

Objectif

Analyser la différence de concentration moyenne en particules fines en région Hauts-de-France entre l'année 2019 et l'année 2020.

Données

Dans ce projet, nous utilisons les données issues du fichier **PM10_NORD_2019-2020.csv**, fournies par l'organisation Atmo Hauts-de-France. Ces données contiennent des informations sur les concentrations de particules fines PM10 (Particulate Matter < 10 µm) exprimées en microgrammes par mètre cube d'air. Les valeurs correspondent aux moyennes annuelles pour les années 2019 et 2020, calculées à partir de mesures horaires.

Chaque enregistrement comprend :

- **Le nom du département** (Department) ;
- **Le code de la station d'observation** (StationCode) ;
- **La typologie du site** (Typology : urban ou suburban) ;
- **Le nom du polluant** (Pollutant) ;
- **Les moyennes annuelles des concentrations de PM10 pour 2019** (PM10_2019) et pour 2020 (PM10_2020) ;
- **Les coordonnées géographiques de la station** : longitude et latitude au format WGS84.

Le jeu de données se compose de 31 lignes et 8 colonnes. Les types de données incluent des chaînes de caractères (Department, StationCode, Typology, Pollutant) et des valeurs numériques (PM10_2019, PM10_2020, Longitude, Latitude).

Résumé statistique (Summary) (Figure 1) :

L'analyse des statistiques résumées montre que la concentration moyenne de PM10 a diminué de 19,65 µg/m³ en 2019 à 17,58 µg/m³ en 2020. La médiane a également baissé, ce qui confirme cette tendance générale à la baisse. Les valeurs minimales et maximales suivent la même dynamique, et l'étendue des données s'est réduite en 2020, indiquant une diminution de la dispersion des valeurs.

```
> summary(data[, c("PM10_2019", "PM10_2020")])
  PM10_2019      PM10_2020
Min.   :16.00   Min.   :14.00
1st Qu.:17.65   1st Qu.:15.65
Median :19.40   Median :17.60
Mean   :19.65   Mean   :17.58
3rd Qu.:21.70   3rd Qu.:19.25
Max.   :24.10   Max.   :21.50
```

Figure 1 : Résumé statistique des concentrations de PM10 pour 2019 et 2020

Histogrammes des distributions (Figure 2) :

Les histogrammes révèlent que la distribution des concentrations de PM10 en 2019 est

décalée vers la droite, indiquant des niveaux globalement plus élevés. En 2020, la distribution est davantage centrée sur des valeurs plus faibles, reflétant une baisse des concentrations. La fréquence des mesures situées entre 16 et 20 $\mu\text{g}/\text{m}^3$ a augmenté en 2020, ce qui souligne une amélioration de la qualité de l'air.

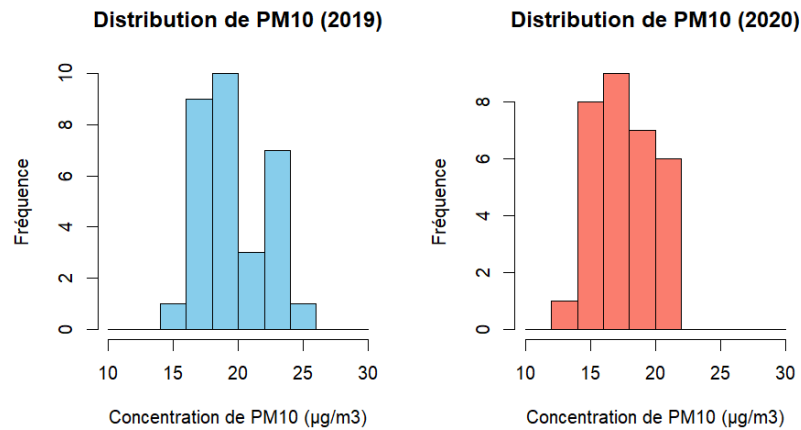


Figure 2 : Histogramme des concentrations de PM10 pour 2019 et 2020

Boxplots (Figure 3) :

Les boxplots illustrent clairement une réduction des valeurs médianes de PM10 entre 2019 et 2020. L'intervalle interquartile (IQR) est également plus étroit en 2020, indiquant une variabilité moindre des concentrations. Les limites supérieures des valeurs extrêmes ont baissé, soulignant une réduction des niveaux de pollution les plus élevés. Ces observations confirment une amélioration significative de la qualité de l'air dans la région analysée.

