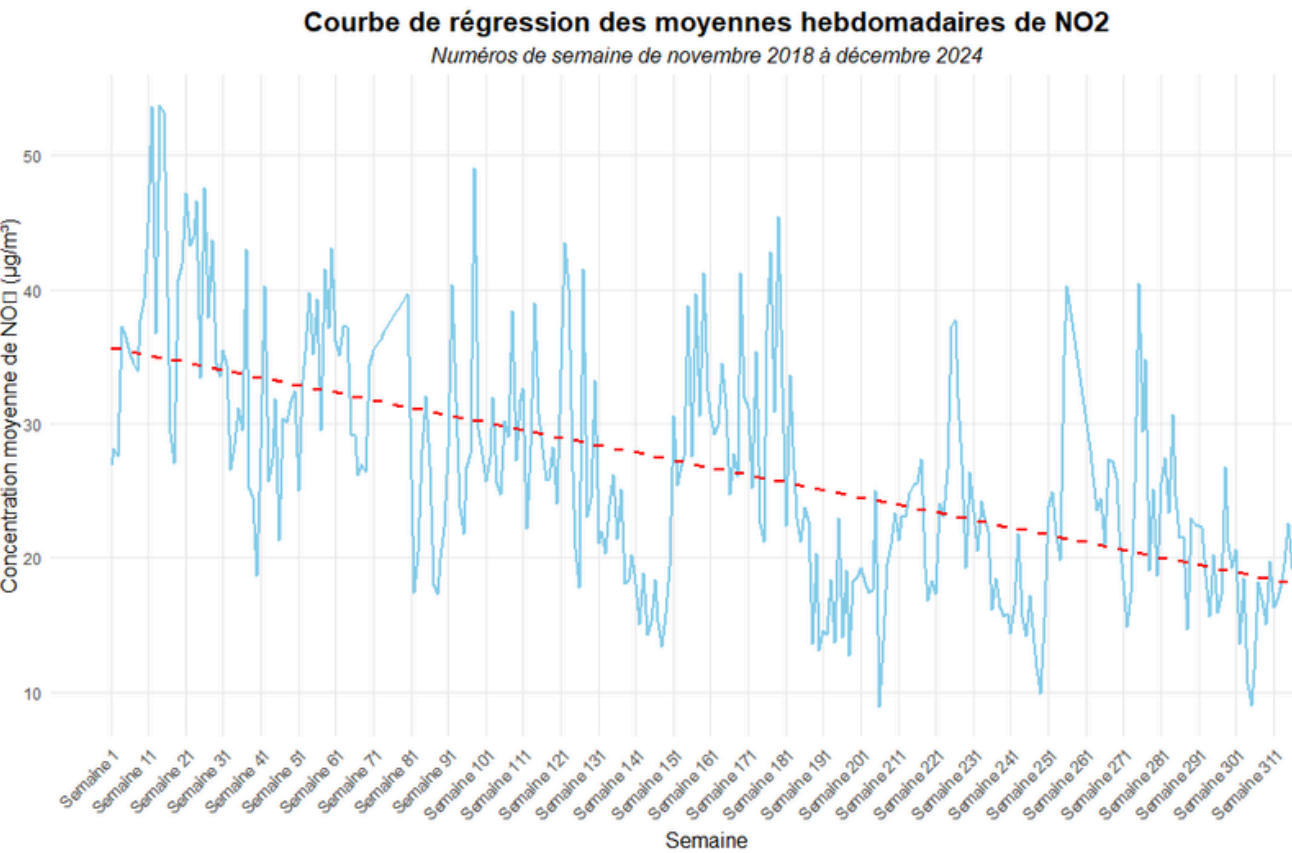
## COURBES DES MOYENNES



On constate que les moyennes journalières de NO<sub>2</sub> sont les plus faibles entre 02h00 et 06h00, périodes de fermeture du métro où l'activité humaine et ferroviaire est quasi nulle. Ces concentrations augmentent pendant les heures de pointe (08h00-09h00 et 18h00-19h00), en raison d'un afflux massif de voyageurs et d'une fréquence accrue des trains. Les frottements mécaniques des trains et la remise en suspension des particules fines amplifient ces niveaux.

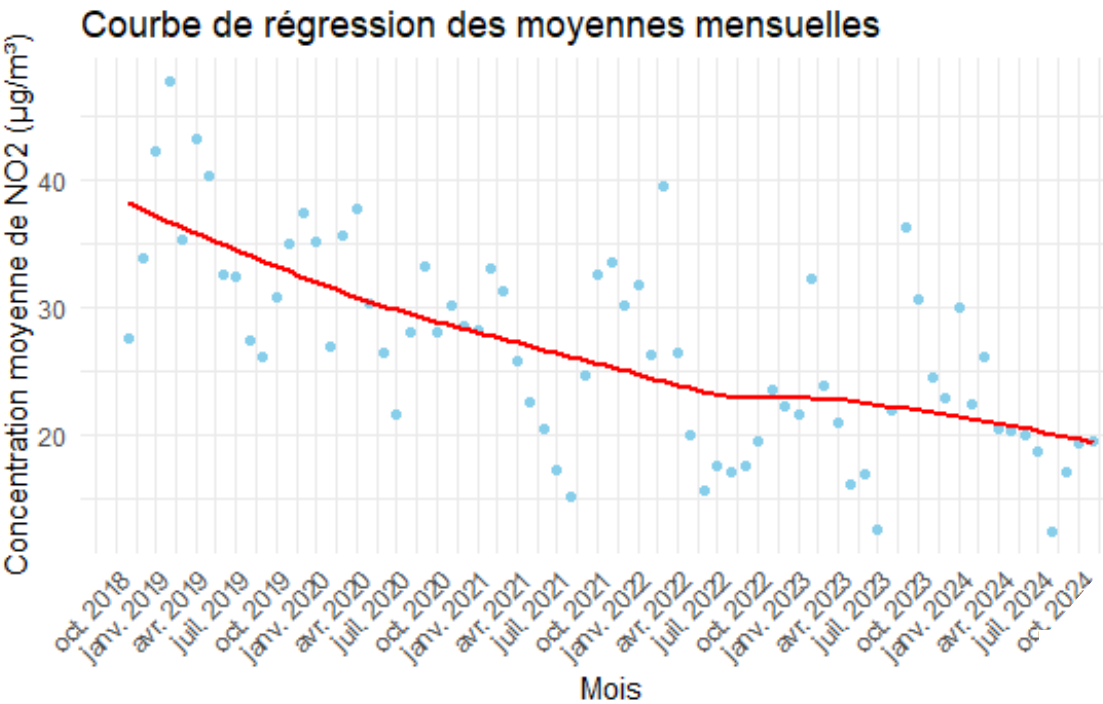
Entre les heures de pointe, une diminution notable est observée, notamment autour de midi, reflétant une réduction de l'activité globale.

