

Impact de la Pollution de l’Air sur la Mortalité : État des Lieux et Approches

Hana Sebia*, Tarik Boumaza**
Marie Le Guilly***, Mohand-Saïd Hacid****, Delphine Maucourt-Boulch[‡]

*hana.sebia@etu.univ-lyon1.fr,
Université Lyon 1

**tarik.boumaza@etu.univ-lyon1.fr,
Université Lyon 1

***marie.le-guilly@univ-lyon1.fr,
LIRIS UMR 5205 CNRS, Université Lyon 1

****mohand-said.hacid@univ-lyon1.fr,

LIRIS UMR 5205 CNRS, Université Lyon 1
[‡]delphine.boulch@chu-lyon.fr

CNRS, UMR5558, Laboratoire de Biométrie et Biologie Évolutive
France

Résumé. L’OMS estime que la pollution de l’air contribue à 7 millions de décès prématurés par an. Notre recherche a pour objectif l’évaluation de ses effets sur la mortalité dans le département du Rhône (France). L’ensemble des données sur la mortalité disponibles entre 2007 et 2019 ont été analysées. En mesurant l’exposition à la pollution de la population, en nous intéressant aux polluants considérés comme les plus dangereux pour l’Homme d’après l’OMS (NO , NO_2 , O_3 , $PM_{2.5}$ et PM_{10}), et en utilisant des méthodes d’analyse linéaires et non linéaires, on démontre une corrélation significative. La liaison entre les deux phénomènes suit une tendance linéaire positive (particulièrement pour les PM et NO_2). On aboutit à un $\rho^2 = 0.44$ pour le NO_2 , signifiant que 44% de la mortalité est liée à ce polluant. Nous avons également montré qu’une augmentation de 1% de ce polluant induisait une hausse de mortalité de 0.87%.

1 Introduction

D’après l’Organisation Mondiale de la Santé, la pollution de l’air est aujourd’hui le principal risque environnemental pour la santé dans le monde. L’OMS estime qu’en cumulant pollution intérieure et pollution extérieure, plus de 7 millions de décès survenus en 2012 sont liés à la pollution de l’air (dont 2.6 millions pour la pollution extérieure)OMS (2014). Par ailleurs, les flux de patients sont en constante progression dans les hôpitaux. De plus, la variation de ceux-ci peut induire une surcharge, du fait des contraintes matérielles (en particulier des lits de réanimations) et humaines.

Nous nous sommes alors fixés pour objectif l’étude des deux phénomènes. On se pose ainsi les questions suivantes : existe-t-il une corrélation entre la pollution extérieure et la mortalité dans le département du Rhône ? Peut-on prédire l’évolution de la mortalité en fonction de la pollution de l’air observée ?

Pour cela, nous avons d'abord effectué un travail sur les données de pollution ainsi que sur les chiffres de la mortalité. Ensuite, nous avons étudié les méthodes les plus pertinentes qui nous permettraient de démontrer un lien entre les deux phénomènes étudiés.

2 État de l'art

À ce jour, de nombreuses études ont démontré les effets néfastes de la pollution sur la mortalité.

Un programme nommé Erpurs (Campagna et al., 2003) a été mis en place en Île-de-France (France) suite à l'épisode de pollution de janvier-février 1989 afin d'évaluer la corrélation entre les variations temporelles des niveaux de pollution et le nombre quotidien de décès disponible de 1987 à 1998. Les polluants étudiés sont le NO_2 , les FN (fumées noires), les PM_{13} et le SO_2 . L'indicateur de mortalité considère les causes principales et immédiates du décès. L'humidité et la température, les niveaux de pollens ainsi que les épidémies de grippe ont été pris en compte dans les méthodes d'analyse utilisées - GAM (Hastie et Tibshirani, 1987) et GLM (Nelder et Wedderburn, 1972) -. Parmi les résultats montrés par l'étude, on peut observer une augmentation pouvant aller jusqu'à 4.7% pour la mortalité pour causes respiratoires, en rapport avec les particules fines lorsque la pollution passe d'un niveau faible à médian.