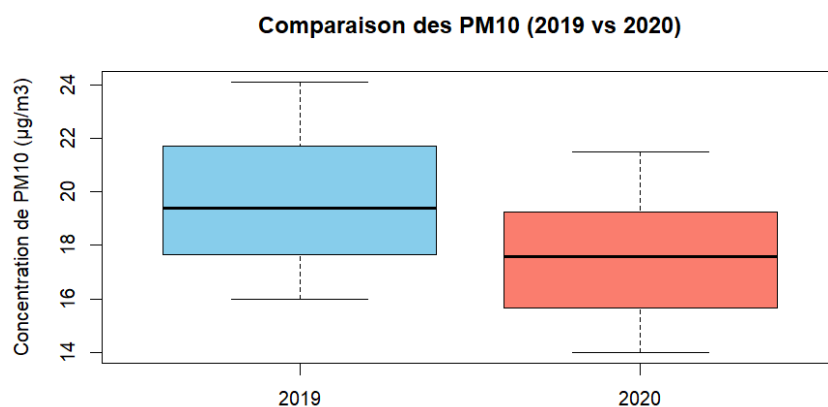


Figure 3 : Comparaison des concentrations de PM10 entre 2019 et 2020 (Boxplots)

t-test (Figure 4) :

Le t-test est utilisé pour vérifier si les moyennes de deux échantillons liés diffèrent de manière statistiquement significative. Dans ce cas, le t-test montre une différence

significative entre les moyennes des concentrations de PM10 en 2019 et 2020 (p-value = 8.212e-12). Cela confirme que la baisse des concentrations moyennes de PM10 en 2020 n'est pas due au hasard.

```
Paired t-test

data: data$PM10_2019 and data$PM10_2020
t = 10.753, df = 30, p-value = 8.212e-12
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 1.675030 2.460454
sample estimates:
mean difference
 2.067742
```

Figure 4 : Résultats du t-test

F-test (Figure 5) :

Le F-test est utilisé pour comparer les variances (dispersion des données) de deux échantillons. Les résultats du F-test indiquent qu'il n'y a pas de différence significative entre les variances des concentrations de PM10 en 2019 et 2020 (p-value = 0.8561). Cela signifie que la variabilité des concentrations est restée similaire entre les deux années.

```
F test to compare two variances

data: data$PM10_2019 and data$PM10_2020
F = 1.069, num df = 30, denom df = 30, p-value = 0.8561
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5154601 2.2171188
sample estimates:
ratio of variances
 1.069035
```

Figure 5 : Résultats du F-test

Corrélation linéaire entre les années (Figure 6) :

Le test de corrélation de Pearson a montré une forte corrélation positive entre les concentrations de PM10 en 2019 et en 2020, avec un coefficient de corrélation (r) de 0.8905. Cette corrélation est hautement significative (p-value = 1.981e-11), ce qui indique que les concentrations de PM10 en 2020 sont fortement liées à celles de 2019. Cela suggère une grande stabilité dans la distribution spatiale des PM10 entre les deux années.

L'intervalle de confiance à 95 % pour le coefficient de corrélation est compris entre 0.7833 et 0.9462, ce qui confirme la robustesse de cette relation linéaire. Ces résultats montrent que, malgré une diminution générale des moyennes en 2020, les concentrations de PM10 dans les différentes zones géographiques ont suivi une tendance similaire à celle de 2019.

```

Pearson's product-moment correlation

data: data$PM10_2019 and data$PM10_2020
t = 10.539, df = 29, p-value = 1.981e-11
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7832919 0.9462470
sample estimates:
      cor
0.8904782

```

Figure 6 : Corrélation linéaire entre les concentrations de PM10 en 2019 et 2020

Indice de Moran (Figure 7) :

L'indice de Moran pour 2019 est de 0.1270, et pour 2020, il est de 0.1207. Ces valeurs positives indiquent une faible autocorrélation spatiale positive. Les p-values respectives sont 0.0386 (2019) et 0.0471 (2020), ce qui confirme que l'autocorrélation spatiale est statistiquement significative au niveau de confiance de 95 %. Cela signifie que les concentrations de PM10 dans les points géographiquement proches ont tendance à être similaires.