On peut dire que les variations horaires des concentrations de NO<sub>2</sub> à Châtelet sont fortement corrélées à l'activité du métro et à l'afflux de voyageurs.



Dans la courbe chaque point correspond à la concentration moyenne hebdomadaire de NO<sub>2</sub> mesurée dans la station Châtelet., les variations semaine par semaine montrent des hausses et des baisses régulières. Cela peut être lié à l'affluence dans la station de métro,;

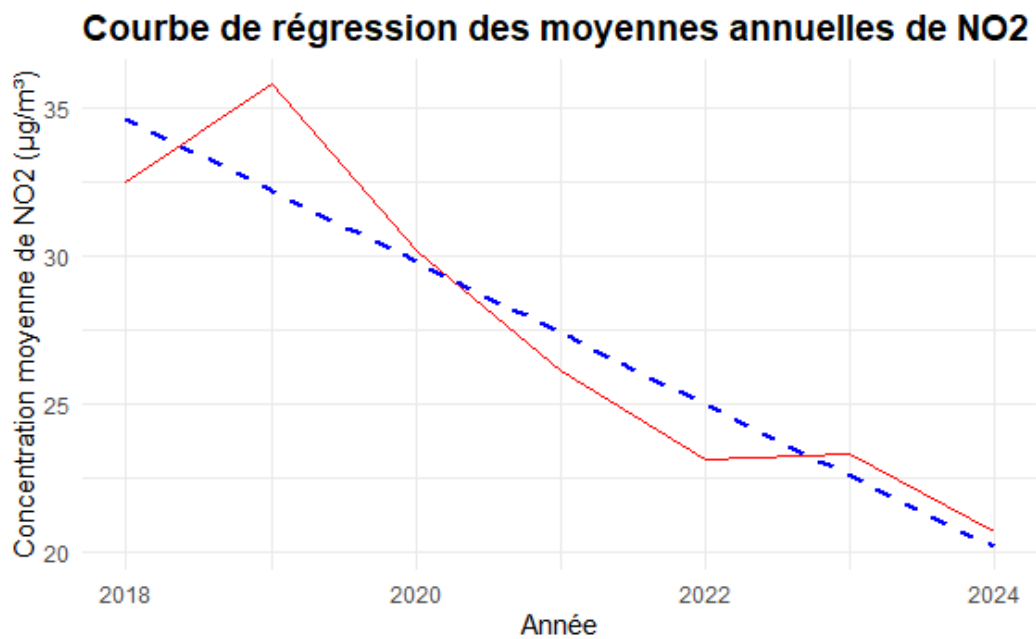
- Semaine haute : Plus d'activité (jours ouvrés, heures de pointe, évènements locaux).
- Semaine basse : Moins d'activité (week-ends, vacances scolaires, jours fériés).



En ce qui concerne les moyennes annuelles les premiers mois (2018-2019) montrent des concentrations élevées de NO2, dépassant souvent les 35-40  $\mu\text{g}/\text{m}^3$ , témoignant d'une qualité de l'air médiocre.

Une baisse nette à partir de 2020 coïncide avec les confinements liés à la pandémie de COVID-19, soulignant l'impact significatif des réductions temporaires d'activité humaine.

Vers la fin de la période, notamment en 2023-2025, les concentrations de NO2 semblent se stabiliser autour de 15-20  $\mu\text{g}/\text{m}^3$ , indiquant une amélioration durable de la qualité de l'air.



Comme pour les moyennes mensuelles, la tendance est clairement décroissante. Entre 2018-et 2019, les concentrations annuelles dépassent les 35  $\mu\text{g}/\text{m}^3$ , ce qui reste élevé.

