ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Instituto Nacional de Ecología

Analyse exploratoire d'un jeu de données de concentrations en ozone troposphérique

Application à la Zone Metropolitaine de la Vallée de Mexico

Elena Andrey, Travail pratique de master

Encadrement scientifique:
Régis Caloz – LaSIG (Laboratoire des Systèmes de l'Information Géographique)
Alain Clappier – LPAS (Laboratoire de Pollution Atmosphérique et Sols)

Lausanne, octobre 2004 – février 2005

Résumé

L’ozone troposphérique est une problématique environnementale de plus en plus répandue à l’échelle de la planète. L’augmentation démesurée de la taille des agglomérations, la croissance démographique et l’augmentation des émissions sont autant de facteurs qui contribuent à cette pollution de nature urbaine.

Le présent travail propose l’exploration statistique complète d’un jeu de données de concentrations en ozone troposphérique, mesuré dans l’une des plus grandes cités au monde : la Zone Métropolitaine de la Vallée de Mexico. La régionalisation des valeurs est effectuée à l’aide de méthodes d’analyse géostatistiques. Les surfaces interpolées sont ensuite interprétées du point de vue de la pollution atmosphérique. Finalement, cette étude tente d’insérer les résultats obtenus dans un contexte environnemental plus large, et essaie de mettre en avant les potentialités des Systèmes d’Information Géographiques pour la gestion de la qualité de l’air de la Zone Métropolitaine de la Vallé de Mexico.

Table des matières

1. Intérêt des Sciences de l'Information Géographique pour la pollution de l'air de la ZMVM.

La pollution atmosphérique par l'ozone troposphérique est un phénomène de nature spatiale pouvant être décrit par de nombreuses variables dont les interdépendances sont complexes. Les Sciences de l'Information géographique permettent d'appréhender ce type de grand système.

En effet, la représentation numérique par des moyens techniques performants de chacune des composantes spatiales de ce phénomène (*météorologie, démographie, topographie, etc.*) permet, outre une simple visualisation de chacune de ces composantes, le croisement de données par des traitments plus approfondis, dans le but de quantifier les interrelations.

Les différentes sciences plus particulièrement concernées par ce travail sont la géostatistique, ainsi que l'analyse et la visualisation multidimensionelle de l'espace géographique. L'atmosphère de cette agglomération étant l'une des plus polluées au monde, le gouvernement mexicain a mis en place dans les années '80 de nombreux outils, dont un très grand réseau de mesures des polluants atmosphériques principaux (*NOx, SO_2, O_3, PM10, CO, etc.*). Ce réseau, fiable et dense, a permis la production d'un volume important de données depuis lors, permettant ainsi l'accomplissement d'études et la publication de nombreux travaux dans les différents domaines relatifs à la pollution atmosphérique. Dès lors, l'exploitation d'un point de vue géostatistique de jeux de données d'une telle qualité, et la mise en relation de ces données avec tout autre variable impliquée ou objet exposé par la pollution de l'air peut être intéressante.

Ce travail traite donc du problème de la spatialisation de la concentration en ozone troposphérique, ainsi que de la représentation intégrée à un contexte environnemental plus large des résultats obtenus. Il n'est pas question de modélisation prédictionnelle, compte tenu de la complexité non seulement des mécanismes physico-chimiques de la formation de l'ozone, mais également de la complexité du comportement des variables qui influencent la propagation de ce polluant secondaire (*notamment des masses d'air au sol*).

Une représentation intégrée de l'information géographique relative à la pollution atmosphérique par l'ozone troposphérique pourrait constituer un outil performant d'aide à la décision, plus spécifiquement du point de vue de l'exposition chronique. Les représentations numériques de l'espace géographique pourraient de manière plus générale devenir nécessaires à la prise de décisions appropriées dans les domaines les plus variés de la gestion d'une ressource naturelle.

2. Contexte et objectifs généraux

Cette étude a été proposée à la suite d'un travail de semestre réalisé par M. Emmanuel Gaillard (*étudiant de l'Ecole des Mines de St-Etienne*) au Laboratoire de Pollution Atmosphérique et des Sols[1], sous l'encadrement scientifique de M. Alain Clappier. Le travail de E. Gaillard a identifié des répartitions spatiales caractéristiques pour l'ozone troposphérique à l'aide d'une méthode de classification statistique appelée la méthode mixte (*annexe 1*). L'idée dans un premier temps étant de réduire le volume de données à analyser (*mesures de concentrations horaires et séries annuelles pour chacune des stations*). Dans un deuxième temps une tentative de mise en correspondance des situations de pollution obtenues avec des données météorologiques regroupées selon la même méthode a été effectuée. Les paramètres météorologiques permettraient dans un troisième temps une approche de type prédictionnelle pour la concentration en ozone. Les répartitions spatiales obtenues ont été représentées au moyen d'outils de SIG.

Autant l'idée et les résultats obtenus par la méthode mixte sont très intéressants du point de vue du traitement d'une quantité aussi importante de données, autant la représentation de données spatiales doit obéir à certaines règles qui ont été omises dans la démarche d'E. Gaillard. Plus précisément :

- la cartographie d'un phénomène devrait représenter des valeurs « *simultanées* »
- le temps est une variable importante de la spatialisation de phénomènes naturels tels que l'ozone, pour lequel les concentrations peuvent varier très rapidement. Or, l'évolution dans le temps des valeurs de concentrations a été écartée pour une raison précise : le travail de E. Gaillard ne s'est pas intéressé à l'amplitude des pics de concentration, mais de manière plus générale à la localisation géographique de ceux-ci.

Le présent travail s'insère dans un contexte différent. La structure temporelle des données y est importante. La notion de gestion implique spécifiquement un suivi temporel de l'évolution d'un phénomène.

Les objectifs proposés à ce stade sont premièrement une analyse statistique exploratoire complète du jeu de données de concentration en ozone, d'un point de vue temporel, multivarié et spatial, afin de mettre en évidence des comportements généraux et plus spécifiques du phénomène. Cette analyse devrait pouvoir isoler dans le temps des épisodes cohérents et comparables. Elle devrait également pouvoir identifier une structure spatiale aux mesures en vue de leur régionalisation.

De plus, le traitement de ce type de données (*volume, valeurs nulles, etc.*) devrait indirectement mener à une « *méthodologie exploratoire* », ainsi qu'à l'utilisation d'outils de calcul et de développement.

