JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE

PIERRE LOUP
MARIETTE GERBER
YOHANN FOUCHER
MONIQUE SAINTOT
MAURICIO HERNANDEZ
ISABELLE ROMIEU
JEAN-PIERRE DAURÈS

Évaluation de l'exposition individuelle au dioxyde d'azote et à l'ozone en fonction d'un mélange de variables agrégées et individuelles : modèles et exemple

Journal de la société française de statistique, tome 145, nº 3 (2004), p. 69-79

<http://www.numdam.org/item?id=JSFS_2004__145_3_69_0>

© Société française de statistique, 2004, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (http://publications-sfds.math.cnrs.fr/index.php/J-SFdS) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

ÉVALUATION DE L'EXPOSITION INDIVIDUELLE AU DIOXYDE D'AZOTE ET À L'OZONE EN FONCTION D'UN MÉLANGE DE VARIABLES AGRÉGÉES ET INDIVIDUELLES : MODÈLES ET EXEMPLE

Pierre LOUP [1], Mariette GERBER [2], Yohann FOUCHER [1], Monique SAINTOT [2], Mauricio HERNANDEZ [3], Isabelle ROMIEU [3] et Jean-Pierre DAURÈS [1]

RÉSUMÉ

Pour déterminer l'impact réel de la pollution atmosphérique sur la santé des populations, une meilleure connaissance de l'exposition individuelle est nécessaire. Parmi les axes de recherche actuels, les modèles d'exposition individuelle semblent constituer l'une des voies d'exploration prioritaires. Dans cet article, nous proposons différentes méthodes susceptibles d'être utilisées pour l'écriture d'un tel modèle, à partir d'un recueil classique de données présentées dans le texte. Nous discutons de leurs attraits, de leurs limites mais aussi de la difficulté d'interprétation des biais écologiques dans les modèles multiniveaux.

ABSTRACT

To determine real health effects of air pollution, a better knowledge of individual exposure is necessary. Among the various research topics, individual exposure modelling seems to be one of the most interesting. In this paper we propose several modelling strategies which could be applied to write such a model from a traditional data collection also presented in this text. We discuss not only their interests and shortcomings but also the difficulties of the ecologic bias interpretation in the multilevel models.

1. Institut Universitaire de Recherche Clinique, Laboratoire d'Épidémiologie et de Biostatistique, 641, avenue du Doyen Gaston Giraud 34093 Montpellier, France
p_loup@iurc.montp.inserm.fr

2. Groupe d'Epidémiologie Métabolique, Centre de Recherche en Cancérologie, INSERM-CRLC, Montpellier, France

3. Institut National de la Santé Publique, Cuernavaca, Mexique

1. Introduction

Les effets aigus de la pollution atmosphérique sont actuellement relativement bien identifiés et même estimés (Desqueyroux et Momas, 1999; Kunzli *et al.*, 2000). Il n'en est rien des effets de la pollution chronique, tels les dommages à la fonction respiratoire ou même le développement de tumeurs pulmonaires.

D'une manière générale, les enquêtes épidémiologiques d'environnement urbain se heurtent à la difficulté d'évaluer les conditions d'exposition (Smith, 1988). Les principales contraintes relèvent de l'hétérogénéité des populations, de la multiplicité des sources d'exposition et de l'évaluation de l'exposition. On sait bien que l'association entre facteurs environnementaux et santé est difficile à mettre en évidence et qu'il est nécessaire, pour renforcer cette association, d'évaluer avec précision l'exposition individuelle et de maîtriser les facteurs de confusion (Hémon, 1995).

L'existence de réseaux de surveillance dans de nombreuses villes peut procurer les premières mesures d'exposition. Cependant, celles-ci sont trop imprécises quand il s'agit d'établir un lien entre exposition et apparition de pathologies dans un groupe de sujets.