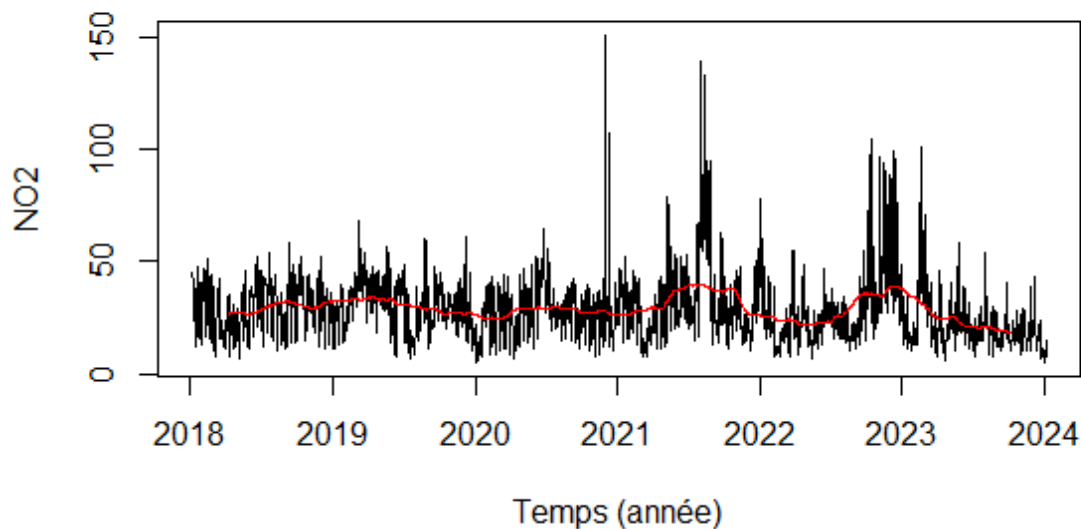
Une légère augmentation est observée en 2019 avant la baisse de 2020. Cela pourrait être lié à des facteurs spécifiques, comme une hausse de l'activité économique ou des conditions météorologiques défavorables..

En 2020, une chute brutale ramène les concentrations annuelles sous les 30  $\mu\text{g}/\text{m}^3$ .

Puis les années suivantes (2021-2024) montrent une amélioration continue, avec des concentrations qui se stabilisent autour de 20  $\mu\text{g}/\text{m}^3$ .

# LES MOYENNES MOBILES

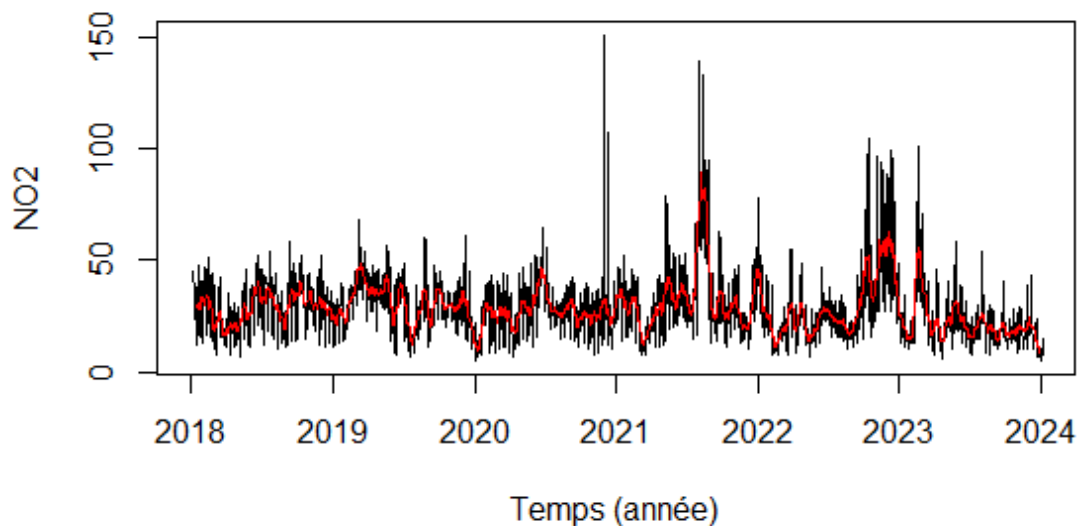
## Moyenne mobile d'ordre 7



On constate que les niveaux de NO2 sont plus élevés en semaine et diminuent légèrement le week-end. Cette différence s'explique par un trafic plus important en semaine. Le nombre de passagers dans la station augmente aussi, ce qui contribue à la pollution (proximité des rails, moteurs, etc.).

On peut dire que les fluctuations hebdomadaires suivent un cycle court, lié directement à l'activité humaine régulière.

## Moyenne mobile d'ordre 30



Les cycles observés, avec des pics réguliers sur plusieurs années, montrent une variation saisonnière forte et répétée. Les niveaux de NO2 augmentent systématiquement en hiver et diminuent en été. Les pics hivernaux sont dus à une augmentation des émissions urbaines (chauffage, trafic) et à une mauvaise dispersion des polluants causée par des températures basses et une ventilation naturelle réduite.