<code>> print(moran_2019)</code>	<code>> print(moran_2020)</code>
\$observed	\$observed
[1] 0.1270478	[1] 0.1207644
\$expected	\$expected
[1] -0.03333333	[1] -0.03333333
\$sd	\$sd
[1] 0.07753187	[1] 0.07760444
\$p.value	\$p.value
[1] 0.03858526	[1] 0.04706868

Figure 7 : Résultats de l'indice de Moran

Test de Mantel (Figure 8) :

Le test de Mantel montre une corrélation modérée entre les distances géographiques et les concentrations de PM10 pour les années 2019 ($r = 0.2249$, $p\text{-value} = 0.003$) et 2020 ($r = 0.2261$, $p\text{-value} = 0.004$). Ces résultats suggèrent que la répartition spatiale des PM10 est significativement influencée par les distances géographiques entre les points. Ces résultats confirment que la structure spatiale des concentrations de PM10 est influencée par des facteurs locaux tels que les sources de pollution, entraînant une similarité des concentrations dans les zones proches géographiquement.

<code>> print(mantel_2019)</code>	<code>> print(mantel_2020)</code>
Mantel statistic based on Pearson's product-moment correlation	Mantel statistic based on Pearson's product-moment correlation
Call: mantel(xdis = DistMat, ydis = Dist_PM10_2019)	Call: mantel(xdis = DistMat, ydis = Dist_PM10_2020)
Mantel statistic r: 0.2249 Significance: 0.003	Mantel statistic r: 0.2261 Significance: 0.004
Upper quantiles of permutations (null model): 90% 95% 97.5% 99% 0.0851 0.1154 0.1379 0.1657 Permutation: free Number of permutations: 999	Upper quantiles of permutations (null model): 90% 95% 97.5% 99% 0.0828 0.1126 0.1415 0.1794 Permutation: free Number of permutations: 999

Figure 8 : Résultats du test de Mantel

indice de Geary (Figure 9) :

L'image présente deux graphiques de l'indice de Geary pour les années 2019 et 2020, montrant comment la statistique de Geary varie en fonction des classes de distance. En 2019, l'indice de Geary montre une augmentation marquée sur les grandes distances, ce qui indique une forte hétérogénéité spatiale de la pollution par les PM10, avec des concentrations élevées dans certaines régions. En 2020, l'augmentation de l'indice est plus progressive, et la statistique atteint des valeurs plus faibles à grande distance, suggérant une répartition plus uniforme de la pollution.