La mesure de l'exposition individuelle de certains polluants est réalisable grâce à des échantillonneurs passifs portés par le sujet. Certains sont bien validés et leur utilisation bien codifiée. C'est le cas du tube de Palmes qui évalue l'exposition au dioxyde d'azote (NO_2) (Palmes *et al.*, 1976; Atkins *et al.*, 1986). Mais la mesure par échantillonnage de l'exposition à l'ozone (O_3) est plus délicate, notamment car les concentrations d'O_3 varient largement au cours de la journée (maximum entre 12 h et 14 h, et faible la nuit). Or les échantillonneurs passifs donnent la mesure de l'exposition cumulée pendant un temps donné, aplatissant ainsi le pic diurne de l'O_3. Quelle que soit la valeur métrologique de la technique, son utilisation en épidémiologie est limitée par le nombre de participants et est en général réservée aux études de panel.

Pour aborder l'étude de l'effet éventuel de la pollution par les photo-oxydants (NO_2 et O_3), nous avons mesuré l'exposition individuelle par échantillonneurs passifs. Nous avons ainsi conduit deux études à Montpellier, et une en Île de France, qui nous ont permis d'identifier les déterminants individuels de l'exposition personnelle à NO_2 et à O_3 (Bernard *et al.*, 1997, Bernard *et al.*, 1999; Saintot *et al.*, 1999, Saintot *et coll.*, 2000).

Il nous a paru important de répéter ces études dans d'autres situations urbaines, où pourraient coexister divers polluants. En conséquence, une étude a été conduite au Mexique.

L'objectif principal de cet article, au regard du plan d'expérience de l'enquête Mexicaine et des mesures individuelles et agrégées qui seront disponibles prochainement, est de discuter des stratégies de modélisation semblant les mieux adaptées pour l'écriture d'un modèle d'exposition individuelle à la pollution atmosphérique, mélangeant des facteurs explicatifs agrégés et non agrégés. D'autre part, les mesures par échantillonnage passif pour l'O_3 étant sujettes à caution, nous proposons une analyse statistique afin d'étudier leur fiabilité.

2. Sujets et méthodes

2.1. Sujets

Quatre zones ont été sélectionnées en raison de leurs différents niveaux d'exposition : une zone caractérisée par un faible trafic automobile, présentée comme subissant une faible pollution par photo-oxydants (Cuernavaca Nord), une zone caractérisée par un fort trafic automobile présentée comme subissant un taux élevé de NO_2 et plus modéré de O_3 (Mexico ville), une zone à trafic automobile modéré présentée comme subissant un taux modéré de NO_2 et un fort taux de O_3 (Mexico Sud Est) et une zone à trafic automobile modéré présentée comme subissant un taux modéré de NO_2 et faible de O_3 (Cuernavaca ville).

Sur l'ensemble des 4 zones, 321 sujets de 14 à 75 ans ont porté des échantillonneurs à NO_2 et O_3, sur des périodes de 5 jours. Parmi les participants, 94 habitaient à Cuernavaca Nord, 98 à Mexico ville, 65 à Mexico Sud Est et 64 à Cuernavaca ville. Nous disposons de données complètes uniquement pour 309 sujets.

2.2. Métrologie

Mesure individuelle

Pour la mesure du NO_2, un système de tubes portatifs, appelé tubes Palmes (Palmes *et al.*, 1976) est utilisé. Cette méthode est maintenant éprouvée (Atkins *et al.*, 1986). Ce tube a été validé dans les conditions d'utilisation montpelliéraines (Bernard *et al.*, 1997).

La mesure de l'O_3 par échantillonneur passif (PASSAM AG., Suisse) est basée sur la décomposition par oxydation du 1,2-Di(4-Pyridyl)-éthylène (DPE) et la détermination de l'aldéhyde résultante (Bernard *et al.*, 1999). L'interférence entre le NO_2 et l'O_3 est inférieure à 10 % par cette méthode. Cependant leur fiabilité est variable selon les conditions climatiques.

Les sujets portent les échantillonneurs passifs débouchés, accrochés à leur ceinture, durant 5 jours et les posent le soir près de leur lit. Les échantillonneurs sont protégés des vêtements par une plaque de PVC pour éviter l'absorption d'O_3 par le tissu des vêtements (Bernard *et al.*, 1999).