Un séjour de trois semaines au Centre National de Recherche et de Formation en Environnement dans la ville de Mexico (*CENICA*) a ensuite permis de préciser le contexte environnemental du travail. Les visites de terrain, informations et données numériques obtenues ont permis de développer et d'étendre les objectifs mentionnés ci-dessus. Il s'agit dès lors de manière plus générale de mettre en évidence les potentialités des SIG pour la gestion de la pollution par l'ozone troposphérique dans la Zone Métropolitaine de la Vallée de Mexico.

[1] *LPAS, EPFL*

3. La qualité de l'air pour la Zone Métropolitaine de la Vallée de Mexico

La ZMVM désigne une surface urbanisée située sur le haut plateau des Etats-Unis du Mexique. Elle s'étend sur environ 1'500km^2 en 2000, et regroupe seize délégations constituant le District Fédéral (*entité territoriale politique, chef lieu du gouvernement fédéral mexicain*), ainsi que 37 autres municipalités de l'état de Mexico et une de l'état de Hidalgo (*voir figure 3.1*). La ville compte actuellement plus de 20 millions d'habitants, ce qui en fait la deuxième plus grande agglomération au monde après Tokyo.

Figure 3.1: représentation de la Zone Métropolitaine de la Vallée de Mexico

Evolution historique de la pollution

Une des principales inquiétudes pour la population de la ZMVM est la qualité de l'air, plus particulièrement la présence de particules et d'ozone troposphérique, dont les impacts sur la santé sont graves. Le problème de la mauvaise qualité de l'air dans la ville a été identifié et reconnu par le gouvernement ainsi que par la population depuis 1960 déjà. Depuis lors l'agglomération n'a cessé de rapidement se développer en termes de surface urbanisée et de nombre d'habitants, et par conséquent en termes de demande en transports et en consommation énergétique (*commerces, services, consommation résidentielle, industries, agriculture, etc.*).
Conscients du problème, dans les années 70 un réseau de mesures a été mis en place. Les valeurs despolluants principaux (*particules, NOx, SO_2, ozone, etc.*) sont depuis ce temps retransmises aux instances politiques et au public sous forme d'un indice d'exposition : l'IMECA (*Indice Metropolitano de Calidad del Aire*). Les premières législations environnementales (*Ley Federal para Prevenir y Controlar la Contaminación Ambiental, 1971*) et structures politiques pour l'environnement (*Subsecretaría de Mejoramiento del Ambiente, 1976-1982*) naissent également dans les années 70.

La crise économique des années 80 ainsi que le tremblement de terre de 1985 auront pour effet de limiter l'intérêt général pour les impacts de la pollution atmosphérique sur l'environnement et la santé publique, et par conséquent le développement de mesures préventives ainsi que le renforcement des normes de qualité de l'air.

Par la suite, au début des années 90 l'évolution de la situation deviendra critique. L'OMS désignera en 1992 dans un rapport du PNUE[2], à la suite d'épisodes extrêmement sévères de pollution, que l'agglomération de la ZMVM est la plus polluée au monde. Les pics d'ozone dépassent alors les 440 ppb ($\approx$ 530 $\mu g/m^3$: normes OPair = 120 $\mu g/m^3$, valeur moyenne sur 24 heures). Une série de mesures concrètes seront aussitôt appliquées.

Ces mesures comprennent principalement la réduction des émissions du secteur des transports (*pose de catalyseurs, élimination du plomb dans l'essence, amélioration de la qualité des carburants par rapport à la teneur en souffre dans le diesel notamment, mise en place de mesures de planification dans le secteur des transports publics urbains, etc.*). L'Etat tente également de limiter les émissions industrielles et commerciales (*fermeture de certains établissements,*

[2] *Programme des Nations Unies pour l'Environnement*

notamment une des plus importantes raffineries située dans le nord-est de la ville, « 18 de Marzo »), ainsi que de développer les outils de la planification territoriale (*délimitaion de zones industrielles et résidentielles, « restauration écologique », contrôle de l'extension sauvage et démesurée des quartiers d'habitation en périphérie, etc.*), éducation environnementale et recherche.

Les résultats en termes d'émissions porteront principalement sur une diminution drastique des concentrations en plomb et oxydes de soufre de l'atmosphère, ainsi qu'une diminution un peu moins marquée des émissions de monoxyde de carbone.

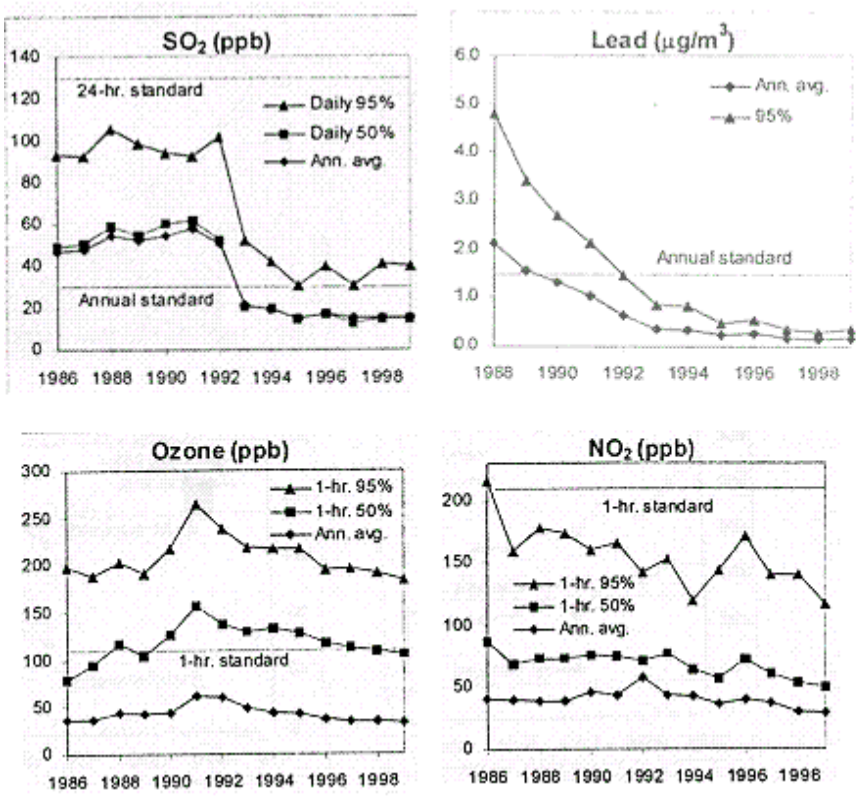


Figure 3.2 : Evolution des concentrations en Pb, SO2, ozone et NOx, Molina et al. 2002

La qualité de l'air reste cependant mauvaise à l'heure actuelle. Les normes pour l'ozone sont encore dépassées plus de 80% des jours (*Molina et al., 2002*).