D'autres recherches évaluant les effets des polluants atmosphériques ont montré que ceux-ci étaient d'importants facteurs contribuant à l'augmentation des maladies respiratoires et à la mortalité prématurée (Hamra et al., 2014). Les effets néfastes sur la santé comprennent également une augmentation des hospitalisations pour maladies pulmonaires chroniques obstructives (Moolgavkar, 2000) ainsi qu'un risque accru de maladies cardio-vasculaires et de cancer du poumon (Valavanidis et al., 2014).

3 Méthode

3.1 Données de pollution

3.1.1 Recueil des mesures pour chaque polluant

Les données de pollution exploitées sont en libre accès et disponibles sur l'API ATMO Auvergne-Rhône Alpes rec. Les polluants suivants, jugés comme les plus dangereux pour la santé Valavanidis et al. (2016), ont été sélectionnés : CO : monoxyde de carbone ; NO : monoxyde d'azote ; NO_2 : dioxyde d'azote ; O_3 : trioxygène (ozone) ; SO_2 : dioxyde de soufre ; $PM_{2.5}$: particules fines de diamètre inférieur à $2,5 \mu m$; PM_{10} : particules fines de diamètre inférieur à $10 \mu m$.

Pour chaque polluant, nous avons recueilli les mesures effectuées par toutes les stations situées dans le département du Rhône de 2007 à 2019 (données mensuelles) et de 2018 à 2019 (données quotidiennes). Les mesures des CO et SO_2 étant trop incohérentes (valeurs négatives expliquées par l'appareillage ou les mauvaises conditions météorologiques d'après l'ATMO), ou incomplètes, nous avons retenu les polluants suivants : NO , NO_2 , O_3 , $PM_{2.5}$, PM_{10} .

3.1.2 Données démographiques

Nous avons ensuite recueilli les données démographiques des communes où se trouvent les stations. Elles se basent sur les populations légales des communes en vigueur à compter du 1er janvier 2021 (date de référence statistique : 1er janvier 2018) INSEE (2020b). En prenant comme hypothèse que la population de chaque commune est stable relativement à la population totale

de la métropole, on peut construire un indicateur de pollution pour le département du Rhône pour n stations :

$$\text{indicateur} = \frac{\sum_{i=1}^n \text{mesure}_{\text{station}_i} * \text{population_commune}_{\text{station}_i}}{\text{population_totale}}$$

Cet indicateur permet de représenter l'exposition à la pollution de la population. En effet, chacune des stations mesure la pollution extérieure mais ne prend pas en compte l'absorption de celle-ci. Cet indicateur, bien qu'imparfait car il considère que la population d'une commune est exposée à la même quantité de pollution à un instant donné, fournit une vision globale de l'exposition. L'évolution de cet indicateur est représenté sur la figure 1.