Mesure environnementale

Les réseaux de surveillance de Mexico et de Cuernavaca nous ont communiqué les mesures réalisées dans les stations proches des résidences des participants. Cuernavaca Nord n'avait pas de station de NO_2.

En outre, nous avons placé des échantillonneurs à proximité de chaque analyseur, lors de chaque campagne de mesures, pour vérifier la capacité des échantillonneurs passifs à mesurer l'O_3 dans les conditions mexicaines.

Des informations complémentaires météorologiques (température, pluviométrie, vitesse du vent, ...) sont aussi à notre disposition.

Questionnaire

Les participants doivent répondre à un questionnaire portant sur les caractéristiques sociodémographiques. Leur budget espace-temps, leurs consommations de tabac, de fruits et légumes pendant les 5 jours de l'enquête sont enregistrés.

3. Méthodologie statistique

Préalablement à l'étape de modélisation de l'exposition individuelle, une étude de fiabilité des mesures micro-environnementales par échantillonneur est nécessaire afin de déceler d'éventuels biais de mesure. En effet, si la mesure de l'exposition au NO_2 par échantillonnage passif est une méthode éprouvée, il n'en est pas de même pour l'O_3 dont les mesures peuvent se révéler plus ou moins fiables selon le contexte (conditions climatiques, interférences avec d'autres polluants,...).

Cette étape permettra de savoir si les variables environnementales ont un effet sur la valeur réelle du polluant (effet de dispersion par le vent, par exemple) ou si elles perturbent directement sa mesure. Dans ce dernier cas, cette information devra être prise en compte lors de l'inclusion de ces variables dans les modèles prédictifs finaux de l'exposition individuelle à la pollution atmosphérique au NO_2 et à l'O_3.

3.1. Étude de la fiabilité des mesures de O_3 par échantillonnage passif

Une comparaison des mesures par échantillonneurs passifs pour l'O_3 par rapport à celles obtenues par analyseurs a été réalisée pour vérifier la capacité des échantillonneurs à bien mesurer l'O_3 dans les conditions climatiques mexicaines. Ainsi, pour chaque campagne de mesure, un échantillonneur passif pour l'O_3 a été positionné à proximité de chaque analyseur.

Pour chaque campagne de mesure et pour un analyseur donné, nous disposerons donc de 5 mesures de l'analyseur appariées sur une même mesure de l'échantillonneur (cumulée sur 5 jours et exprimée en $\mu g/m^3$).

Soit X_{ijk} la variable aléatoire représentant les mesures réalisées par l'analyseur i ($i = 1, \ldots, 4$), au cours de la campagne de mesures j ($j = 1, \ldots, m_i$) au jour k ($k = 1, \ldots, 5$) ; m_i symbolisant le nombre de campagnes réalisées sur l'analyseur i. De même, on pose Y_{ij} la variable aléatoire représentant les mesures réalisées par l'échantillonneur à proximité de l'analyseur i lors de la campagne j.

Enfin, définissons Z_{ijk} comme l'écart entre la mesure de référence de l'analyseur et la mesure réelle de l'échantillonneur, soit $Z_{ijk} = X_{ijk} - Y_{ij}$, l'étude de fiabilité consistera en l'analyse de cette nouvelle variable. Dans un premier temps, nous nous attacherons à réaliser son étude descriptive, pour ensuite nous orienter vers une modélisation plus précise.

Analyse descriptive

Tout d'abord, un histogramme de la variable Z_{ijk} nous donnera un aperçu de sa distribution et de la nécessité ou non d'une normalisation. Nous pourrons estimer le centrage de cette distribution autour de 0 (validité de la mesure) ainsi que sa dispersion (fiabilité de la mesure).

Dans un second temps, l'analyse descriptive sera approfondie, notamment par un nuage de points de Z_{ijk} en fonction de X_{ijk}. Ce dernier pourra, par exemple, retranscrire un phénomène de saturation de l'échantillonneur ou, au contraire, mettre en évidence la difficulté des échantillonneurs à mesurer des concentrations de faibles niveaux (Figure 1).