Facteur géographique

Les facteurs de pollution sont nombreux et leurs interdépendances complexes. La ZMVM se distingue toutefois des autres mégalopoles de par sa situation géographique. Plus précisément par ses caractéristiques topographiques et climatiques.

Le plateau mexicain s'étend à une altitude d'environ 2'240m. La teneur en oxygène de l'atmosphère y est diminuée de 23% par rapport au niveau de la mer, ce qui implique des dépenses en carburant généralement plus importantes, ainsi que des effets sur la santé plus graves, compte tenu du fait que pour une même quantité d'oxygène une plus grande quantité de polluants sont inspirés.

Aussi, la ZMVM se situe dans un bassin topographique délimité par une chaîne de sommets montagneux, dont les plus importants culminent à plus de 5'200 m d'altitude au Sud-Ouest de la zone : le Popocatepetl (5452 m) et l'Ixtaccíhualtl (5'256 m). L'ouverture de cette chaîne au Nord permet l'intrusion d'un vent géostrophique, dont l'origine est liée aux conditions de pression à très grande échelle. Ce courant déplace généralement les émissions primaires industrielles du Nord-Est de la ville vers le Sud. Une plus petite ouverture au Sud-Est à ~2'500-2'700 m laisse déborder une brise de vallée formée sur les versants Sud de la chaîne, où l'ensoleillement est plus important. De la convergence Est-Ouest de ces deux masses d'air résulte une situation localement complexe, qui semblerait-il aurait tendance à limiter le phénomène de transport de la pollution.

D'un point de vue climatique, la latitude subtropicale à laquelle la ZMVM se trouve (*19°25'N, 99°10'W*) permet la distinction de deux saisons bien définies lors desquelles la formation d'ozone est différenciée:

- l'hiver, les conditions météorologiques sont de type anticylonique, avec vent léger et ciel dégagé en permanence. Les inversions nocturnes persistent plusieurs heures après le lever du soleil, et le mélange vertical de la troposphère est faible. La saison est qualifiée de sèche et la pollution plus ou moins concentrée ne peut donc pas être évacuée par déposition humide.

- l'été, de juin à septembre, un ciel généralement nuageux empêche aux réactions photochimiques de développer leur potentiel. Les averses fréquentes éliminent bonne partie de la pollution par déposition humide. Les épisodes de pics sont moins fréquents.

Autres facteurs de pollution

La croissance urbaine démographique :
La population de la ZMVM est passée en l'espace de 60 ans (*1940 à 2000*) de 1.76 mio à 17.87 mio d'habitants (*INEGI [3], 2000*). La densité moyenne y est estimée à ~ 12'000 hab/km^2, ce qui est une des plus grandes valeur de densité au monde, excepté pour l'Asie. L'incertitude de ces chiffres est principalement liée au « irregular settlement ».

Un telle dynamique est attribuée à un taux important d'immigration des Etats voisins en raison de l'augmentation de la quantité d'emplois dans l'agglomération grandissante, combiné avec une fertilité moyenne à élevée de la population.

La croissance urbaine physique :

La surface urbanisée de la ZMVM est passée de 118 km^2 en 1940 à plus de 1'500 km^2 en 2000 (*Ward, 1998*). La rapide expansion physique de l'agglomération, une augmentation des distances entre quartiers résidentiels et zones industrielles et services, a provoqué un impact direct sur la désorganisation spatiale et sociale de la cité. Cette désorganisation s'est reportée sur le secteur des transports. En effet, la rapide inefficience globale des transports publics de masse (*métropolitain et trolley bus*) face à la soudaine augmentation des distances a provoqué une augmentation importante de la flotte de transport du secteur informel, à savoir les « *colectivos* ». L'utilisation de la voiture individuelle s'est également fortement répandue, ce qui a finalement participé à une augmentation incontrôlée des émissions. Aussi la planification urbaine est un outil développé au début des années 90. Elle n'a donc joué aucun rôle dans la structuration de la croissance urbaine.

La croissance urbaine économique :

L'augmentation du pouvoir d'achat de la population (*voir augmentation du Produit Intérieur Brut, ou Gros Domestic Product – GDP au tableau 3.1*) a provoqué une augmentation de la consommation. Le Mexique étant un pays producteur de pétrole (*PEMEX [4]*), 91 à 92% de la demande énergétique interne est satisfaite par les combustibles fossiles (*SENER [5], 2000*).

[3] *Instituto Nacional de Estadística, Geografía y Informàtica*
[4] *Petróleos Mexicanos*
[5] *Secretaría de Energía*

Tableau 3.1 : Croissance économique mexicaine, INEGI 1999

Period	GDP	GDP per capita
1950 - 1954	4.3	1.8
1954 - 1958	4.4	1.9
1960 - 1964	5.4	2.7
1965 - 1969	4.9	2.2
1970 - 1974	5.1	2.4
1975 - 1979	5.2	2.5
1980 - 1984	1.5	-0.1
1985 - 1989	0.8	-0.8
1990 - 1994	2.8	0.9
1995 - 1999	4.0	2.5

Effets de l'ozone troposphérique sur la santé et exposition

Tandis que les effets sur la santé des particules sont relativement bien étudiés (*maladies cardiovasculaires, pneumonies, affections chroniques telles que bronchites, obstructions diverses, asthme, etc.*), les impacts sur la santé humaine de la pollution par l'ozone restent plus incertains, en termes de toxicité chronique ou de toxicité aigüe (*indices, seuils, etc.*). Les affections liées à la présence d'ozone semblent toutefois être moins importantes que celles pour les particules. Les effets immédiats sur la santé humaine jusqu'à présent observés à la suite font appel à la notion de toxicité aigüe. Les répercussions observées sont une irritation des muqueuses du nez, des yeux et de la gorge, des douleurs respiratoires, sensations d'oppression et toux ainsi qu'une diminution de la fonction pulmonaire. Ces symptômes sembleraient se multiplier avec l'intensité et la durée de l'exposition. L'ozone attaque les cellules animales et végétales.

