

PRÉVISIBILITÉ A 24 HEURES DES CONCENTRATIONS EN OZONE A DIJON ET CHALON-SUR-SAONE

M.L. HOUZE ⁽¹⁾, Y. RICHARD ⁽¹⁾ et S. MONTEIRO ⁽²⁾

(1) Centre de Recherches de Climatologie, UMR5080 CNRS – Université de Bourgogne,
6 Bd Gabriel, 21000 Dijon, Tel 03 80 39 38 22 Fax 03 80 39 57 41,
E-mail: mlhouze@u-bourgogne.fr

(2) ATMOSF'air Bourgogne Centre-Nord, 5 rue Pasteur, 21000 Dijon, Tel 03 80 38 92 32
Fax 03 80 36 22 17

Résumé:

Conformément aux objectifs fixés par la loi sur l'air, les réseaux de surveillance de la qualité de l'air ATMOSF'air Bourgogne Centre-Nord (Dijon) et Bourgogne Sud (Chalon-sur-Saône) souhaitent se doter d'un outil de prévision au lendemain de la pollution. La collaboration entre Université de Bourgogne et réseaux s'inscrit dans ce contexte. Les premiers travaux montrent qu'en Bourgogne le risque principal concerne l'ozone (Houzé, 1999 ; Richard *et al.*, 2000). Ces études mettent en évidence la prévalence des conditions météorologiques sur les variations locales des concentrations des précurseurs.

Pour sept stations dijonnaises et chalonnaises les données quotidiennes d'ozone, variable expliquée ou prédicte, sont confrontées à des prédicteurs constitués de l'ozone prévu par le Laboratoire de Météorologie Dynamique (Vautard *et al.*, 2000), de précurseurs (CO, NO, NO₂) relevés dans ces mêmes stations, ainsi que de données météorologiques mesurées le lendemain dans les stations Météo-France de Dijon et de Chalon-sur-Saône. Les modèles de prévision utilisent deux méthodes linéaires : régression multiple pas à pas et analyse discriminante. Nous avons établi un modèle pour chacune des sept stations en restreignant le nombre de prédicteurs à cinq. Les premiers résultats, environ 65% de variance expliquée pour les régressions et 68% de bonne classification pour l'analyse discriminante, montrent qu'une prévision de ce type, en mode opérationnel, est envisageable. Pour cela il faudrait disposer en temps réel des prévisions Météo-France et LMD.

Abstract:

The ATMOSF'air networks monitoring air quality would like to be able to predict ozone concentrations 24 hours in advance in accordance with the air quality law of 1996. The University of Burgundy is collaborating with the networks on this project. Primary studies show that tropospheric ozone is the main pollutant in Burgundy and that the meteorological conditions are very important when considering ozone concentrations (Houzé, 1999 ; Richard *et al.*, 2000).

We predict the ozone hourly maximal value (predictand) using three types of predictors. First are ozone level from Laboratory of Meteorology Dynamic (LMD) calculated from air mass back trajectories and precursor integration (Vautard *et al.*, 2000), second next day meteorological data measured in Dijon and Chalon-sur-Saône Météo-France stations, and third the pollution variables (precursors): CO, NO, NO₂. Predicting models use two linear methods: Linear Multiple Regression and Discriminant Analysis. We obtained one model for each ATMOSF'air station. We used five predictors. Once models have been established, improvement can be made. First results show that this predicting is possible in operational use. Regression models explain about 65% variance and Discriminant Analysis models reach about 68% good classification. We

need Météo-France and LMD data in real time in order to provide an operational model.

Mots-clés: ozone, prévisibilité à 24h, méthodes statistiques linéaires, Bourgogne

Key-Words: ozone, predicting 24 hours in advance, statistical linear methods, Burgundy

Introduction

Les activités humaines (agricoles, industrielles, transports...) engendrent une modification de la composition chimique de l'atmosphère. La décomposition chimique ou photochimique des polluants primaires à différentes altitudes conduit à la production d'espèces chimiques actives, tel l'ozone troposphérique, qui vont à leur tour perturber les équilibres physico-chimiques de l'atmosphère (Mégie, 1988). Dans l'hémisphère nord, les concentrations de fond en ozone étaient, au début du siècle, d'environ 20 $\mu\text{g}/\text{m}^3$. Elles peuvent atteindre aujourd'hui 60, 80, voire 100 $\mu\text{g}/\text{m}^3$ en moyenne (Mégie *et al.*, 1993). Il en résulte des problèmes de santé publique, de conservation du patrimoine, des dommages sur la végétation et une dégradation de la visibilité.