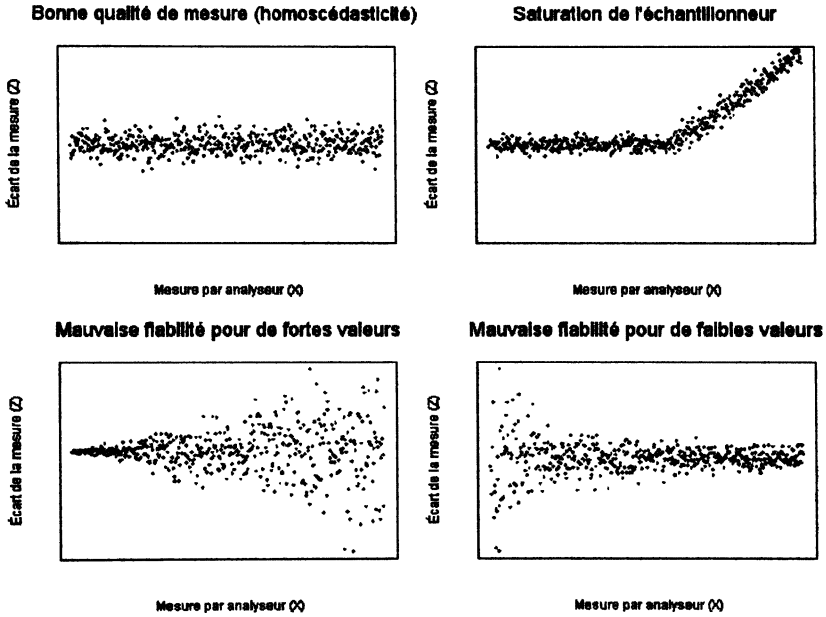

FIG 1. — Qualité de la mesure par échantillonnage passif par rapport à un référentiel : quelques exemples.

Modélisation

Notre objectif étant d'expliquer le biais de mesure dû à l'échantillonneur, nous avons choisi de modéliser Z_{ijk} en fonction de covariables environnementales agrégées :

$$E[Z_{ijk}] = f(A_{ijk})$$

où A_{ijk} est le vecteur des p covariables agrégées, $A_{ijk} = (A_{ijk}^{(1)}, A_{ijk}^{(2)}, \ldots, A_{ijk}^{(p)})$. Or, il apparaît clairement qu'il existe une forte corrélation entre les mesures issues d'un même analyseur i et pour une même campagne j. L'hypothèse d'indépendance des observations n'est donc vraisemblablement pas respectée.

Pour analyser ce type de données agrégées, différentes modélisations sont rencontrées dans le domaine écologique et statistique. Nous présentons succinctement les principales et discutons de leur applicabilité à notre problématique.

(i) Modèles stratifiés par grappes ou modèles régionaux

Cette méthode de modélisation consiste à estimer un modèle spécifique à chaque grappe. Pour la grappe ij (i représentant l'analyseur et j la campagne de mesure) l'écriture du modèle est la suivante :

$$Z_{ijk} = \beta_{0,ij} + \beta_{1,ij} A_{ijk}^{(1)} + \beta_{2,ij} A_{ijk}^{(2)} + \ldots + \beta_{p,ij} A_{ijk}^{(p)} + \varepsilon_{ijk} \text{ avec } \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_{ij})$$

et où $(\beta_{0,ij}, \ldots, \beta_{p,ij})$ est le vecteur des coefficients de régression propre à la grappe ij.

Seule l'hypothèse d'indépendance des observations au sein de chaque grappe est réalisée. Cependant, uniquement des variables d'un niveau d'agrégation inférieur à celui des grappes peuvent être prises en compte, ce qui rend ce modèle inadapté à notre problématique. De plus, ce modèle est souvent peu satisfaisant du fait de la répartition des effectifs dans les différentes grappes, synonyme de perte de puissance.