Pour ce qui est de la notion d'exposition, une des caractéristiques importantes dont il faut tenir compte est que l'ozone troposphérique est un polluant secondaire, issu d'une réaction photochimique fortement non linéaire entre les composés organiques volatils (*VOC*) et les oxydes d'azote (*NOx*) (*voir figures 3.3 et 3.4*). Les réactions simplifiées de la formation son les suivantes :

$$NO_2 + O_2 \leftrightarrow O_3 + NO$$
$$NO + VOC \leftrightarrow NO_2 + VOC'$$

Où les VOC servent de carburant à la réaction et où NO recycle NO_2 pour former O3.

Il est donc difficile de mettre en relation émissions et présence d'ozone en vue d'identifier géographiquement les concentrations moyennes auxquelles la population est exposée. Cette relation est encore plus difficile à définir si l'on considère que la persistance de l'ozone en zone urbaine est directement influencée par la présence de fortes concentrations de NOx. En effet, de jour on assiste à une décomposition chimique des molécules selon la réaction suivante :

$$NO + O_3 \leftrightarrow NO_2 + O_2$$

Lorsque la concentration en monoxyde d'azote augmente fortement, O_3 est très vite dissocié.

Ces réactions expliquent pourquoi l'ozone troposphérique se situe plus favorablement dans des zones suburbaines ou rurales. En ZMVM, et malgré la forte variabilité spatiale du polluant, il semblerait que les principaux pics d'ozone reportés soient localisés dans le Sud-Ouest de la ville, où se situent les principaux quartiers résidentiels. Cette distribution serait inverse à celle des particules, polluants primaires dont les principales émissions se situent dans les zones industrielles au Nord de la zone métropolitaine.
Il semblerait aussi que la ZMVM soit plus sensible aux variations de NOx (*voir figure 3.4, point C*).

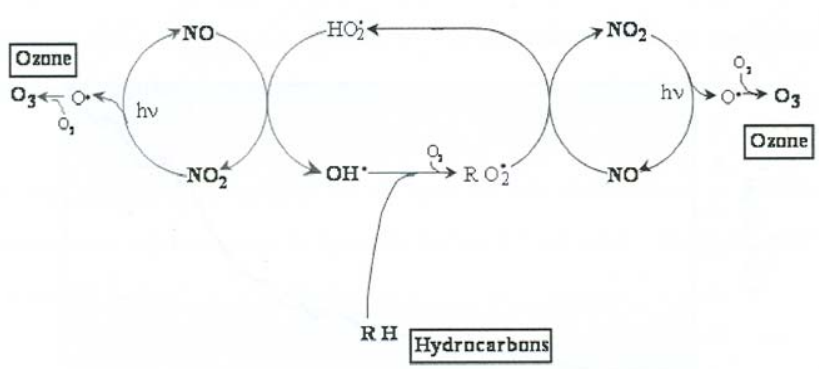


Figure 3.3: mécanisme simplifié de la formation d'ozone

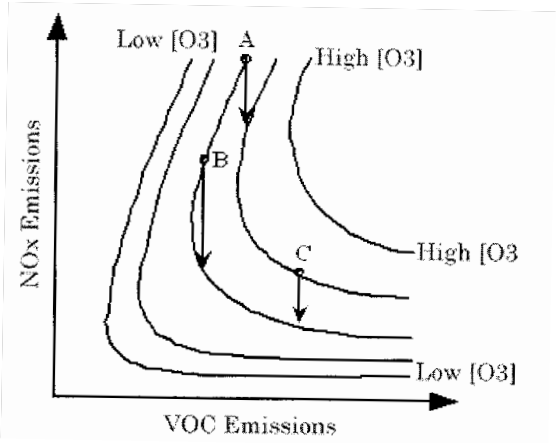


Figure 3.4: isoplètes d'ozone en fonction des émissions de NOx et VOC, illustration de la non linéarité de la réaction

Finalement, toujours en termes d'exposition, l'ozone est un composé fortement oxydant, réagissant directement avec tous types de surfaces. Les individus ne seraient donc affectés que lorsqu'ils se trouvent à l'air libre.

Sources de pollution: l'inventaire des émissions

L'inventaire des émissions est un instrument de gestion conçu par les instituts fédéraux de recherche pour la ZMVM. Il permet de connaître le volume ainsi que le type de contaminant émis par chacun des secteur s (*voire sous-secteur*) de production afin d'identifier des mesures d'amélioration prioritaires. Cet outil est destiné aux autorités chargées de la gestion de la qualité de l'air ainsi qu'aux autorités gouvernementales en général et à la population. Il constitue en outre une des bases du modèle de simulation de la qualité de l'air développé et utilisé par un groupe de travail formé entre l'INE[6] et le DDF[7], le « *Multiscale Climate Chemistry Model* » (*MCCM*). Ce groupe est entre autre responsable de la collecte et du traitement de données issu du réseau de monitorage RAMA décrit plus loin. Le dernier inventaire paru est le « *Inventario de Emisiones a la Atmósfera, Zona Metropolitana de la Valle de México, Secretaría del Medio Ambiente, 2000* ».

[6] *Instituto Nacional de Ecología*
[7] *Departamento del Distrito Federal*

Suite aux défauts de méthodologie qui ont en partie rendu les résultats des années précédentes plus qu'incertains, cet inventaire, de type « *bottom-up* [8] », a été établi sur la base des recommendations de Dr. Mario Molina Pasquel et son groupe de recherche ainsi que par la compagnie Eastern Research Group Inc.[9], et se veut plus juste. Un résumé des principaux résultats est présenté aux figures 3.5 et 3.6.

Sector	Emisiones [ton /año]								
	PM_{10}	$PM_{2.5}$	SO_2	CO	NOx	COT	CH_4	COV	NH_3
Fuentes puntuales	2,809	572	10,288	10,004	24,717	22,794	181	22,010	216
Fuentes de área	509	492	45	6,633	10,636	418,586	168,549	197,803	12,969
Fuentes móviles	5,287	4,589	4,348	2,018,788	157,239	210,816	11,593	194,517	2,261
Vegetación y suelos	1,736	380	N/A	N/A	859	15,425	N/A	15,425	N/A
Total	**10,341**	**6,033**	**14,681**	**2,035,425**	**193,451**	**667,621**	**180,323**	**429,755**	**15,446**

Figure 3.5: émissions par type de source, extrait de « Inventario de Emissiones a la Atmosféra, Zona Metropolitana del Valle de México, 2000