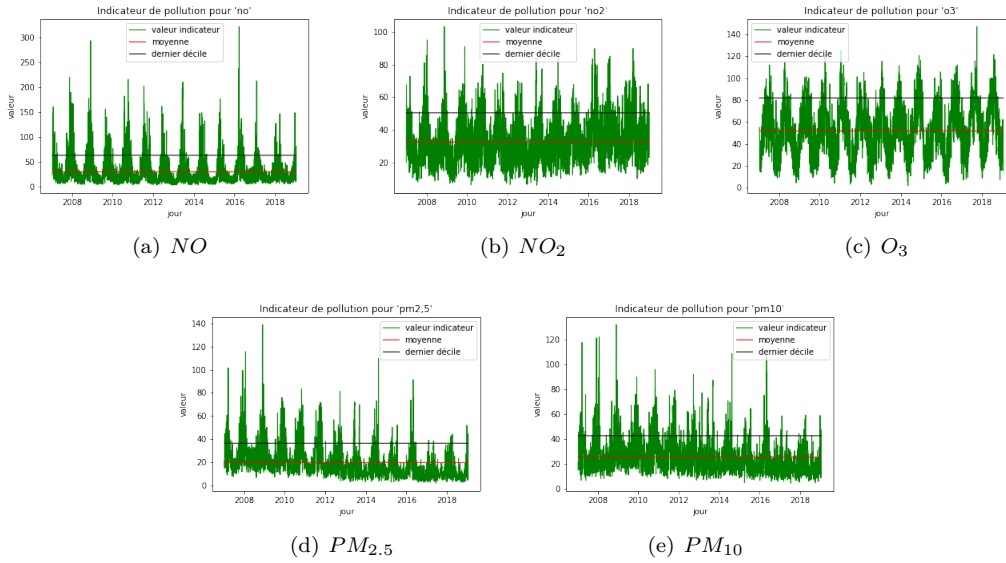


FIG. 1 – Évolution de la valeur de l'indicateur entre 2007 et 2019

3.2 Données de mortalité

Pour les chiffres de la mortalité, nous avons étudié les bases de données de l'INSEE pour les années de 2007 à 2019 (incluses) INSEE (2020a). Nous avons alors collecté les chiffres des décès quotidiens pour les années 2018 et 2019 (seuls chiffres quotidiens disponibles), ainsi que les décès mensuels de 2007 à 2019 (pour l'étude d'une période plus longue). Ces données sont représentées sur la figure 2. Ensuite, nous n'avons retenu que les décès des personnes ayant pour lieu de résidence une commune du département du Rhône (code INSEE 69), puisque c'est le département où l'on mesure la pollution extérieure.

3.3 Étude des corrélations

Afin de quantifier la relation entre nos deux variables (*i.e.*, la mortalité et la pollution) de manière à mettre en évidence le sens de la liaison et son intensité, nous explorons deux pistes : la linéarité et la non linéarité de la relation.

Impact de la pollution de l'air sur la mortalité

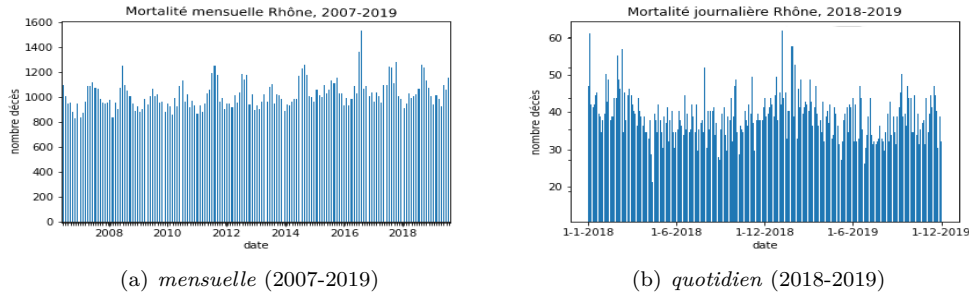


FIG. 2 – Chiffres de la mortalité

Le coefficient de corrélation de Pearson constitue une mesure de l'intensité de liaison linéaire et monotone entre deux variables. Il représente une normalisation de la covariance mesurant la tendance de deux variables à être simultanément au dessus ou en dessous de leurs espérances respectives. Or, la non-normalité des données de pollution et de mortalité ne nous permet pas de tirer de résultats fiables avec ce test paramétrique. En effet, les deux tests probabilistes (*Shapiro-Wilk* et *Anderson Darling*) effectués sur les données indiquent que les distributions ne suivent pas une loi normale (respectivement $p\text{-value} < 10^{-6} < 0.05$ et $p\text{-value} < 10^{-6} < 0.05$ pour la mortalité mensuelle ; $p\text{-value} < 10^{-4} < 0.05$ et $p\text{-value} < 10^{-3} < 0.05$ pour la mortalité quotidienne). Nous nous tournons donc vers une variante non paramétrique de ce dernier : le coefficient de corrélation de Spearman.

3.3.1 Corrélation de Spearman

Le coefficient de Spearman est calculé à partir des substitutions des valeurs observées des variables originelles par leurs rangs :

$$\rho = \frac{\text{cov}(R_X, R_Y)}{\sigma_{R_X} * \sigma_{R_Y}}$$

où $\text{cov}(R_X, R_Y)$ représente la covariance des variables de rang et σ_{R_X} et σ_{R_Y} représentent les écarts-types.

La mesure est normalisée et varie entre -1 (forte liaison négative) et 1 (forte liaison positive), 0 indiquant une absence de liaison entre les deux variables.