(ii) Modèles à effets fixes ou modèles contextuels

L'intégration des variables agrégées dans ce type de modèle se fait en tant qu'effets fixes. L'écriture du modèle est donc classique aux modèles linéaires généralisés (Mc Cullagh and Nelder, 1983) :

$$Z_{ijk} = \beta_0 + \beta_1 A_{ijk}^{(1)} + \beta_2 A_{ijk}^{(2)} + \ldots + \beta_p A_{ijk}^{(p)} + \varepsilon_{ijk} \text{ avec } \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma)$$

Comparativement aux modèles régionaux, ces modèles contextuels sont plus puissants puisqu'ils s'appuient sur l'ensemble de l'échantillon. Cependant, mesurer l'effet réel des grappes ne comporte aucun intérêt au vue de notre problématique, et nécessiterait l'introduction dans le modèle d'un nombre de variables indicatrices de la grappe trop important (22 grappes au total).

(iii) Modèles à effets mixtes (modèles multiniveaux)

Les modèles à effets mixtes constituent un moyen d'étudier des données non indépendantes par l'introduction d'effets aléatoires (Pinhero and Bates, 2000; Courgeau, 2004). La variance totale est alors expliquée par des effets moyens sur l'échantillon (effets fixes) et des effets variant selon la grappe (effets aléatoires).

Ce modèle, pourrait être défini de la façon suivante :

$$Z_{ijk} = (\beta_0 + \beta_{0,ij}) + (\beta_1 + \beta_{1,ij}) A_{ijk}^{(1)} + \beta_2 A_{ijk}^{(2)} + \ldots + \beta_p A_{ijk}^{(p)} + \varepsilon_{ijk}$$

avec $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_\varepsilon)$, $\beta_{0ij} \sim \mathcal{N}(0, \sigma_0)$, $\beta_{1ij} \sim \mathcal{N}(0, \sigma_1)$ et β_{0ij}, β_{1ij}, ε_{ij} indépendants.

Dans ce cas particulier, deux effets aléatoires ont été introduits sur les deux premiers coefficients de régression. Ils permettent de mesurer la variabilité de

l'intercept et d'un effet fixe selon les grappes ij. Ce modèle est dit à grappe spécifique.

Il possède un réel intérêt lorsque le nombre de grappes est important (seules quelques variances sont estimées) et que ces dernières modifient l'effet des facteurs non agrégés (biais écologique). Or, dans notre problématique d'évaluation de la fiabilité des échantillonneurs à l'O_3, seul l'effet moyen des variables agrégées nous intéresse.

(iv) Modèles marginaux

Ce type de modèle moyenné sur l'échantillon permet la prise en considération d'éventuelles corrélations (Diggle, Liang and Zeger, 1994). Le modèle s'écrit de la façon suivante :

$$Z_{ijk} = \beta_0 + \beta_1 A_{ijk}^{(1)} + \beta_2 A_{ijk}^{(2)} + \ldots + \beta_p A_{ijk}^{(p)} + \varepsilon_{ijk} \text{ avec } \varepsilon \sim \mathcal{N}(0, V)$$

Où ε est le vecteur des résidus de la régression de dimension $(5\sum_i m_i)$: $\varepsilon = (\varepsilon_{111}, \varepsilon_{112}, \ldots, \varepsilon_{115}, \varepsilon_{121}, \ldots, \varepsilon_{4m_45})'$ et où V est la matrice de variance-covariance des résidus. V est diagonale par blocs (5×5) ; pour chaque bloc, les termes non diagonaux ne sont pas nuls ce qui traduit le fait que seules les observations appartenant à 2 blocs différents sont considérées comme indépendantes.

Ainsi, l'interprétation des coefficients de la régression se fait en fonction de l'espérance marginale et est identique à celle des effets fixes d'un modèle linéaire à effets mixtes. Dans les deux modèles, les coefficients représentent l'effet moyen des covariables sur la variable d'intérêt quelle que soit la grappe.