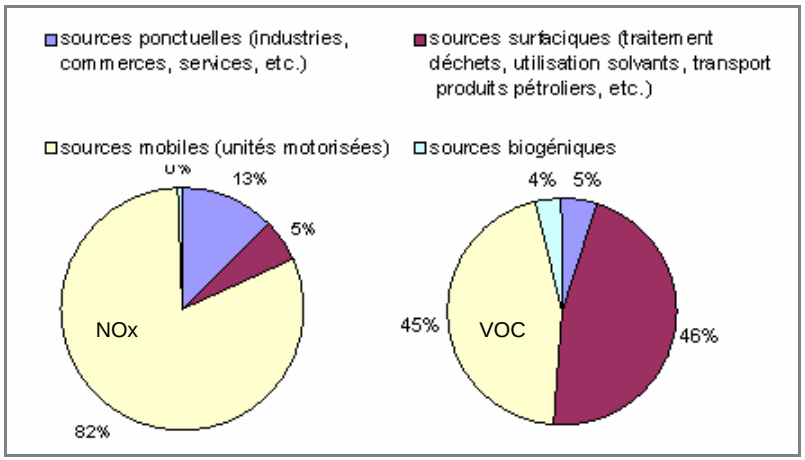


Figure 3.6 : répartition des émissions par type de source

[8] Un inventaire de type « bottom-up » évalue les émissions totales annuelles sur la base d'une estimation du nombre d'entreprises actives en ZMVM, multipliée par un facteur d'émission correspondant à une émission unitaire par secteur d'activité. Il apparaît dans ce type de démarche pour une ville comme celle-ci que le recencement des différentes untiés (industrielles, services, commerces, nombre et type de véhicules, etc.) est extrêmement fastidieux et source d'incertitudes.

[9] Análisis y Diagnóstico del Inventario de Emisiones de la Zona Metropolitana del Valle de México. M.J. Molina, L.T Molina, G. Sosa, J. Gasca y J. West. Instituto Tecnológico de Massachusetts. Agosto 2000

Les chiffres ci-dessus montrent que le secteur des transports (*sources mobiles*) est responsable pour plus de 82% des émissions de NOx en 2000, ainsi que pour 45% des émissions de VOC. Celui ci est également responsable pour plus de 20% des émissions de SO_2 et 35% des émissions de PM10, ce qui justifie des stratégies de gestion qui portent sur la réduction des émissions du secteur des transports.

Le cas particulier du secteur des transports

Les émissions à l'origine de la formation de l'ozone (*NOx et VOC*) sont en premier lieu dues au secteur des transports, auquel on associe donc généralement les programmes et mesures de réduction des émissions. La gestion des transports est actuellement complexe, car très peu prise en compte dans la rapide croissance urbaine des dernières décennies (physique, démographique et économique). Les différentes conséquences en relation avec l'augmentation des émissions sont les suivantes :

La détérioration du système de transports publics :

Les transports public à *«haute occupation »*[10] se sont fortement déteriorés dans les années 90. Le réseau du métropolitain, habituellement fortement fréquenté, s'est spatialement très peu développé. Les différentes raisons politiques et économiques de cet échec ont encouragé d'une part une explosion du *« secteur informel »* des transports collectifs, a savoir les *« colectivos »* (*voir figure 3.7*), et d'autre part l'aquisition de véhicules privés. Ces transports étant de type moyenne à faible occupation (*COMETRAVI [11] estime l' occupation des véhicules à 1.5 individu en moyenne en 1999*) le nombre de kilomètres parcourus journalièrement dans la zone métropolitaine a fortement augmenté. Le nombre de trajets serait passé de env. 20 mio en 1988 à plus de 35 mio en 1996 (*COMETRAVI, 1999*).

[10] « haute-occupation» désigne les transports permettant l'acheminement d'un grand nombre de personne pour un faible kilométrage parcouru par le véhicule.

[11] Comisión Metropolitana de Transporte y Vidalidad

Hormis l'augmentation du kilométrage parcouru, les autres types de problèmes posés par le système de *« colectivos »* comprennent :

- Un manque d'efficacité général : il n'existe pas « *d'organe de gestion* » des différentes companies privées ou associations de propriétaires de minibus, ce qui provoque chroniquement un manque de synchronisation dans le temps des « *colectivos* », un excès de véhicules durant les heures creuses, etc.
- Une forte compétitivité entre les différents propriétaires de minibus : difficultés à faire face aux opérations de maintenance, réparations et remplacements de pièces (*ou de véhicule !*), inobservance des règles de la route, temps d'attente long au terminus afin de remplir les véhicules au maximum, augmentation des tarifs, etc.

En plus de participer à une augmentation des émissions, ces problèmes seraient également responsables d'une augmentation des accidents de la route.

Figure 3.7: pesero de la ZMVM

Les taxis :

Le nombre de taxis enregistrés est estimé à ~ 69'000 (*COMETRAVI, 1999*), dont seulement 8'000 env. font partie de companies rescensées. Les incertitudes de ces chiffres (*dont les valeurs varient en fonction des sources*) proviennent de la procédure d'enregistrement des véhicules. Le problème est celui de l'âge moyen des véhicules, avec lequel les émissions ont tendance à augmenter.

Le transport de marchandises :

La ZMVM a une grande importance dans l'activité économique du pays. Mexico D.F. concentre plus d'1/5 du PIB (*INEGI, 1999*). Elle compte bon nombre de grandes industries et de commerces (*voir figure 3.8*), ce qui accentue le transport de marchandises (*intra et inter-urbain*).