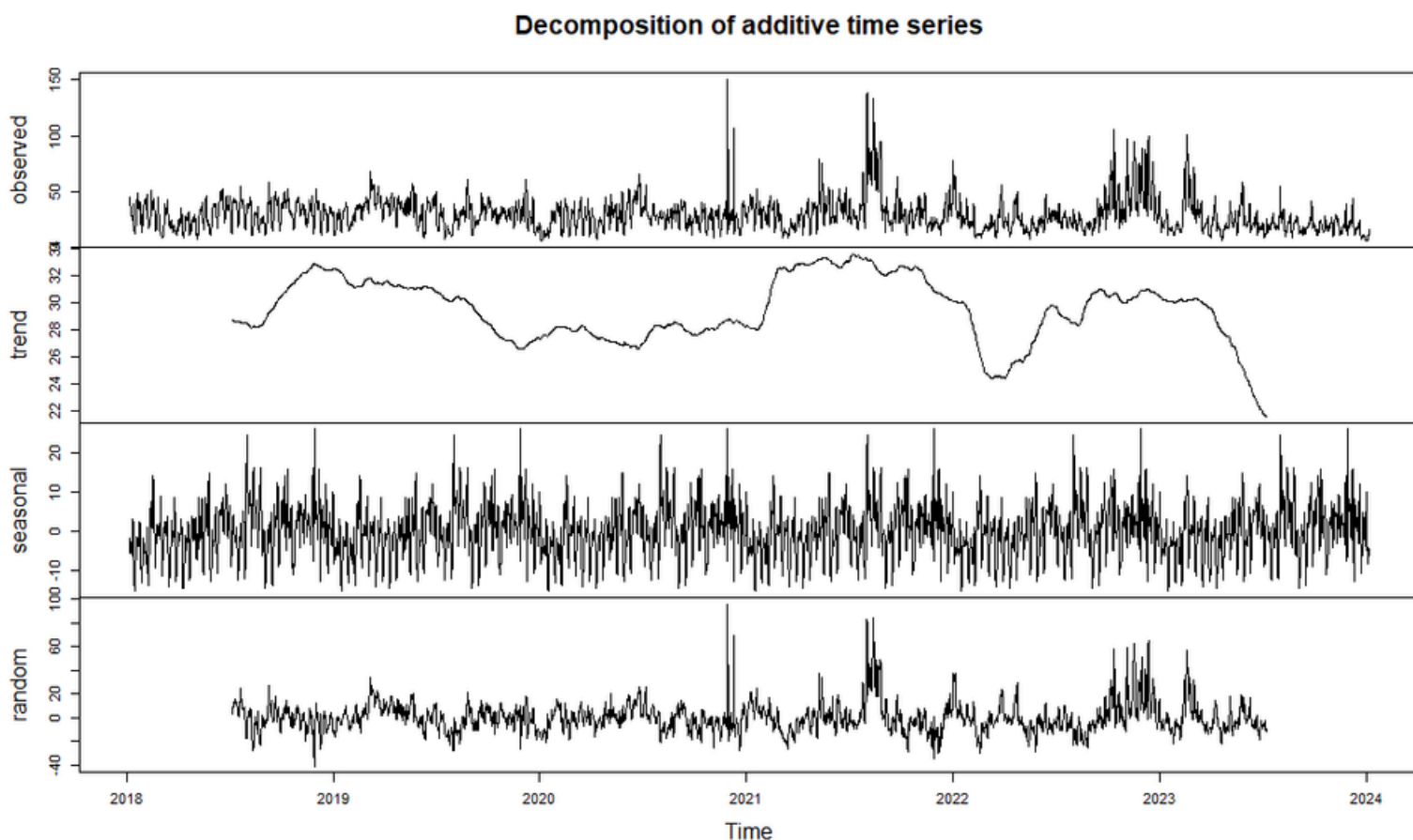
La diminution estivale reflète une meilleure dispersion des polluants grâce à une ventilation accrue et une activité urbaine moindre.

Contrairement à la moyenne mobile d'ordre 30, qui révèle les tendances saisonnières à long terme, celle d'ordre 7 met en lumière les effets immédiats des variations hebdomadaires.



# DÉCOMPOSITION DE LA SÉRIE TEMPORELLE

Ce graphique décompose la série temporelle de NO<sub>2</sub> (dioxyde d'azote) mesurée à la station Châtelet en 4 composantes principales.

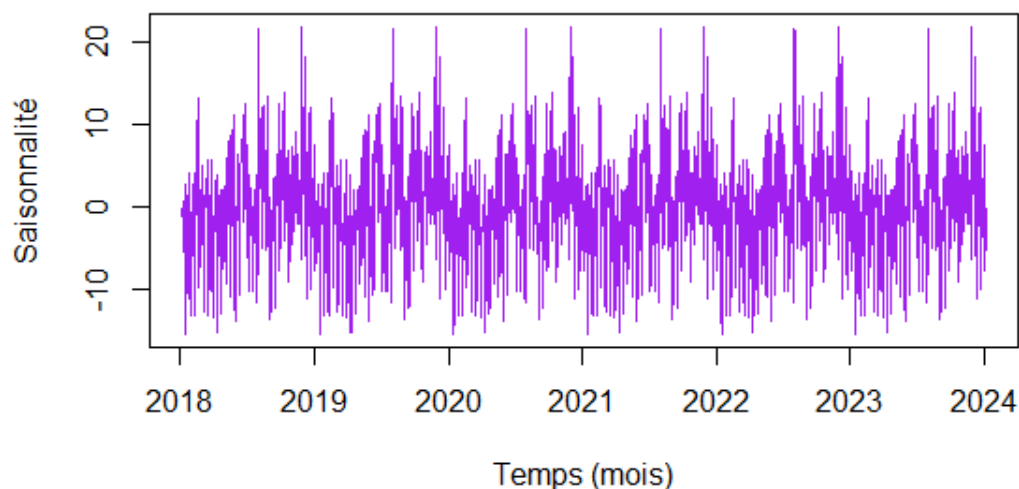


- **Observation globale (Observed):** On observe une variabilité importante sur toute la période, avec des pics notables de NO<sub>2</sub>, en particulier entre 2020 et 2022.
- **Tendance (Trend) :** La courbe de tendance indique une légère diminution des niveaux de NO<sub>2</sub> sur plusieurs années, probablement liée à des efforts pour réduire la pollution. Cependant, cette baisse reste modérée, et des pics hivernaux persistants montrent que la pollution intérieure est encore problématique.
- **Saisonnalité (Seasonal) :** la saisonnalité est clairement définie, avec une amplitude stable sur la période. Cela confirme que les concentrations de NO<sub>2</sub> suivent un cycle prévisible, influencé par des facteurs externes (climat, trafic...).
- **Résidus (Random) :** les résidus, qui représentent les anomalies, montrent des pics ponctuels. Ces variations peuvent correspondre à des événements spécifiques, comme des pannes de ventilation, des travaux dans la station ou des épisodes de forte pollution extérieure.