Ce dernier modèle est donc particulièrement adapté à nos données. L'estimation des paramètres peut être basée sur les Equations d'Estimation Généralisées (GEE) (Liang and Zeger, 1986). Celles-ci s'appuient sur des matrices de corrélation de travail, $R_{ij}(\zeta)$, introduites dans l'écriture de la matrice de variance-covariance, permettant de tenir compte de corrélations entre les observations. ζ représente un vecteur de paramètres de corrélation inconnu. En fonction des hypothèses de corrélation, différentes matrices de travail ont été développées (identité, auto-régressive d'ordre 1, ...). Il s'agit donc de trouver celle qui s'adapte le mieux à notre problématique, sans que le nombre de paramètres à estimer soit trop important.

Dans notre situation, il semble que la matrice échangeable, pour laquelle tous les termes de corrélation des campagnes sont égaux pour un même analyseur, soit la mieux adaptée. En effet, la corrélation des mesures ne dépend que des caractéristiques propres à chaque lieu (analyseur) et à chaque campagne. Elle s'écrit pour un bloc donné :

$$R_{ij}(\zeta) = \begin{pmatrix} 1 & \zeta_i & \zeta_i & \zeta_i & \zeta_i \\ \zeta_i & 1 & \zeta_i & \zeta_i & \zeta_i \\ \zeta_i & \zeta_i & 1 & \zeta_i & \zeta_i \\ \zeta_i & \zeta_i & \zeta_i & 1 & \zeta_i \\ \zeta_i & \zeta_i & \zeta_i & \zeta_i & 1 \end{pmatrix}$$

3.2. Modélisation de l'exposition individuelle à la pollution atmosphérique

L'objectif principal de cette étude est la modélisation de l'exposition individuelle à la pollution atmosphérique en fonction d'un nombre limité de variables individuelles et agrégées. Les déterminants de cette exposition pouvant être différents selon le polluant considéré (NO_2 ou O_3), un modèle spécifique à chacun d'eux sera construit.

Pour chaque participant, nous disposons de deux mesures obtenues par les échantillonneurs passifs (une par polluant). Les concentrations sont exprimées en $\mu g/m^3$ et constituent les variables à expliquer lors de l'écriture des modèles prédictifs.

Au préalable, une analyse descriptive des variables à expliquer sera réalisée pour évaluer la nécessité d'une normalisation.

Soit Y_i l'exposition individuelle de l'individu i à l'un des deux polluants. La modélisation de la variable Y_i se fera en fonction de variables individuelles (recueillies à partir du questionnaire) et de variables agrégées (données du réseau de pollution et données météorologiques) :

$$E[Y_i] = f(I_i, A_i)$$

où $I_i = (I_i^{(1)}, I_i^{(2)}, \ldots, I_i^{(q)})'$ est le vecteur des q covariables individuelles pour l'individu i et où $A_i = (A_i^{(1)}, A_i^{(2)}, \ldots, A_i^{(p)})'$ est le vecteur des p covariables agrégées.

Dans cette problématique d'analyse de l'exposition individuelle, nous considérons l'hypothèse d'indépendance entre les individus (les mesures n'étant pas répétées sur une même personne).

Compte-tenu des arguments cités dans la section précédente nous appliquerons la théorie classique des modèles linéaires généralisés. En effet, la seule source de corrélation potentielle entre individus est la mesure du polluant faite à partir de l'analyseur. Or cette variable sera prise en compte dans le modèle en tant qu'effet fixe.

Le modèle peut ainsi s'écrire :

$$Y_i = \beta_0 + \beta_1 I_i^{(1)} + \ldots + \beta_q I_i^{(q)} + \beta_{q+1} A_i^{(1)} + \ldots + \beta_{q+p} A_i^{(p)} + \varepsilon_i \text{ avec } \varepsilon_i \sim \mathcal{N}(0, \sigma)$$

Dans le domaine environnemental, la notion de biais écologique est une donnée importante à prendre en considération (Greenland, 2002 ; Courgeau, 2004). En effet, les effets individuels peuvent être liés au contexte. Cette notion de modification d'effet peut être prise en considération par l'ajout d'une interaction dans notre modèle de régression. Par exemple :

$$Y_i = \beta_0 + \beta_1 I_i^{(1)} + \ldots + \beta_q I_i^{(q)} + \beta_{q+1} A_i^{(1)} + \ldots + \beta_{q+p} A_i^{(p)} + \beta_{q+p+1} I_i^{(2)} A_i^{(1)} + \varepsilon_i$$

avec $\varepsilon_i \sim \mathcal{N}(0, \sigma)$

Dans ce cas l'effet de la variable individuelle $I^{(2)}$ est modifié par la variable contextuelle $A^{(1)}$ et β_{q+p+1} mesure cette modification. Le test d'hypothèse $\beta_{q+p+1} \neq 0$ permet de tester la présence du biais écologique.