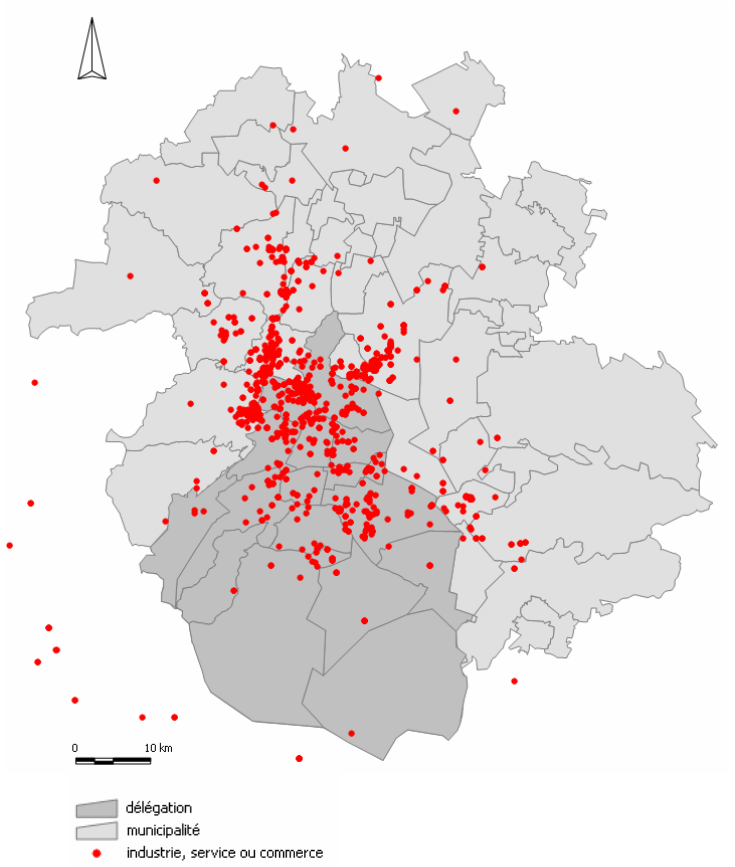


Figure 3.8 : localisation des installations industrielles, commerciales et services

La densité du réseau routier et la fréquence des congestions contribuent à l'inefficacité du transport de marchandises.

Bases légales environnementales et gestion de la qualité de l'air en ZMVM

Le principal texte de loi environnementale est le *«Ley General del Equilibrio Ecologico y Protección al Ambiente, 1988 »*. Les principales structures légales environnementales sont présentées à la figure 3.9.

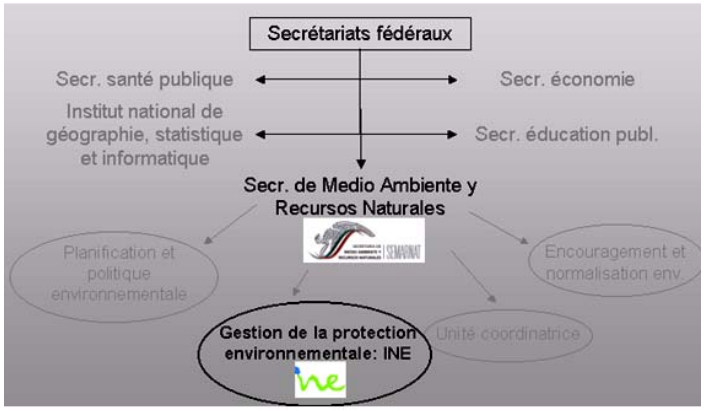


Figure 3.9 : ministères et sous sectrétariats de l'Environnement

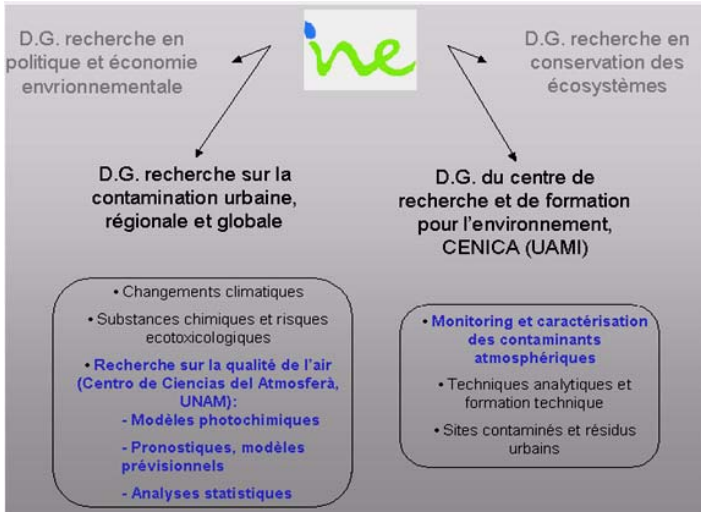


Figure 3.10 : structure et fonctionnement INE

Les ministères (*ou secrétariats fédéraux*) dirigent les différents secteurs politiques. Ils regroupent les sous-secrétariats. Pour l'Environnement, l'Institut National pour l'Ecologie (*INE*) est l'un des organes principaux (*voir figure 3.10*).
Les normes légales dans le domaine de la qualité de l'air sont établies pour tous les contaminants ayant un effet sur la santé :

Tableau 3.2 : Normes mexicaines officielles de qualité de l'air (Secretaría de Salud, 1994)

	Exposition aigue		***Exposition chronique***
Com-posé	***Concentration moyenne et intervalle de temps***	***Fréquence maximum acceptable***	
O_3	*0.11 ppm (1h)*	*1 fois/3 ans*	
SO_2	*0.13 ppm (24h)*	*1 fois/ an*	*0.03 ppm (moy. an.)*
NO_2	*0.21 ppm (1h)*	*1 fois/ an*	---
CO	*11 ppm (8h)*	*1 fois/ an*	---
PM_{10}	*150 g/m^3 (24h)*	*1 fois/ an*	*50 g/m^3 (moy. an.)*
Pb	---	---	*1.5 g/m^3 (moy. s/3*
TSP	*260 g/m^3 (24h)*	*1 fois/ an*	*75 g/m^3 (moy. an.)*

Ces normes sont édictées par le ministère de la santé (*Secretaría de Salud, 1994*), à la suite d'études épidémiologiques, de toxicité. Elles tiennent plus particulièrement compte de la tranche vulnérable de la population (*enfants et personnes âgées*).

D'un point de vue légal, les différentes stratégies et programmes de gestion ont commencé à être mis en place dans les années 90. Ils s'intéressent tout particulièrement au secteur des transports (*réduction des émissions par le « Hoy No Circula, 1989 », amélioration de la qualité des carburants, notamment élimination du plomb et réduction de la teneur en soufre dans l'essence, introduction du catalyseur et de carburants alternatifs tels que le Liquid Petroleum Gas, le Compressed Natural Gas, etc., inspection de l'état de service des véhicules, amélioration du service des transports publics*). Deux des programmes les plus importants sont *PICCA*[12] et *PROAIRE*[13]. Le suivi interannuel de la pollution étant réalisé par la représentation du nombre de jours

[12] *Program Integral contra la Contaminación del Aire, 1990-1995*
[13] *Progama Para Mejorar la Calidad del Aire en el Valle de México, 1995-2000*

pour lesquels les normes du tableau 3.2 sont dépassées, le programme PROAIRE aura pour principal objectif de réduire les pics d'ozone mesurés, en amplitude et nombre.

En 2000, la qualité de l'air de la ZMVM dépasse encore pour le 80% des jours les normes officielles mexicaines de la qualité de l'air pour l'ozone.