# LES COEFFICIENTS SAISONNIERS

Composante saisonnière de NO2



La courbe des composantes saisonnières montre une variation périodique de la concentration en NO<sub>2</sub> au fil des mois, avec des pics et des creux réguliers. Cette périodicité peut indiquer des phénomènes récurrents influençant les niveaux de NO<sub>2</sub>, comme les conditions météorologiques (température, vent, humidité) ou des variations d'activité humaine (trafic, chauffage).

L'amplitude reste relativement stable entre 2018 et 2024. Cela suggère que la saisonnalité est un facteur constant dans la dynamique de la pollution au NO<sub>2</sub> dans cette station.

Aucun changement visible de la moyenne saisonnière au fil des années, cela peut signifier l'absence d'une diminution ou augmentation notable de la composante saisonnière elle-même.

Concernant les coefficients saisonniers, les valeurs positives correspondent aux périodes où la pollution au NO<sub>2</sub> est supérieure à la moyenne annuelle, tandis que les valeurs négatives correspondent aux périodes où la pollution est inférieure à la moyenne.

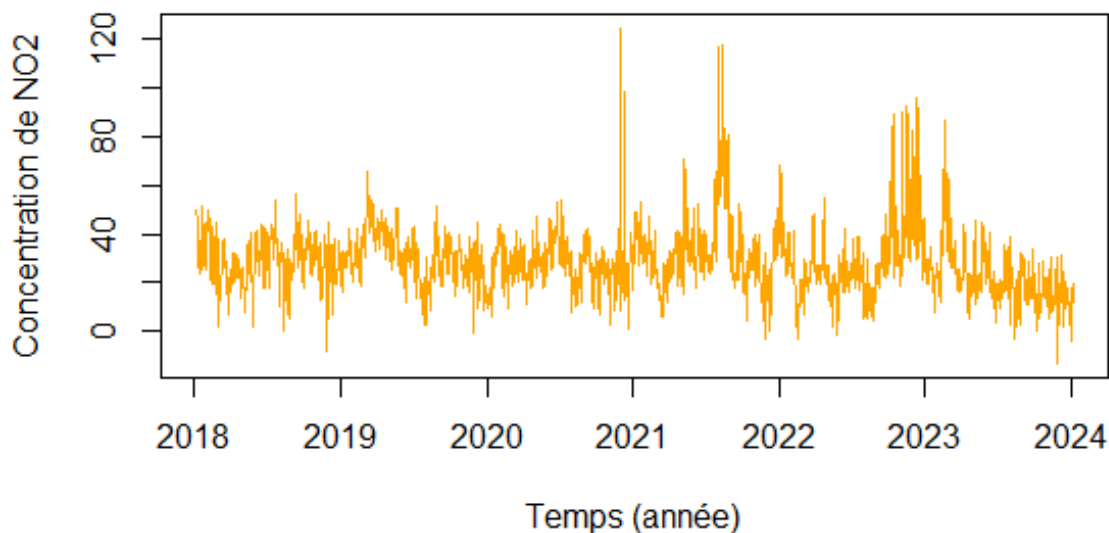
On peut dire que chaque année, les concentrations de NO<sub>2</sub> fluctuent de manière prévisible, avec des pics de pollution suivis de phases moins polluées.

Frequency = 720

[1]	-4.49317175	-5.48789397	-4.08275508	-1.67719952	-3.07136619
[6]	-4.06497730	-4.65775508	-5.25025508	-6.04247730	-6.43511619
[11]	-5.22872730	-1.02275508	2.98363381	2.59030048	1.19682825
[16]	-0.39692175	-7.79219952	-13.19081064	-14.79178286	-15.39233841
[21]	-13.39303286	-12.59372730	-8.39150508	-6.78719952	-4.78386619
[26]	-2.78081064	-1.37553286	-1.76844952	-3.76025508	-3.35150508
[31]	-2.34442175	-3.53803286	-3.53122730	-3.92567175	-3.32081064
[36]	-0.71539397	2.49043936	2.09655048	-0.69719952	-4.29150508
[41]	-9.48706064	-10.48400508	-11.68178286	-12.48150508	-13.08344952
[46]	-10.48608841	-7.28817175	-5.49164397	-3.29831064	-1.90469952
[51]	-3.70553286	-3.70247730	-3.29900508	-4.49456064	-5.29122730
[56]	-6.28858841	-8.28539397	-8.28261619	-8.47997730	-6.07775508
[61]	-4.67567175	-2.27275508	1.13016159	-1.66664397	-8.46275508
[66]	-13.65997730	-14.05858841	-12.45831064	-10.05914397	-6.45983841
[71]	-5.85969952	-1.45914397	1.34210603	5.34432825	6.54724492
[76]	7.74988381	3.55168936	3.55349492	0.95557825	0.75807825
[81]	4.96071714	4.56307825	13.56405048	14.36432825	12.76530048
[86]	11.16585603	10.36641159	3.96863381	-4.62747730	-9.42372730
[91]	-6.02122730	-6.81942175	-6.01789397	-0.61706064	-1.01706064
[96]	-0.81706064	1.38363381	1.58460603	0.78516159	-0.21456064
[101]	-1.21442175	-1.01428286	1.38599492	-1.01414397	0.78502270
[106]	-0.21594952	0.18335603	3.78363381	8.98418936	9.18349492
[111]	6.98210603	1.18071714	-5.22053286	-10.22122730	-10.82136619
[116]	-11.42150508	-9.62192175	-5.62192175	-5.02094952	-1.81886619
[121]	-0.41594952	2.58793936	0.99363381	0.59974492	0.80696714
[126]	-0.38247730	0.82877270	-1.16136619	0.44752270	0.05418936
[131]	2.25960603	2.86613381	2.47321714	3.88057825	2.68849492
[136]	-2.70358841	-8.49706064	-11.09358841	-14.09289397	-13.09303286
[141]	-10.89289397	-10.29261619	-9.69219952	-7.69067175	-5.68747730