Le principe de précaution nécessite d'anticiper les épisodes de pollution pour mettre en œuvre des mesures d'urgence en cas de dépassement du seuil d'alerte. Ceci implique une prévision à court terme (24h), voire à plus longue échéance (2 à 5 jours). L'action sur les causes de la pollution tient à l'importance de la prévision (Chartier, 1999). Par conséquent, le développement de modèles de prévision devient une nécessité (Gasso, 1999). Aux USA et en Allemagne, la prévision à 24h est déjà obligatoire (Greene *et al.*, 1999). Les méthodes de prévision les plus utilisées font appel aux outils statistiques, méthodes pertinentes (Simpson et Layton, 1983 ; Robeson et Steyn, 1990) qui ont la caractéristique de ne pas intégrer dynamiquement la chimie de l'ozone, particulièrement délicate. Ce sont en général des modèles d'usage local (Ghattas, 1998). Le Centre de Recherche en Automatique de Nancy (CRAN) a effectué des comparaisons entre différentes techniques statistiques (Rouil et Wroblewski, 1999). Entre les résultats obtenus avec les modèles linéaires et ceux issus de réseaux neuronaux ou basés sur la logique floue, peu de différences ont été notées.

Dans ce contexte, la collaboration entre le Centre de Recherches en Climatologie (CNRS - Université de Bourgogne) et les réseaux ATMOSF'air, a pour but le développement de modèles statistiques simples de prévisibilité de l'ozone pour le lendemain.

1. Données

Nous disposons d'une base de données historiques, commune à sept stations dijonnaises et chalonnaises, allant du 01 avril au 30 septembre 1999. Peu de pics de pollution à l'ozone sont observés du fait de la brève période couverte, de la faible densité de population et de la faiblesse du tissu industriel.

1.1 Le prédictand : L'ozone troposphérique à Dijon et Chalon-sur-Saône

Le prédictand est la valeur à prévoir. Elle prend la forme finale de la concentration

maximale d'ozone du lendemain (J+1). Nous avons sélectionné les valeurs journalières maximales d'ozone de 10h à 18h TU inclus, tranche horaire où sont concentrées les valeurs les plus élevées (Houzé, 1999).

1.2 Les prédictesurs chimiques ou précurseurs

L'ozone du jour (J) est utilisé comme prédictesur afin de prendre en compte un éventuel phénomène d'accumulation (Vergez, 1999). Nous avons pris en compte la vitesse de production de l'ozone de 06h à 12h TU ($(\max O_3 - \min O_3)/7$), ainsi que les valeurs moyennes et maximales de 10h à 18h TU le jour J. Les oxydes d'azote constituent l'un des principaux précurseurs de la pollution photochimique, ils sont mesurés dans toutes les stations des réseaux ATMOSF'air. Les COVNM (Composés Organiques Volatiles Non Méthaniques) ne sont pas mesurés en Bourgogne. Dans les milieux riches en NO_x (>5ppb) la production d'ozone augmente avec celle des COVNM. Le fait que les modèles statistiques établis ne prennent pas en compte les émissions biogéniques ou anthropogéniques de COVNM pourrait expliquer certaines approximations des résultats. Sont intégrées aux modèles statistiques les données de CO, NO, NO₂ et NO_x ($NO_x = NO + (1.25/1.91) NO_2$), selon deux valeurs :

- la moyenne 10h - 18h TU du jour J, afin de minimiser l'impact d'éventuelles erreurs (valeurs anormalement fortes) ;
- le maximum du jour J, un changement de temps rapide pouvant rendre la moyenne peu pertinente.

1.3 Les prédictesurs météorologiques : données locales observées à J+1

L'intérêt d'utiliser des données météorologiques observées dans un modèle (Tabl. 1)

Tableau 1 : Paramètres météorologiques locaux et conséquences sur les concentrations en ozone

Paramètres	Effets
Température	→ conditionne la stabilité des basses couches de l'atmosphère et la dispersion verticale des polluants (Mégie <i>et al.</i> , 1993)
Rayonnement solaire	→ effet direct sur la production d'ozone en accélérant les réactions chimiques d'oxydation (NO, CH ₄ , COVNM) (Wark et Warner, 1976)
Vitesse du vent	→ dilue continuellement, dans la direction du transport du panache, les polluants libérés au point d'émission → brasse l'air, favorise les advections et les échanges verticaux
Précipitations	→ peuvent induire un phénomène de <i>lessivage</i> de l'air
Humidité relative	→ corrélée avec la condensation de la vapeur d'eau sur le sol (rosée), celle-ci augmente le dépôt de l'ozone

est de s'affranchir des erreurs de prévision météorologique. Ces données sont donc pertinentes dans la phase de développement du modèle. Bien entendu en mode opérationnel elles ne sont pas utilisables.