Ce coefficient est calculé pour les données mensuelles et quotidiennes de pollution et de mortalité.

On cherche également d'autres corrélations possibles, notamment par la sélection des données. En effet, l'impact de l'évolution du niveau de pollution n'est peut-être pas le même en considérant un niveau élevé relativement à la moyenne de celui-ci. On étudie alors deux méthodes. Dans un premier temps, on calcule la moyenne glissante de la mortalité quotidienne et des mesures de pollution sur N jours. Dans un second temps, on étudie la sélection par seuil : on ne sélectionne que les jours dont la mesure de pollution est supérieure à un seuil donné.

3.3.2 Modèles Additifs Généralisés

Une modélisation de la relation entre les indicateurs de mortalité et de pollution a également été réalisée via le développement de modèles additifs généralisés (GAM) Hastie et Tibshirani (1987). Contrairement à la méthode de corrélation de Spearman (qui mesure la monotonie de la

liaison), ces modèles permettent une plus grande flexibilité quant à la relation non linéaire que peuvent avoir certaines covariables explicatives avec une variable dépendante. Pour n variables, un modèle additif généralisé a la forme :

$$g[E(Y)] = \sum_{i=1}^n S_i(X_i)$$

où E désigne l'espérance, g est une fonction de lien et les S_i sont des fonctions de lissage. Cette modélisation prend en compte les indicateurs de pollution atmosphérique uniquement (pris individuellement).

Afin d'estimer un coefficient attribuable à chaque polluant étudié, nous nous appuyons sur le modèle statistique GLM : Generalized Linear model Nelder et Wedderburn (1972) :

$$h[E(Y)] = X\beta$$

où h est une fonction quelconque, $E(Y)$ est l'espérance de la variable à prédire, X est la matrice des variables explicatives et β le vecteur des paramètres à estimer.

Le risque relatif (RR) est calculé à partir de l'exponentielle des éléments de β et le pourcentage de variation du risque de mortalité anticipé (MA) est calculé à partir du risque relatif : $(RR - 1) \cdot 100$. Ainsi, un pourcentage positif indique une augmentation du risque tandis qu'un pourcentage négatif indique une diminution de celui-ci. Les deux modèles ont été implémentés avec R (package 'mgcv'¹), le code est disponible en ligne²

4 Résultats

4.1 Données mensuelles

On détermine la corrélation de Spearman entre la mortalité et chacun des polluants étudiés sur la période 2007-2019. Les résultats obtenus démontrent que celles-ci ne sont pas négligeables : $\rho > 0.52$ pour NO et $\rho > 0.66$ pour NO_2 notamment (cf. table 1).

ρ^2 s'interprète comme une proportion de variance expliquée. Ceci permet d'affirmer alors que la pollution en NO_2 est liée à 44% des décès ($\rho^2 \approx 0.44$). Notons que la certitude p-value est inférieure à 10^{-3} pour tous les résultats présentés, ce qui indique que la probabilité que la valeur nulle soit correcte est très faible (et que les résultats sont aléatoires).

	NO	NO_2	O_3	$PM_{2.5}$	PM_{10}
mortalité	0.527262	0.665121	-0.478647	0.238333	0.238671

TAB. 1 – Coefficients de corrélation de Spearman entre les polluants et la mortalité mensuelle 2007-2019

Le développement des modèles additifs généralisés pour les données mensuelles montre une liaison significative (*edf* avec *p-value* $< 10^{-3}$) entre la mortalité et l'indicateur de pollution pour le NO_2 . La courbe suit une tendance linéaire positive : *edf* ≈ 1 (cf. figure 3).

Les autres courbes indiquent des corrélations non linéaires mais monotones (*edf* > 2), à l'exception du polluant O_3 pour lequel on n'observe pas de corrélation significative. Un modèle

1. <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>.