# LA SÉRIE DÉSAISONNALISÉE

Série désaisonnalisée de NO<sub>2</sub>

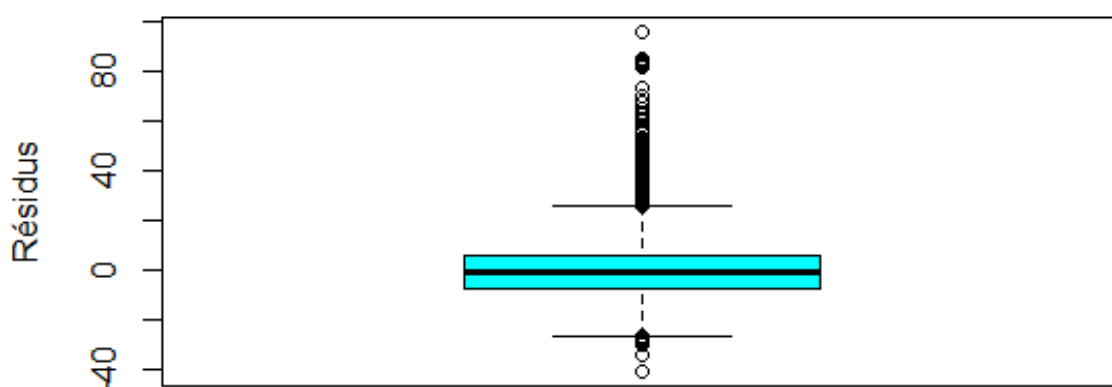


Cette courbe met en évidence les variations non saisonnières des concentrations de NO<sub>2</sub> entre 2019 et 2024. Contrairement au graphique précédent, qui montrait une forte récurrence annuelle, cette série révèle les tendances générales.

Les pics observés en 2021-2022 montrent clairement l'impact de la reprise des activités humaines après la période de confinement, ils mettent en lumière la corrélation entre trafic et pollution au NO<sub>2</sub>.

On constate également que la tendance est à la baisse entre 2023-2024 cela pourrait refléter des améliorations durables de la qualité de l'air, signalant une efficacité croissante des mesures adoptées contre la pollution.

## Boxplot des résidus de la série NO2

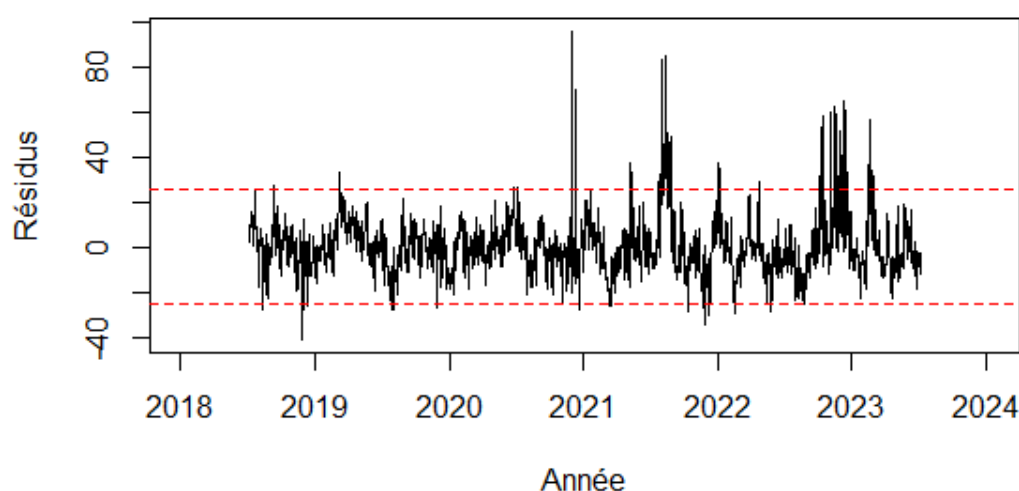


```
> summary(NO2_resid)
      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
-41.0581  -7.3634  -1.2372   0.2098   5.7716  96.0846
```

La majorité des résidus se situe entre -7,63 et 5,77, ce qui montre que la concentration réelle de NO2 reste globalement proche de la tendance désaisonnalisée, avec des écarts relativement faibles. Cela suggère que les variations de pollution dans la station Châtelet sont généralement stables et proches des attentes. Les résidus positifs (jusqu'à 96,08) indiquent des journées où la concentration de NO2 est beaucoup plus élevée que la tendance attendue. Ces journées représentent des épisodes critiques de pollution, potentiellement dus à des facteurs comme une ventilation insuffisante ou une forte affluence de passagers.

Pour les résidus négatifs (jusqu'à -41,05), ils signalent des journées où la concentration de NO2 est significativement plus basse que la tendance. Ces périodes pourraient correspondre à des moments où l'activité dans la station était réduite (par exemple, pendant des jours fériés, des grèves ou des périodes de faible affluence).