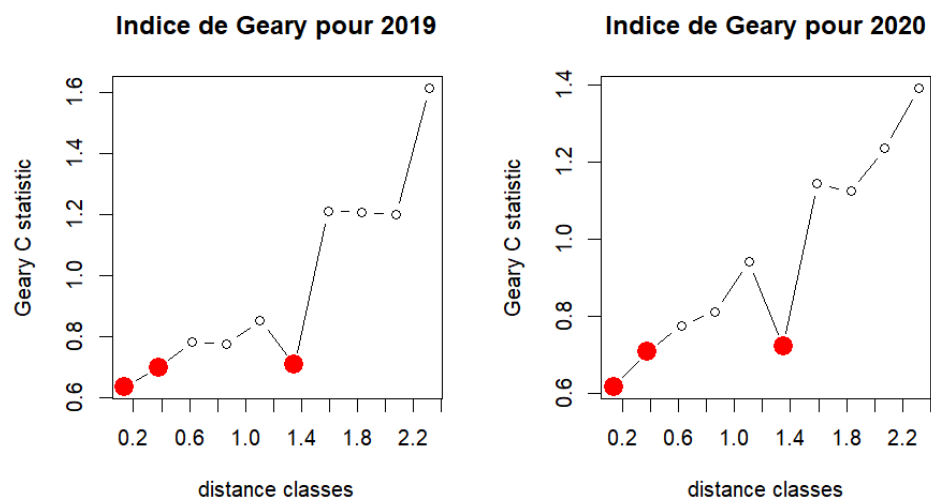


Figure 9 : Évolution de l'indice de Geary pour les concentrations de PM10 en 2019 et 2020

test de Shapiro-Wilk (Figure 10) :

Les résultats du test de Shapiro-Wilk pour les résidus des années 2019 et 2020 montrent que la distribution des résidus est proche de la normale. Pour les deux années, les valeurs p dépassent largement 0.05 (pour 2019, p-value = 0.776, pour 2020, p-value = 0.9279), ce qui signifie qu'il n'y a pas suffisamment de preuves pour rejeter l'hypothèse de normalité des résidus. Cela témoigne de la bonne qualité du modèle de Krigeage, car les résidus (erreurs) de l'ajustement suivent une distribution normale, ce qui est une condition importante pour la validité des conclusions statistiques.

```

> print(shapiro_test_2019)

Shapiro-Wilk normality test

data: residuals_2019
W = 0.97872, p-value = 0.776

> print(shapiro_test_2020)

Shapiro-Wilk normality test

data: residuals_2020
W = 0.98483, p-value = 0.9279

```

Figure 10 : Résultats du Shapiro-Wilk

Analyse des résidus

Les histogrammes des résidus pour les années 2019 et 2020 montrent que leur distribution est proche d'une distribution normale, avec des valeurs concentrées autour de zéro. Cela indique que les modèles utilisés pour l'interpolation des concentrations de PM10 sont fiables et ne présentent pas de biais systématique. Ces résultats confirment la validité des modèles pour l'analyse spatiale de la pollution.

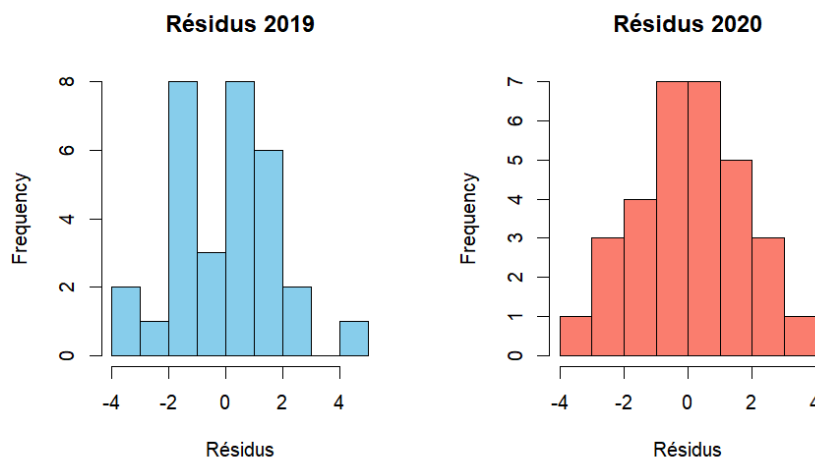


Figure 11 : Distribution des résidus pour 2019 et 2020

L'interpolation Kriging (Figure 12) :

Les cartes d'interpolation des PM10 pour les années 2019 et 2020 montrent une structure spatiale cohérente. En 2019, les concentrations de PM10 varient entre **17 et 21 $\mu\text{g}/\text{m}^3$** , avec des niveaux plus élevés situés dans la partie nord-ouest de la région. En revanche, en 2020, les concentrations sont légèrement plus faibles, oscillant entre **15 et 19 $\mu\text{g}/\text{m}^3$** , avec une baisse globale observée sur toute la région. La comparaison des deux cartes met en évidence une réduction globale des concentrations de PM10 en 2020, ce qui pourrait être

attribué à des changements dans les activités anthropiques, tels que les mesures de confinement liées à la pandémie, ou à d'autres facteurs environnementaux.