2. <https://forge.univ-lyon1.fr/m1-grp26/impact-pollution-mortalite>

Impact de la pollution de l'air sur la mortalité

	<i>NO</i>	<i>NO₂</i>	<i>O₃</i>	<i>PM_{2.5}</i>	<i>PM₁₀</i>
mortalité	0.188265	0.197158	-0.219550	0.150273	0.106504

TAB. 2 – Coefficients de corrélation de Spearman entre les polluants et la mortalité quotidienne 2018-2019

polluant	r	p-value	seuil	N jours	L données	méthode
<i>NO</i>	0.53354	$< 10^{-5}$	0	7	730	moyenne glissante
<i>NO₂</i>	0.51777	$< 10^{-5}$	0	7	730	moyenne glissante
<i>O₃</i>	0.22844	0.039	87	5	82	sélection par seuil
<i>PM_{2.5}</i>	0.46977	$< 10^{-5}$	0	7	730	moyenne glissante
<i>PM₁₀</i>	0.44715	$2 \cdot 10^{-5}$	29	7	86	sélection par seuil

TAB. 3 – Coefficients de corrélation de Spearman entre les polluants et la mortalité quotidienne 2018-2019 et méthode associée

linéaire généralisé appliqué au polluant *NO₂*, paramétré par une loi de Poisson, indique un coefficient 0.0086629. D'où :

$$MA_{NO_2} = (RR_{NO_2} - 1) \cdot 100 = (e^{0.0086629} - 1) \cdot 100 \approx 0.87\%$$

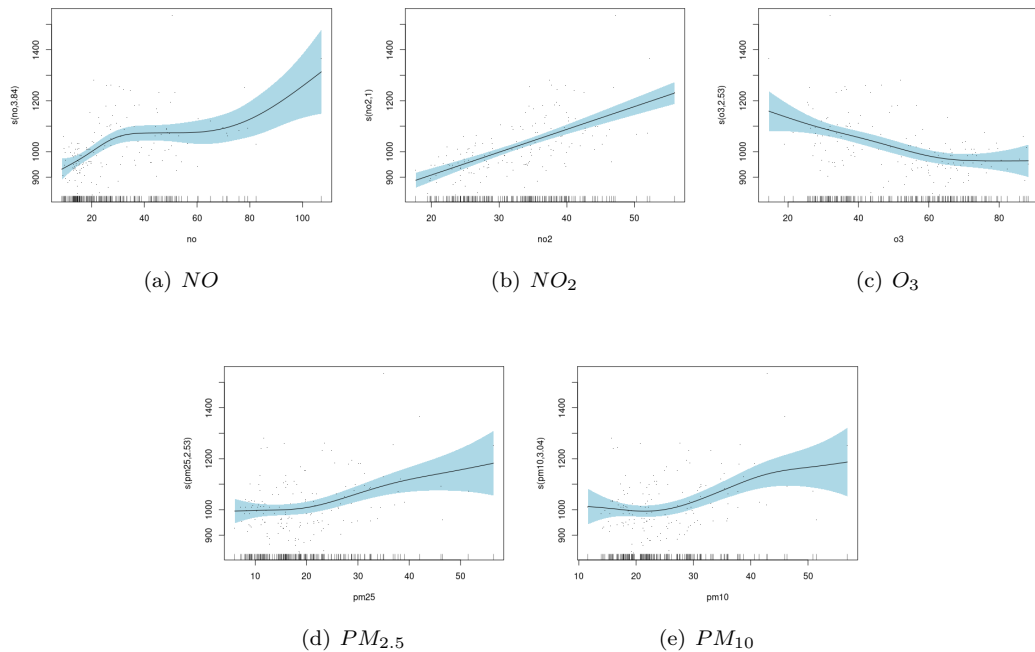


FIG. 3 – Modèle GAM pour la mortalité mensuelle et chacun des polluants étudiés

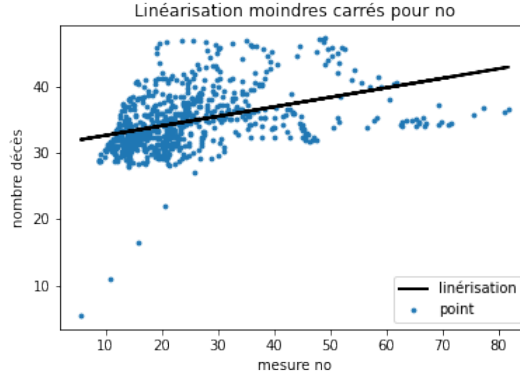


FIG. 4 – Corrélation entre NO et mortalité : droite approchant le nuage de points par la méthode des moindres carrés