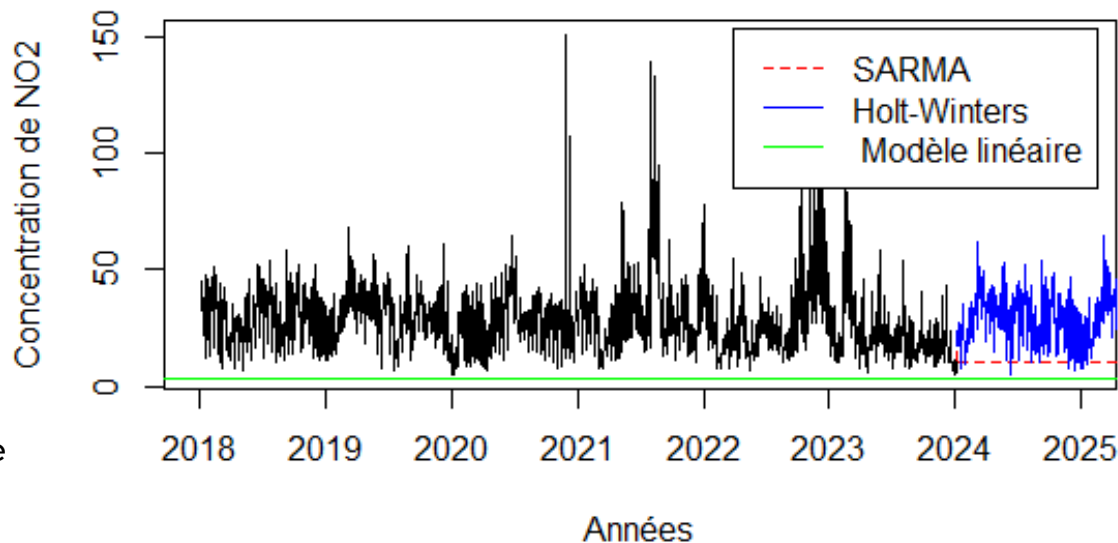
## Graphique des résidus



Le graphique des résidus met en évidence des fluctuations autour de zéro, il confirme les observations du boxplot. Avec des pics positifs atteignant plus de 80 en 2021, ce qui indique des épisodes de pollution anormalement élevés à la station Châtelet, cohérents avec les anomalies observées dans la distribution des résidus.

## Prévision des concentrations de NO2

Le modèle **SARMA** projette une prévision presque plate pour les concentrations de NO2, indiquant qu'il anticipe une stabilité sans variations significatives dans les mois à venir. Cela suggère que, selon ce modèle, les niveaux de NO2 ne fluctueront pas.



Le modèle **Holt-Winters** propose une prévision plus nuancée, avec des fluctuations régulières et une capture apparente des cycles saisonniers. Cela reflète mieux la réalité des concentrations de NO2 dans la station, où les variations sont influencées par des éléments comme la densité de passagers, la ventilation naturelle ou mécanique, et les apports extérieurs de pollution. Ce modèle montre que la pollution pourrait augmenter ou diminuer de manière cyclique, en phase avec ces facteurs.

Enfin, le **modèle linéaire** donne une prévision constante et basse, suggérant une hypothèse de stabilité des concentrations de NO2. Bien que simple à interpréter, ce modèle ne tient pas compte des variations importantes observées dans les données.

```
> mse_l_no2 <- mean((test_data - prevision_l_no2)^2)
> cat("Erreur quadratique moyenne (MSE) du modèle linéaire: ", mse_l_no2, "\n")
Erreur quadratique moyenne (MSE) du modèle linéaire: 100.8307
> mse_sarma <- mean((test_data - prediction_auto_sarma)^2)
> cat("Erreur quadratique moyenne (MSE) du modèle SARIMA: ", mse_sarma, "\n")
Erreur quadratique moyenne (MSE) du modèle SARIMA: 8.602417
> mse_hw <- mean((test_data - prediction_hw)^2)
> cat("Erreur quadratique moyenne (MSE) du modèle Holt-winters: ", mse_hw, "\n")
Erreur quadratique moyenne (MSE) du modèle Holt-winters: 8.688002
```

L'erreur quadratique moyenne (MSE) mesure la moyenne des écarts quadratiques entre les valeurs prédites et les valeurs réelles. Plus le MSE est faible, meilleure est la qualité des prédictions du modèle, car cela indique que les prédictions sont plus proches des observations réelles. Le modèle linéaire (MSE = 100.83) a la plus grande erreur parmi les trois. Cela suggère que le modèle linéaire a des limites importantes pour capturer les tendances et variations des concentrations de NO2.

Les résultats montrent clairement que les modèles SARIMA et Holt-Winters sont beaucoup plus performants que le modèle linéaire pour prédire les concentrations de NO2. Ces modèles offrent donc des prévisions plus précises et fiables sur la période de test, avec des erreurs de prédiction beaucoup plus faibles que le modèle linéaire.

# CONCLUSION

Notre étude sur la qualité de l'air à la station de Châtelet, centrée sur les concentrations de NO<sub>2</sub> a révélé une tendance générale à la baisse des niveaux de NO<sub>2</sub> depuis 2018, probablement en lien avec des améliorations dans la gestion de la ventilation et une réduction des sources de pollution internes à la station. Une saisonnalité stable a également été observée, avec des pics de concentration en hiver, liés à une moindre dispersion des polluants dans l'air confiné, et des baisses en été, lorsque les conditions favorisent une meilleure circulation de l'air. Toutefois, des variations ponctuelles, ce qui reflète des événements locaux comme des travaux ou des changements temporaires dans les conditions environnementales, persistent.

En conclusion, bien que des progrès notables aient été réalisés, ces résultats soulignent la nécessité de poursuivre les efforts pour améliorer la qualité de l'air, notamment par l'optimisation de la ventilation dans les espaces souterrains et par la réduction continue des émissions polluantes, afin de garantir un environnement plus sain pour les usagers et les travailleurs.