1.4 Un prédicteur synthétique : l'ozone prévu pour Dijon via les rétrotrajectoires des masses d'air

Les données de prévision des concentrations en ozone effectuées au Laboratoire de Météorologie Dynamique (LMD-CNRS) dans le cadre du projet communautaire de recherche sur la pollution atmosphérique ESQUIF (Etude et Simulation de la Qualité de l'air en Ile-de-France) sont utilisées (Vautard *et al.*, 2000). Les données d'ozone prévu sont issues des rétrotrajectoires qui permettent de déterminer l'origine de la masse d'air arrivant à 12hTU à Dijon. Elles permettent de prendre en compte la contribution d'échelle européenne mais négligent les émissions locales de précurseurs et les conditions météorologiques locales.

2. Méthodes

Les modèles statistiques proposés concerneront :

- l'ensemble des jours, et non uniquement les pics, définis généralement comme dépassements des normes (UE ou OMS) ;
- la prévisibilité des concentrations maximales quotidiennes (les moyennes étant trop influencées par les faibles concentrations nocturnes) ;
- la prévisibilité par station, et non par agglomération, pour prendre en compte les différences importantes entre centre et périphérie (Richard *et al.*, 2000). En outre, ce choix qui conduit à effectuer sept études de prévisibilité distinctes, indépendantes les unes des autres, est le moyen de tester la robustesse de la méthode.

2.1 Régression linéaire multiple pas à pas

La régression multiple pas à pas consiste à expliquer un phénomène Y, ici les concentrations maximales d'ozone à J+1, par la conjonction des variations de plusieurs phénomènes. Suivant les différentes bibliographies, il a été décidé de prendre en compte cinq prédicteurs (Vergez, 1999).

2.2 Analyse discriminante

L'analyse discriminante est une méthode linéaire. Les fonctions linéaires discriminantes sont les combinaisons linéaires de ces variables, dont les valeurs séparent au mieux des classes connues *a priori*. Parmi toutes les combinaisons linéaires de variables, nous cherchons celles qui ont une variance externe maximale et une variance interne minimale (Lebart *et al.*, 1988). Il a été décidé d'utiliser trois classes équiprobables si possible. La classe 1 regroupe les faibles valeurs d'ozone, la classe 2 les valeurs moyennes et la classe 3 les valeurs les plus fortes. Ces notions de fortes, moyennes et faibles valeurs sont liées à la distribution statistique. Elles sont relatives aux concentrations de chaque station et n'ont rien de commun avec les valeurs guides.

⇒ Pour les deux méthodes, dont les performances seront confrontées, nous avons utilisé la procédure de test dite *Jack-knife* (validation croisée ou cross-validation). Cette méthode, rendue obligatoire du fait du faible historique commun des sept stations, permet d'évaluer la performance réelle d'un modèle de prévision avant qu'il ne soit appliqué (Thiria *et al.*, 1997). Elle consiste en une resélection d'échantillon.

3. Résultats

3.1 Régression

Dans tous les modèles, sauf à *Daix*, les valeurs d'ozone de la veille (maxima ou moyennes) sont prises en compte et ont un coefficient positif. Ceci s'explique par le fait que les variations d'ozone d'un jour à l'autre sont en général peu importantes. Néanmoins les performances des modèles (Tabl. 2) sont bien supérieures à celles des expériences réalisées sur les seules valeurs du jour *J* (situation $J+1 = aO_3J + b$) pour lesquelles les coefficients de détermination oscillent entre 0,13 et 0,26.

Les sept modèles prennent également en compte les valeurs d'ozone prévu par le LMD (Vautard *et al.*, 2000). La seule variable météorologique présente dans les sept modèles est la moyenne journalière d'humidité relative. Elle a toujours un coefficient négatif. Les autres paramètres pris en compte sont : la température, l'amplitude thermique, la vitesse du vent, l'amplitude des concentrations en NO, les concentrations en NO ou en NO_x. Pour les sept stations étudiées, les prédicteurs sélectionnés sont les mêmes d'un jour à l'autre (sauf quelques cas rares), ceci traduit leur robustesse.