4. Discussion

Des études montrent qu'il existe une corrélation entre l'exposition individuelle des sujets et les mesures d'un analyseur situé à moins de 5 km de la zone de vie du sujet. Cependant les déterminants personnels modifient cette relation comme cela a été observé à Montpellier et en Ile de France (Bernard *et al.*, 1997, Bernard *et al.*, 1999; Saintot *et coll.*, 2000).

La mesure de l'exposition au NO_2 par échantillonnage passif est une méthode éprouvée, il n'en est pas de même pour l'O_3 dont les mesures peuvent se révéler plus ou moins fiables selon le contexte. Les conditions climatiques (chaleur, humidité) peuvent être à l'origine de cette variabilité dans les mesures. De plus, concernant l'étude Mexicaine, certaines zones d'études (Mexico ville notamment) constituent une zone de trafic intense, ainsi une interaction avec d'autres polluants, comme les particules, est envisageable.

La nature des données et la difficulté manifeste de considérer plusieurs covariables agrégées et corrélées au sein d'un modèle, nous orienteront vers certaines méthodes en particulier. Ce choix *a priori* est cependant pédagogique. En effet, les différentes modélisations devront être comparées, le choix d'un modèle final pouvant être guidé par l'utilisation d'un critère de comparaison de modèles (critère d'information d'Akaike, critère d'information de Bayes).

Ainsi, pour l'analyse de fiabilité, nous privilégierons dans un premier temps une modélisation par GEE, pour laquelle le choix de la matrice de travail impose des hypothèses plus ou moins fortes quant à la structure de la matrice de variance-covariance. Ce choix devra alors être validé par la minimisation des critères d'information. Pour la modélisation de l'exposition individuelle, un modèle linéaire généralisé semble le mieux adapté. Cependant l'apport d'une éventuelle structure de corrélation pourra être évalué.

L'inclusion des variables météorologiques fera l'objet d'une attention particulière. En effet, beaucoup d'entre elles sont corrélées, la valeur des unes pouvant être modifiée par celle des autres. Ainsi, les Modèles Additifs Généralisés, permettant la prise en compte de formes fonctionnelles complexes entre la variable réponse et les covariables, pourront être utilisés (Hastie and Tibshirani, 1990; Dominici *et al.*, 2002).

D'une manière générale, quelle que soit la méthode de modélisation choisie, le mélange de variables individuelles et de variables agrégées conduit à des difficultés d'interprétation des coefficients du modèle. Tout d'abord, il est important d'avoir conscience que les effets individuels et les effets contextuels sont indissociables. Les effets individuels ne doivent jamais être ignorés, même si l'objectif est uniquement contextuel. En effet, lorsque les variables individuelles sont résumées au travers des groupes, la variable mesurée sur le groupe est expliquée par la somme d'effets individuels et d'effets contextuels.

C'est ce que l'on appelle le biais écologique. *A contrario*, la seule prise en compte de données individuelles conduit à l'erreur atomiste.

Un autre problème sérieux réside dans la manière artificielle et arbitraire de créer les groupes, sans lien obligatoire avec la variable d'intérêt. Une solution peut être l'utilisation des groupes déjà disponibles et le choix de niveaux plus pertinents (Greenland, 2002).

Au cours de cette étude, des marqueurs biologiques comme critères d'exposition ont été retenus. Ainsi, une prise de sang a été réalisée au cinquième jour de l'enquête sur les sujets à jeun. Une modélisation des marqueurs, comme variable de substitut à expliquer, pourrait nous permettre de vérifier si les variables incluses sont analogues à celles du modèle d'exposition à la pollution (puisque dans ce cas la variable à expliquer est mesurée par l'échantillonneur passif) et de tester ainsi la robustesse de nos résultats.