Mesure de la qualité de l'air : Red Autmomàtica de Monitoreo Atmosférico

Le réseau de mesures RAMA est opérationnel depuis 1986, en partie grâce à l'assistance technique de l'EPA[14]. Il est constitué de 37 stations au total, dont une partie mesure les polluants atmosphériques ordinaires et l'autre les paramètres météorologiques ordinaires (*humidité relative, température, intensité et direction du vent au sol*).

19 stations au total, dont la distribution spatiale est représentée à la figure 3.11 mesurent les concentrations horaires en ozone. La technique utilisée est la photometrie UV. Le nombre ainsi que la localisation des stations s'accorde aux critères édictés par l'OMS. Cependant, aucune des stations ne se situe dans une zone rurale du bassin atmosphérique.

Les mesures brutes sont ensuite automatiquement envoyées vers un centre de traitements de données, où elles sont contrôlées et validées. Elles sont finalement diffusées librement sur le web (http://www.sima.com.mx/sima/df/index.html) sous formes de tables de concentrations horaires par station. L'indice IMECA mentionné précédemment (*voir p. 6, « Evolution historique de la pollution »*) permet la diffusion au public des données de concentrations. Il est calculé sur la base des concentrations validées, selon un algorithme dont le but est l'obtention d'une valeur de 100 IMECA égale à la valeur admise par la norme, ceci pour n'importe lequel des polluants. Le système (*réseau + validation*) est opéré par CENICA[15], un institut de recherche fédéral faisant partie de l'INE, ainsi que par des scientifiques pour le gouvernement du DF, responsables de la mise à jour d'un site internet, le SIMAT[16], permettant la diffusion de l'IMECA à l'intention du grand public et des autorités.
Un programme d'entretien et de calibrage des instruments est observé. Les stations et procédures de mesures sont aussi certfiées chaque année par l'EPA. La fiabilité de l'appareillage est donc bonne, et ce depuis la mise en service du réseau, et le degré d'imprécision des mesures pour l'ozone est de 3%.

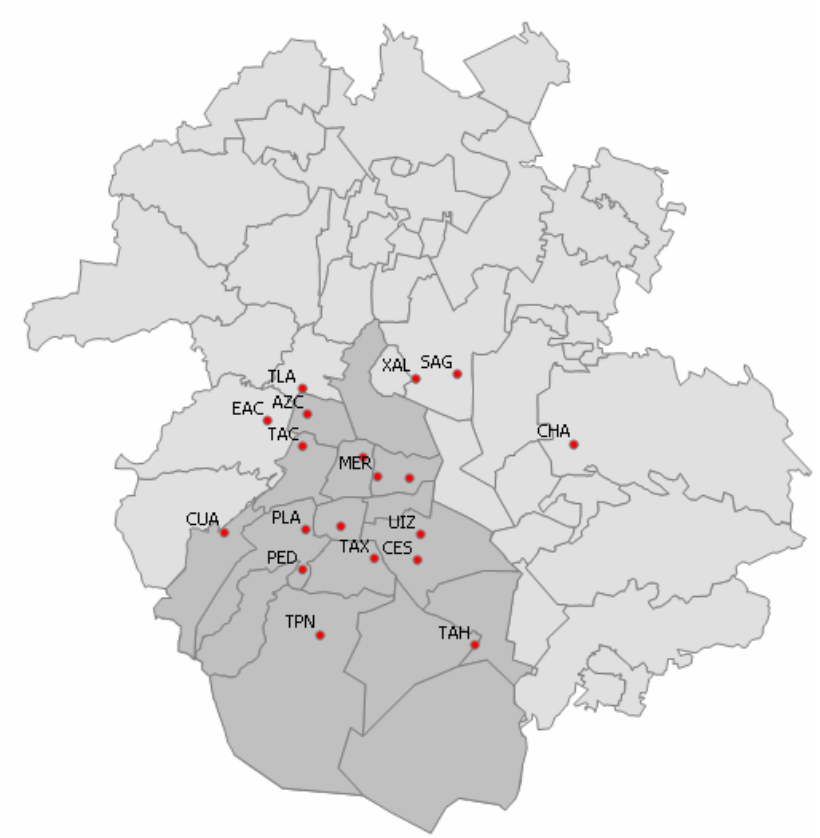


Figure 3.11 : distribution spatiale des stations de mesure de RAMA

Suivi de l'évolution de la pollution atmosphérique et prédiction

Le DF essaie d'étudier les tendances à long terme de la qualité de l'air afin de pouvoir évaluer l'impact des mesures de réduction entreprises. Le suivi est notamment réalisé à l'aide de l'utilisation d'un modèle de qualité de l'air, le *« Multiscale Climate Chemistry Model » (MCCM)*, choisi pour sa capacité à reproduire les caractéristiques topographiques du bassin atmosphérique. Les paramètres météorologiques sont obtenus par un couplage avec un modèle météorologique à méso-échelle (*MM5).*

[14] *Environmental Protection Agency, USA*
[15] *Centro National de Investigación y Capacitación Ambiental (INE)*
[16] *Sistema de Monitoreo Atmosférico (Gobierno del Distrito Federal)*, *http://www.sma.df.gob.mx/simat/*

Ce type de modélisation est également utilisé pour tenter de réaliser des prévisions à court terme (*3 à 5 jours*). Des étude sont actuellement en cours d'étude au CCA[17].

Les émissions sont tirées de l'inventaire des émissions publié par le SMA. Les différentes données spatiales (*sources ponctuelles, surfaciques et linéaires*) sont produites par l'INEGI.

L'utilisation des Systèmes d'Information Géographique reste marginale, mais semble pourvue d'un fort intérêt.

Conclusions

Les connaissances dans le domaine de la qualité de l'air pour la ZMVM réunies au cours de ces dernières années sont nombreuses. Les questions posées par la recherche sont donc toujours plus précises. Les affirmations suivantes sont à retenir dans le cas de l'ozone :

- l'environnement géographique de la ZMVM est un facteur déterminant : bassin atmosphérique, inversions thermiques, altitude et système anticyclonique sont les caractéristiques de celui-ci.
- Les inversions thermiques sont fortes le matin et disparaissent progressivement l'après-midi. Elles définissent l'évolution du profil vertical des concentration en ozone.
- Le secteur des transports est la principale source d'émissions menant à la formation d'ozone.
- La ZMVM émet d'énormes quantités de VOC. Le rapport entre VOC et NOx semblerait prouver que la cité est plus sensible aux modifications des émissions de NOx.
- Les effets sur la santé de l'ozone sont encore mal connus, les indices d'exposition difficiles à définir, et par conséquent l'exposition en fonction du lieu également.