Les erreurs moyennes au carré n'excèdent pas 18 µg/m³ (erreurs moyennes de l'ordre de +/- 4 µg/m³). Que ce soit à Dijon ou à Chalon-sur-Saône, les modèles les moins performants sont relatifs aux stations situées en centre ville : Pasteur pour Dijon, Centre pour Chalon-sur-Saône. Le meilleur modèle est celui d'*Hilaire* avec une performance de 72 % (Tabl. 2).

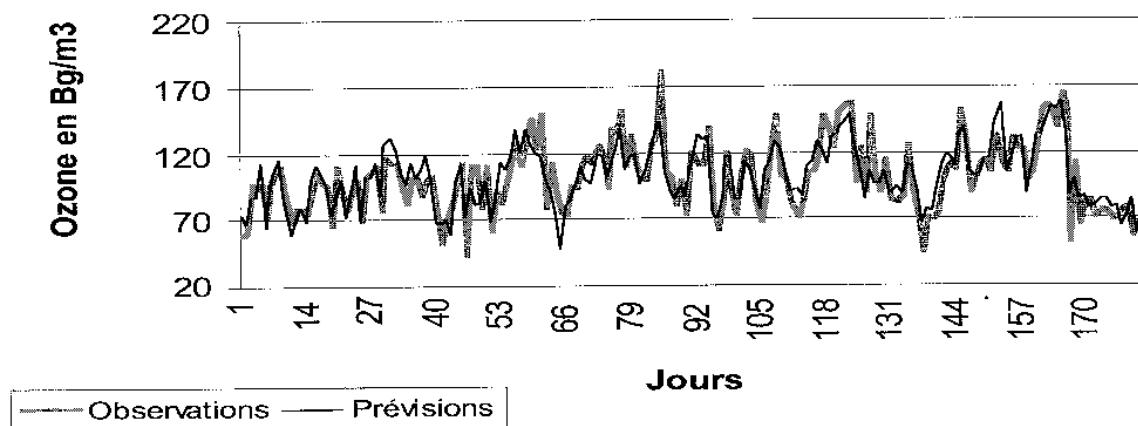


Figure 1: Observations et prévisions des concentrations en ozone à Hilaire

Tableau 2: Résultats des régressions

Station	Situation	Coefficient de détermination	Erreur moyenne au carré ajusté
Pasteur	(Dijon, centre)	0.4768	15.2454
Tarnier	(Dijon, centre)	0.6891	13.773
Daix	(Dijon, périphérie)	0.7475	17.948
Balzac	(Dijon, centre)	0.6762	14.3287
Centre	(Chalon, centre)	0.6345	14.6895
Champforgeuil	(Chalon, périphérie)	0.6938	15.1562
Hilaire	(Chalon, centre)	0.7204	14.504
Moyenne		0,6626	15,0921

En contrepartie, pour chaque station, les écart-types calculés sur les observations sont systématiquement plus importants que ceux des estimations. Le modèle montre ses limites pour les prévisions des valeurs exceptionnelles, dont les pics, surtout si ceux-ci sont éphémères. Ainsi les pics isolés sont-ils moins bien prévus que les pics qui s'inscrivent dans une séquence à fortes concentrations (Fig. 1).

3.2 Application d'un correctif en fonction du rythme hebdomadaire

Les modèles étant établis, nous avons cherché à savoir s'il était possible d'en améliorer les résultats. L'hypothèse est que les modèles obtenus ne peuvent pas prévoir les hausses de concentration en ozone qui se produisent lorsque le trafic automobile est moindre, c'est à dire le plus fréquemment le week-end (Houzé, 1999). Dans ce cas, la destruction locale d'ozone par le NO est plus faible. Nous avons donc calculé les moyennes des résidus pour chaque jour de la semaine et testé les différences entre résidus des jours ouvrés et résidus de week-end (test de Student). Seules les stations de *Daix* et *Hilaire* ont des résidus qui ne sont pas différents entre le week-end et la semaine. Pour toutes les autres stations, une valeur correctrice est affectée selon que l'on est un jour de semaine ou un jour de week-end (Tabl. 3). En pratique le correctif semaine/week-end intervient après que le modèle ait tourné. Il faut ajouter ou retrancher ces valeurs (moyennes des résidus) aux valeurs d'ozone prévues par le modèle.