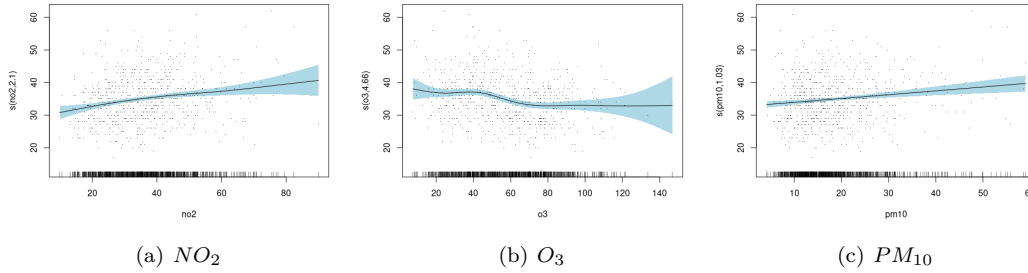


FIG. 5 – Modèle GAM pour la mortalité quotidienne et chacun des polluants étudiés

4.2 Données quotidiennes

On détermine la corrélation de Spearman entre la mortalité et chacun des polluants étudiés sur la période 2018-2019. Les résultats obtenus ne sont pas concluants : ($\rho_i^2 < 0.05$, négligeable) (cf. table 2).

En utilisant les méthodes décrites précédemment (sélection par seuil et moyenne glissante), on démontre une corrélation entre certains polluants et la mortalité quotidienne (cf. Table 3).

En effet, on obtient un $\rho^2 \in [0.2, 0.28]$ pour le NO , O_3 , $PM_{2.5}$, PM_{10} , avec des paramètres particuliers. Cette relation permet notamment, via la méthode des moindres carrés, d’approcher la droite d’équation $y = ax + b$ déterminée par le nuage de points : $a = 0.143$, $b = 31.220$ (cf. figure 4).

Le développement des modèles additifs généralisés pour les données quotidiennes montre une faible liaison linéaire positive entre la mortalité et les polluants $PM_{2.5}$ et PM_{10} ($edf_{PM_{2.5}} \approx 1.001$ et $edf_{PM_{10}} \approx 1.027$), une liaison quadratique pour les polluants NO et NO_2 ($edf_{NO} \approx 2.779$ et $edf_{NO_2} \approx 2.101$) et une absence de liaison pour le polluant O_3 (approchant des fonctions constantes) (cf. table 5).

$$MA_{PM_{2.5}} = (RR_{PM_{2.5}} - 1) \cdot 100 = (e^{0.004589} - 1) \cdot 100 \approx 0.46\%$$

$$MA_{PM_{10}} = (RR_{PM_{10}} - 1) \cdot 100 = (e^{0.0030908} - 1) \cdot 100 \approx 0.31\%$$

5 Discussion

Les résultats explicités dans la section précédente montrent une corrélation entre pollution extérieure et mortalité. On retrouve des coefficients de corrélation de Spearman significatifs pour l'analyse des données mensuelles comme quotidiennes (en particulier pour les polluants NO et NO_2).

De plus, les modèles développés montrent une hausse de la mortalité due à l'augmentation de la pollution. D'après l'analyse des données mensuelles, une augmentation d'un pourcent de la pollution en NO_2 conduit à une hausse de la mortalité de 0.87%.

Ces résultats sont cohérents avec la littérature disponible (dont l'étude décrite dans la section 2). Cependant, les corrélations sont ici moins puissantes. En effet, l'étude démontrant elle aussi le lien entre les deux phénomènes étudiés Campagna et al. (2003) formule une conclusion plus forte, puisqu'elle aboutit à des pourcentages plus importants. Plusieurs facteurs peuvent l'expliquer, tels que l'exclusion des décès accidentels et intérêt particulier porté à la mortalité pour causes respiratoires ou cardio-vasculaires et la prise en compte d'autres facteurs environnementaux ainsi que les variations à court et à long terme. Il aurait été également intéressant de pouvoir effectuer un travail équivalent pour les polluants CO et SO_2 considérés parmi les polluants les plus dangereux sur la santé d'après l'OMS. Aussi, la distinction des patients par âge, sexe, pathologie et antécédents (informations qui ne sont pas fournies par l'INSEE et soumises au secret médical) pourraient produire des résultats plus pertinents. Enfin, l'hypothèse non réaliste d'uniformité des durées d'exposition aux polluants étudiés pourrait constituer un biais important.