The report analyzes air quality in the Châtelet metro station in Paris between 2018 and 2024, with a focus on NO<sub>2</sub> (nitrogen dioxide) as a key indicator of pollution. Data from over 100,000 observations was used to assess pollution trends, seasonal variations, and the impacts of specific events. The study revealed a general decrease in NO<sub>2</sub> concentrations over the years, with notable improvements beginning in 2020, likely driven by reductions in human activity during the COVID-19 lockdowns. Annual averages of NO<sub>2</sub> dropped from over 35-40 µg/m<sup>3</sup> in 2018-2019 to around 15-20 µg/m<sup>3</sup> in 2023-2024, demonstrating significant progress in air quality improvement.

Seasonal variations were clearly identified, with pollution levels peaking in winter due to increased heating and lower ventilation efficiency in confined spaces, and dropping in summer, when better ventilation and reduced urban activity improve air quality. Weekly cycles showed higher NO<sub>2</sub> concentrations on weekdays, driven by increased passenger traffic and train operations, while weekends displayed lower levels due to reduced activity. Despite these overall improvements, occasional pollution spikes were observed, attributed to specific events such as ventilation system failures, increased passenger loads, or external environmental factors.

The report highlights the influence of seasonal, weekly, and human activity patterns on air pollution in the station. Advanced forecasting models, including SARMA, Holt-Winters, and linear models, were employed to predict future trends, confirming that air quality is expected to continue improving with proper interventions. However, persistent winter peaks and occasional anomalies emphasize the need for further measures to address pollution challenges.

The conclusions underline the progress made in reducing pollution levels in the station but also stress the importance of sustained efforts to optimize ventilation systems, control internal pollution sources, and ensure a healthier environment for both passengers and metro workers. These recommendations aim to further enhance air quality and promote long-term environmental sustainability in the station.



# TÂCHES RÉALISÉES

Après avoir importé les données, j'ai d'abord remplacé les valeurs manquantes par les valeurs de la série lissée. Ensuite, j'ai calculé la série temporelle avec la fonction **ts** afin de tracer les courbes de régression des différentes moyennes, ainsi que les moyennes mobiles d'ordre 7 et 30.

```
# Remplacer les valeurs NA par la tendance + saisonnalité
data$NO2 <- na.approx(data$NO2, rule = 2)
data_remp1 <- NO2.ts
na_indices <- which(is.na(NO2.ts))
data_remp1[na_indices] <- trend[na_indices] + seasonal[na_indices]

# Moyenne mobile d'ordre 7
MM7 <- stats::filter(NO2.ts, filter = rep(1/7, 7))
plot(NO2.ts, ylab = "NO2", xlab="Temps (année)", xlim= c(2018,2024), main = "Concentration de NO2 - Station de",
lines(MM7, col="red"))
```

→ 

```
> sum(is.na(data$NO2))
[1] 0
```

Ensuite, j'ai calculé les coefficients saisonniers ainsi que la décomposition de la série, puis je les ai tracés grâce à la fonction **plot**.

```
# les coefficients saisonniers
coeff_saisonniers <- decomposition_NO2$time.series[, "seasonal"]
coeff_saisonniers

plot(coeff_saisonniers, type = "l", main = "Composante saisonnière",
      ylab = "Saisonnalité", xlab = "Temps (mois)", col = "purple")

# Décomposition de la série temporelle
decomposition_NO2 <- decompose(NO2.ts)
plot(decomposition_NO2)
```

J'ai tracé la courbe de la série désaisonnalisée, ainsi que le boxplot et le graphique des résidus.

```
# La série désaisonnalisée
NO2_désaisonnalisée <- NO2.ts - decomposition_NO2$seasonal

plot(NO2_désaisonnalisée, type = "l", main = "Série désaisonnalisée de NO2",
      ylab = "Concentration de NO2", xlab = "Temps (année)", col = "orange")

# Les résidus
NO2_resid <- decomposition_NO2$random
summary(NO2_resid)

boxplot(NO2_resid, main = "Boxplot des résidus de la série NO2",
        ylab = "Résidus", col = "cyan")

# Graphique des résidus
plot(decomposition_NO2$random, main="Graphique des résidus", xlab="Date", ylab="Résidus")
sd_residus <- sd(decomposition_NO2$random, na.rm = TRUE)
abline(h = 2 * sd_residus, col = "red", lty = 2)
abline(h = -2 * sd_residus, col = "red", lty = 2)
```

Pour réaliser les prévisions, j'ai utilisé trois modèles : SARMA, Holt-Winters et le modèle linéaire. Ensuite, j'ai évalué leur qualité en calculant l'erreur quadratique et le MSE.

```
# Modèle linéaire
t_test <- seq_along(data$NO2)
mod1_no2 <- lm(log(NO2) ~ t_test, data = data)
prevision_l_no2 <- mod1_no2$coefficients[1] + mod1_no2$coefficients[2] * t_test

# le modèle SARMA
auto_sarma <- forecast::auto.arima(NO2.ts)
prediction_auto_sarma <- predict(auto_sarma, n.ahead = 45*24)
prediction_auto_sarma$pred

HW <- Holtwinters(NO2.ts)
prediction_hw <- predict(HW, n.ahead = 45*24)
prediction_hw

t_test <- seq_along(data$NO2)
t_fut <- seq(1, 45 * 24) # jours de novembre à décembre

plot(NO2.ts, xlim = c(2018, 2025), xlab = "Années", ylab = "Concentration de NO2",
      main = "Prévision des concentrations de NO2 ")

# les prédictions des différents modèles
lines(prediction_auto_sarma$pred, col = "red", lty = 2) # Prévisions SARMA
lines(prediction_hw, col = "blue") # Prévisions Holt-winters
lines(prevision_l_no2, col = "green") # Modèle linéaire
```