Tableau 3 : Correctifs apportés aux valeurs d'ozone prévues (en $\mu\text{g}/\text{m}^3$)

Stations	Semaine	Week-end
Pasteur	-2.06	+5.47
Tarnier	-1.08	+2.79
Balzac	-1.68	+4.14
Centre	-2.13	+5.33
Champforgeuil	-1.05	+2.63

3.3 Analyse discriminante

Les variables utilisées pour effectuer l'analyse discriminante sont les mêmes que celles sélectionnées par la régression linéaire multiple pas à pas. Les classes créées sont équiprobables, non représentatives des seuils réglementaires. Elles sont relatives aux niveaux de pollution de chaque station. Les meilleurs résultats sont obtenus à *Daix* et à *Centre* (Tabl. 4). Les résultats les plus faibles, tout comme ceux obtenus avec les régressions linéaires multiples, sont observés à *Pasteur* (centre ville de Dijon).

Tableau 4: Résultats des analyses discriminantes (en %)

Station	Bonne classification	Erreur d'une classe	Erreur de deux classes
Pasteur	59.87	38.21	1.91
Tarnier	68.15	30.59	1.26
Daix	76.05	23.94	0
Balzac	65.18	31.85	2.96
Centre	75.27	22.52	2.19
Champforgeuil	67.58	31.32	1.09
Hilaire	69.78	29.12	1.09
Moyenne	68.41	29.99	1.59

Il n'y a pas de différence de qualité de prévisibilité entre les agglomérations dijonnaise et chalonnaise. Nous observons une majorité de bonnes classifications, peu d'erreurs de deux classes mais un pourcentage d'erreurs d'une classe assez élevé. Celles-ci peuvent relever d'une véritable erreur de prévision, mais peuvent également être le fruit d'un effet de seuil. Pour les jours où le modèle d'analyse discriminante se trompe de deux classes sur au moins une station, les moyennes des paramètres météorologiques (vitesse du vent, température, humidité relative, précipitations, rayonnement solaire) ne sont pas significativement différentes des jours où le modèle ne se trompe pas ou ne se trompe que d'une classe. Si les erreurs de deux classes, non-détection ou fausse alerte, sont synchrones sur plusieurs stations, elles ne peuvent pas être imputées à des phénomènes locaux ou/et de mesure, mais au modèle de prévision. Tel est le cas pour deux non-détections simultanées à Dijon : le 18 avril 1999 (*Pasteur* et *Tarnier*) et à Chalon-sur-Saône, le 18 mai 1999 (*Hilaire* et *Centre*). Mais aucune journée ne voit les modèles d'analyse discriminante se tromper simultanément sur l'ensemble des stations de Dijon ou de Chalon-sur-Saône.

Des erreurs de classification qui ne sont pas synchrones selon les stations, et des paramètres météorologiques qui ne sont pas statistiquement différents selon le type d'erreur, font que nous n'avons pas réussi à améliorer les résultats des modèles d'analyse discriminante. L'analyse discriminante donne une majorité de bons classements et très peu d'erreurs de type 2. *A priori* ce sont de bons résultats. Mais un probl-

ème se pose : en mode opérationnel les critères statistiques utilisés pour la création des classes sont peu pertinents. Il faudrait utiliser cette méthode avec les valeurs des seuils réglementaires. Ici c'est impossible car l'échantillon est trop réduit (période courte et dépassements des seuils trop rares).

Conclusion

Il est difficile d'établir une comparaison entre nos résultats et ceux des autres études car celles-ci sont généralement effectuées sur des zones plus polluées, donc en relation avec les seuils réglementaires.

Les résultats des modèles de prévisibilité développés à l'aide de l'analyse discriminante sont statistiquement de qualité. Deux stations, *Daix* et *Centre*, ont un taux de bonne classification supérieur à 70%, les fausses alertes et les non-détections sont peu nombreuses et asynchrones. Mais les classes utilisées n'ont pas de signification réelle vis à vis des réglementations en vigueur et ne permettent pas de déclencher une quelconque procédure. Il faudrait pouvoir faire des classes de valeurs représentatives de la pollution atmosphérique à l'ozone et non des classes équiprobables. Une solution serait d'utiliser les valeurs des classes utilisées pour calculer l'indice ATMO.