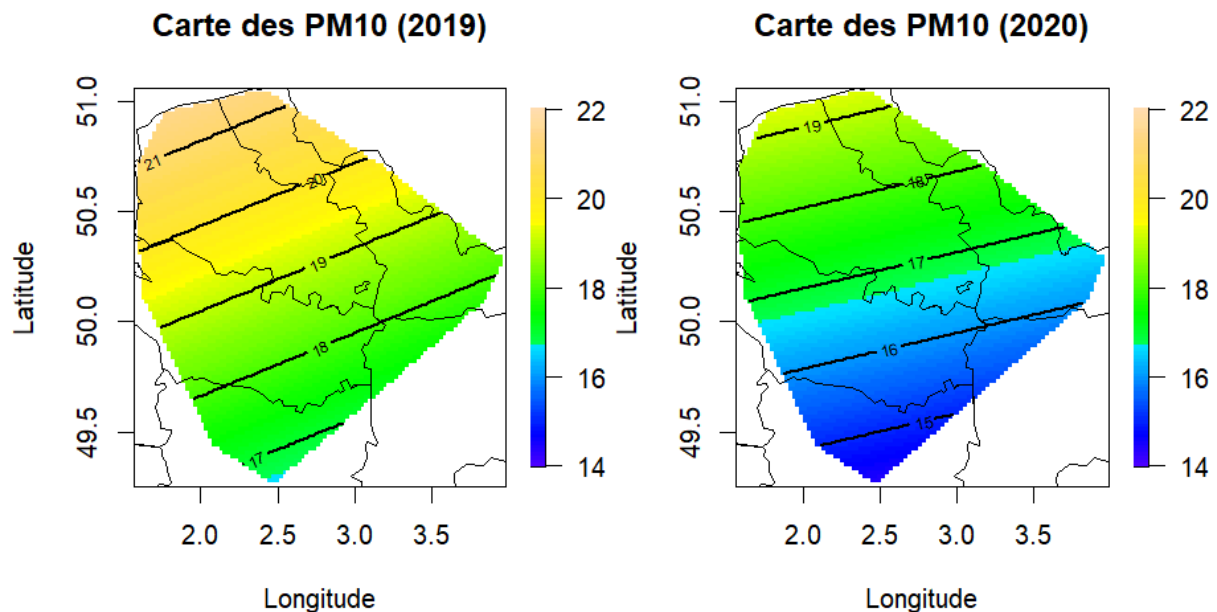


Figure 12 : Cartes finales d'interpolation des concentrations de PM10 en 2019 et 2020

Les résultats de l'analyse montrent une corrélation modérée et positive entre les valeurs observées et prédites de PM10, avec un coefficient de **0.649** pour 2019 et **0.662** pour 2020. Les p-values, respectivement **7.709e-05** et **4.905e-05**, indiquent que ces corrélations sont statistiquement significatives. Cela valide l'efficacité des modèles de krigeage pour décrire la distribution spatiale de PM10. Par ailleurs, le test de Shapiro-Wilk appliqué aux résidus des modèles montre que leur distribution suit une loi normale, avec des p-values de **0.776** pour 2019 et **0.9279** pour 2020. Ces résultats confirment que les modèles décrivent correctement les données sans biais systématique.

Conclusion

L'analyse réalisée a montré une réduction significative des concentrations annuelles moyennes de PM10 en 2020 par rapport à 2019, comme le confirment les résultats des tests statistiques. L'indice de Moran et le test de Mantel ont révélé des dépendances spatiales significatives, indiquant une stabilité dans la structure spatiale de la répartition des PM10 entre les deux années. Les modèles d'interpolation par krigeage ont démontré une grande précision, validée par le test de Shapiro-Wilk et l'analyse des résidus. Ces résultats suggèrent que l'amélioration de la qualité de l'air dans la région.