Remerciements

L'étude conduite au Mexique a été subventionnée par le Ministère de l'Environnement (INERIS). Nous souhaitons remercier le Dr. Eric Parent pour ses précieux conseils et commentaires qui ont contribué à l'amélioration de ce texte.

Références

ATKINS D. H. F., LAW D. V., SANDALLS J., HOUGH A. M. and STEVESSON K. J. (1986). The measurement of nitrogen dioxide in the outdoor environment using passive diffusion samplers. *United Kingdom Atomic Energy Authority, environmental and medical sciences division*, Harwell Laboratory eds., Oxfordshire (UK) (Report AERE R.12133).

BERNARD N. L., ASTRE C. M., VUILLOT B., SAINTOT M. J. and GERBER M. J. (1997). Measurement of background urban nitrogen dioxide pollution levels with passive samplers in Montpellier, France. *Journal of Exposure Analysis and Environmental Epidemiology*, **7(2)**, 165-178.

BERNARD N., GERBER M. and ASTRE C. (1999). Ozone measurement with passive samplers : validation use for ozone pollution assessment in Montpellier, France. *Environmental Science and Technology*, **33**, 217-222.

COURGEAU D. (2004). *Du groupe à l'individu – Synthèse multiniveau.* Paris, Édition de l'Institut National d'Études Démographiques.

DESQUEYROUX H. et MOMAS I. (1999). Synthèse des études épidémiologiques de panel de 1987 à 1998. *Pollution Atmosphérique et Santé*, **47**, 361-375.

DIGGLE P. J., LIANG K-Y. and ZEGER S. L. (1994). *Analysis of Longitudinal Data.* New-York, Oxford University Press.

DOMINICI F., MCDERMOTT A., ZEGER S. L and SAMET J. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology*, **156 (3)**, 193-203.

GREENLAND S. (2002). A review of multilevel theory for ecologic analyses. *Statistics in Medicine*, **21**, 389-395.

HASTIE T. J. and TIBSHIRANI R. J. (1990). *Generalised Additive Models.* London, Chapman and Hall.

HÉMON D. (1995). Recherche épidémiologique sur l'environnement et la santé : quelques aspects méthodologiques. *Revue d'Épidémiologie et de Santé Publique*, **43**, 395-411.

KUNZLI N., KAISER R., MEDINA S., STUDNICKA M., CHANEL O., FILLIGER P., HERRY M., HORAK F., PUYBONNIEUX-TEXIER V., QUENEL P., SCHNEIDER J., SEETHALER R., VERGNAUD J-C. and SOMMER H. (2000). Public-health impact of outdoor and traffic-related air pollution : a European assessment. *The Lancet*, **356**, 795-801.

LIANG K.-Y. and ZEGER S. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121-130.

MC CULLAGH P. and NELDER J. A. (1983). *Generalized Linear Models.* London, Chapman and Hall.

PALMES E. D., GUNNISON A. F., DIMATTIO J. and TOMCZYK C. (1976). Personal sampler for nitrogen dioxide. *American Industrial Hygiene Association Journal*, **37**, 570-577.

PINHERO J. C. and BATES D. M. (2000). *Mixed-effects models in S and S-PLUS.* New-York, Springer-Verlag, New-York.

SAINTOT M., BERNARD N., ASTRE C. and GERBER M. (1999). Ozone exposure and blood antioxidants : a study in a periurban area in Southern France. *Archives Environmental Health*, **54**, 34-39.

SAINTOT M., BERNARD N., GERBER M., ASTRE C., GALAN P. and HERCBERG S. (2000). Déterminants des polluants atmosphériques photo-oxydants (NO_2 et O_3) dans une population de l'Ile de France. *Revue d'Épidémiologie et de Santé Publique*, **48**, 2S54-61.

SMITH A. H. (1988). Epidemiologic input to environmental risk assessment. *Archives Environmental Health*, **43**, 124-9.