Les modèles de régression linéaire multiple correspondent plus à l'attente des réseaux, à savoir la possibilité de prévoir une valeur maximale journalière de concentration en ozone, tandis que l'analyse discriminante n'indique qu'un ordre de grandeur. Les modèles de régression linéaire donnent les meilleurs résultats dans les stations les plus éloignées des centres villes : *Daix*, *Champforgeuil* et *Hilaire* ont un coefficient de détermination de 70% environ. Mais ces modèles sont peu performants pour la prévision des valeurs sensibles, à savoir les pics de pollution.

L'objectif est de passer de modèles de prévisibilité à des modèles de prévision. Pour que ceux-ci soient exploitables par les réseaux ATMOSF'air, ils doivent intégrer les prévisions météorologiques à J+1. Il manque à notre étude des données à différents niveaux géopotentiels. En effet Roussel *et al.* (1997) soulignent l'importance du gradient thermique vertical. En été, il peut signer les nuits à ciel clair et la mise en place d'une inversion de température nocturne entraînant une mauvaise dispersion verticale des polluants.

Bibliographie

- CHARTIER P., 1999: Les outils de prévention à court terme de la pollution atmosphérique urbaine, *Théories Sciences Méthodes*, 1, janvier, p 61.
- GASSO G., 1999: *Elaboration de modèles de prévision des pointes d'ozone (sites de Nancy)*, Centre de Recherche en Automatique de Nancy, Ecole Nationale Supérieure de Géologie, Vandoeuvre, 22 p.
- GHATTAS B., 1998: Prévision des pics d'ozone par arbres de régression simples et agrégés par bootstrap, article soumis à *Statistique Appliquée*, 16 p.

- GREENE J.S., KALKSTEIN L.S., YE H. et SMOYER K., 1999: Relationships between Synoptic Climatology and Atmospheric Pollution at 4 US Cities, *Theoretical and Applied Climatology*, 62, 163-174.
- HOUZE M.-L., 1999: *Concentrations en ozone dans les agglomérations dijonnaise et chalon-naise et conditions météorologiques (avril-août 1998)*, Mémoire de maîtrise de géographie, Université de Bourgogne, 67 p.
- LEBART L, MORINEAU A. et FENELON J.-P. 1988: *Traitement des données statistiques, méthodes et programmes*, Ed. Dunod, Paris, 510 p.
- MEGIE G., 1988: Les modifications chimiques de l'atmosphère et leurs effets sur l'environnement, *Pollution Atmosphérique*, 177, 5-11.
- MEGIE G., BONTE J., CARLIER P., CHAVAUDRA J., DIZENGREMEL P., FEUGIER A., GRANIER C., HAUGLUSTAINE D., KANAKIDOU M., LE BRAS G., MARENCO A., MOUVIER G., TISSOT B., TOUPANCE G., TRUHAUT R., 1993: *Rapport NÆ30, Ozone et propriétés oxydantes de la troposphère*, Paris : Lavoisier, Coll. TEC & DOC, 253 p.
- RICHARD Y., HOuze M.-L., CHABIN J.-P. et THEVENIN D., 2000: Pollution en ozone en Bourgogne orientale, soumis à AIC, 8p.
- ROBESON S.M. et STEYN D.G., 1990: Evaluation and comparaison of statistical forecast models for daily maximum ozone concentrations, *Atmospheric Environment*, 24B, 303-312.
- ROUIL L. et WROBLEWSKI A., 1999: *Retour d'expérience en modélisation*, mesures de la qualité de l'air dans le cadre du LCSQA, 35 p.
- ROUSSEL I., ROUCHOUSE S. et BATTON-HUBERT M., 1997: Pollution atmosphérique urbain et climat local : l'exemple de l'agglomération stéphanoise en 1997, *Revue de Géographie de Lyon*, 72, 315-321.
- SIMPSON R.W. et LAYTON A.P., 1983: Forecasting peak ozone levels, *Atmosphéric Environment*, vol. 17, 9, 1649-1654.
- THIRIA S., LECHEVALLIER Y., GASCUEL O., et CANU S., 1997: *Statistiques et méthodes neuronales*, Ed. Dunod, Paris, 311 p.
- VAUTARD R., BEEKMANN M., ROUX J. et GOMBERT D., 2000: Validation of a deterministic forecasting system for the ozone concentrations over the Paris area, soumis à *Atmospheric Environment*, février, 30 p.
- VERGEZ G., 1999: *Prévisions des pointes d'ozone dans l'agglomération lyonnaise*, Lyon : Note de travail de l'école nationale de la météorologie, 669, 92 p.
- WARK K. et WARNER C.F., 1976: *Air pollution. Its origin and controle*, New-York, 519 p.