6 Conclusion

Les résultats obtenus démontrent une corrélation entre les niveaux de pollution atmosphérique et la mortalité dans le département du Rhône entre 2007 et 2019 pour les polluants NO , NO_2 , $PM_{2.5}$, et PM_{10} . Les méthodes d'analyse employées permettent de montrer qu'une augmentation de la mesure de certains polluants entraînait une hausse de la mortalité à court terme.

Cependant, ces conclusions ne sont pas suffisamment puissantes pour la conception d'un modèle permettant la prédiction de la mortalité. Une étude de la mortalité quotidienne sur une période plus importante, en disposant d'informations privilégiées sur les patients, de sorte à les distinguer par âge, sexe, durée d'exposition moyenne et antécédents médicaux pourraient néanmoins le permettre. Par la suite, on pourra également étudier le développement de modèles prédictifs pour tenter de prédire la mortalité et les hospitalisations en se basant sur les mesures de qualité de l'air.

Références

Api atmo auvergne-rhône-alpes.

Campagna, D., A. Lefranc, C. Nunes-Odasso, et R. Ferry (2003). Évaluation des risques de la pollution urbaine sur la santé en île-de-france (erpurs) : liens avec la mortalité 1987-1998. *VertigO-la revue électronique en sciences de l'environnement* 4(1).

Hamra, G. B., N. Guha, A. Cohen, F. Laden, O. Raaschou-Nielsen, J. M. Samet, P. Vineis, F. Forastiere, P. Saldiva, T. Yorifuji, et al. (2014). Outdoor particulate matter exposure and lung cancer : a systematic review and meta-analysis. *Environmental health perspectives*.

- Hastie, T. et R. Tibshirani (1987). Generalized additive models : some applications. *Journal of the American Statistical Association* 82(398), 371–386.
- INSEE (2020a). Base de données, mortalité.
- INSEE (2020b). Recensement de la population.
- Moolgavkar, S. H. (2000). Air pollution and hospital admissions for chronic obstructive pulmonary disease in three metropolitan areas in the united states. *Inhalation Toxicology* 12, 75–90.
- Nelder, J. A. et R. W. Wedderburn (1972). Generalized linear models. *Journal of the Royal Statistical Society : Series A (General)* 135(3), 370–384.
- OMS (2014). 7 millions de décès prématurés sont liés à la pollution de l’air chaque année.
- Valavanidis, A., V. Thomais, et F. Konstantinos (2016). Air pollution as a significant cause of diseases and premature death. *Ambient Air Pollution in Urban Areas and Indoor Air Pollution are Associated with Adverse Health Effects and Premature Mortality*.
- Valavanidis, A., T. Vlachogianni, et K. Fiotakis (2014). Airborne particulate matter in urban areas and risk for cardiopulmonary mortality and lung cancer : Dietary antioxidants and supplementation for prevention of adverse health effects. *Pharmakeftiki* 26(4), 139–156.

Summary

The WHO estimates that air pollution contributes to 7 million premature deaths per year. Our research aims to evaluate its effects on mortality in the Rhone department (France). Available mortality data between 2007 and 2019 were analyzed. By measuring the exposure to pollution of the population, focusing on the pollutants considered to be the most dangerous for humans according to the WHO (NO , NO_2 , O_3 , $PM_{2.5}$ and PM_{10}), and using linear and non-linear analysis methods, a significant correlation is demonstrated. The link between the two phenomena follows a positive linear trend (particularly for the PM and NO_2). We measured a $\rho^2 = 0.44$ for the NO_2 , meaning that 44 % of the mortality is linked to this pollutant. We have also shown that an increase of 1% of this pollutant induces a 0.87% increase of mortality.