Finalement une question d'un intérêt particulier peut être soulevée : quel serait le comportement de polluants tels que l'ozone, dans le cas ou la ZMVM et les localités environnantes extérieures au bassin se rejoignaient ? Et quelles en seraient donc les conséquences sur l'environnement pour Puebla, Cuernavaca, Toluca, Querétaro, comprenant zones agricoles et forêts, sachant qu'en plus d'atteindre une plus grande portion de la population mexicaine, l'ozone serait également responsable de la diminution des rendements des cultures, et de la biodiversité de manière plus générale ?

[17] *Centro de Ciencias del Atmosferà, UNAM, INE .*

4. Données pour la ZMVM

4.1. Jeu de données de concentrations en ozone

Le jeu de données de concentrations en ozone présenté ci-dessous est l'objet principal de ce travail.

Le RAMA fournit des données horaires validées de concentrations en ozone pour les 19 stations ci-dessous.

Tableau 4.1 : énumération des stations de mesure

N° station	***Nom station***	***Abréviation***
1.	*Lagunilla*	*LAG*
2.	*Tacuba*	*TAC*
3.	*ENEP Acatlán*	*EAC*
4.	*San Agùstin*	*SAG*
5.	*Azcapotzalco*	*AZC*
6.	*Tlalnepantla*	*TLA*
7.	*Xalostoc*	*XAL*
8.	*Merced*	*MER*
9.	*Pedregal*	*PED*
10.	*Cerro de la Estrella*	*CES*
11.	*Plateros*	*PLA*
12.	*Hangares*	*HAN*
13.	*UAM Iztapalapa*	*UIZ*
14.	*Benito Juárez*	*BJU*
15.	*Taxqueña*	*TAX*
16.	*Cuajimalpa*	*CUA*
17.	*Tlalpan*	*TPN*
18.	*Chapingo*	*CHA*
19.	*Tláhuac*	*TAH*

L'emplacement des stations est représenté à la figure 3.11.
L'unité de mesure est le [ppm] (*partie par million*).
L'incertitude de la mesure est de l'ordre de 3%.
Une description plus générale du réseau est présentée dans le chapitre précédent.

La série temporelle utilisée dans ce travail a été choisie par E. Gaillard. Elle comprend les concentrations du début du mois de juillet 1992 à la fin du mois de décembre 2002, et constitue l'intervalle le plus long et le plus complet possible.

2	FECHA	HORA	LAG	TAC	EAC	SAG	AZC	TLA	XAL	MER	PED	CES
3	01.01.1996	1	0.024	0.012	0.036	0.015	0.034	0.034	0.008	0.015	0.007	0.018
4	01.01.1996	2	0.022	0.014	0.029	0.017	0.029	0.026	0.008	0.019	0.005	0.016
5	01.01.1996	3	0.017	0.019	0.031	0.02	0.031	0.029	0.007	0.01	0.006	0.015
6	01.01.1996	4	0.017	0.02	0.026	0.023	0.034	0.019	0.007	0.009	0.005	0.016
7	01.01.1996	5	0.017	0.027	0.026	0.018	0.029	0.018	0.007	0.009	0.005	0.017
8	01.01.1996	6	0.016	0.029	0.026	0.023	0.024	0.017	0.009	0.008	0.006	0.016
9	01.01.1996	7	0.015	0.011	0.02	0.023	0.023	0.014	0.018	0.012	0.018	0.013
10	01.01.1996	8	0.021	0.014	0.017	0.016	0.024	0.019	0.013	0.011	0.011	0.012
11	01.01.1996	9	0.027	0.021	0.024	0.019	0.028	0.016	0.017	0.02	0.018	0.014
12	01.01.1996	10	0.048	0.044	0.044	0.046	0.043	0.042	0.033	0.03	0.039	0.019
13	01.01.1996	11	0.06	0.064	0.047	0.065	0.059	0.061	0.073	0.049	0.058	0.02
14	01.01.1996	12	0.075	0.058	0.046	0.044	0.047	0.055	0.055	0.057	0.064	0.026
15	01.01.1996	13	0.08	0.048	0.046	0.049	0.047	0.048	0.056	0.069	0.064	0.028
16	01.01.1996	14	0.058	0.047	0.044	0.048	0.045	0.049	0.066	0.065	0.05	0.035
17	01.01.1996	15	0.052	0.049	0.047	0.072	0.048	0.052	0.051	0.049	0.044	0.039
18	01.01.1996	16	0.048	0.045	0.042	0.047	0.039	0.045	0.043	0.032	0.042	0.037
19	01.01.1996	17	0.045	0.043	0.038	0.042	0.036	0.043	0.039	0.028	0.039	0.025
20	01.01.1996	18	0.034	0.037	0.035	0.028	0.034	0.037	0.027	0.023	0.035	0.025
21	01.01.1996	19	0.02	0.023	0.03	0.02	0.031	0.03	0.015	0.016	0.028	0.025
22	01.01.1996	20	0.013	0.014	0.027	0.014	0.026	0.027	0.012	0.012	0.019	0.019
23	01.01.1996	21	0.017	0.012	0.028	0.014	0.023	0.015	0.014	0.014	0.012	0.017
24	01.01.1996	22	0.024	0.013	0.017	0.014	0.023	0.017	0.011	0.018	0.011	0.015
25	01.01.1996	23	0.021	0.012	0.011	0.018	0.024	0.013	0.013	0.012	0.005	0.015
26	01.01.1996	24	0.023	0.013	0.028	0.016	0.02	0.01	0.013	0.015	0.004	0.016
27	02.01.1996	1	0.022	0.01	0.016	0.015	0.021	0.01	0.004	0.016	0.004	0.02
28	02.01.1996	2	0.021	0.009	0.014	0.021	0.022	0.007	0.003	0.009	0.004	0.015

Figure 4.1: extrait de la série horaire de concentrations pour 1996, http://www.sma.df.gob.mx/simat/

4.2. Données météorologiques

La météorologie est l'un des principaux paramètres qui influence la formation de l'ozone. RAMA mesure les variables météorologiques ordinaires (*température, humidité relative, intensité et direction du vent au sol*) à intervalle horaire pour dix points de la ZMVM. Ces stations font partie du réseau de mesure des concentrations d'ozone (*voir figure 4.2*).

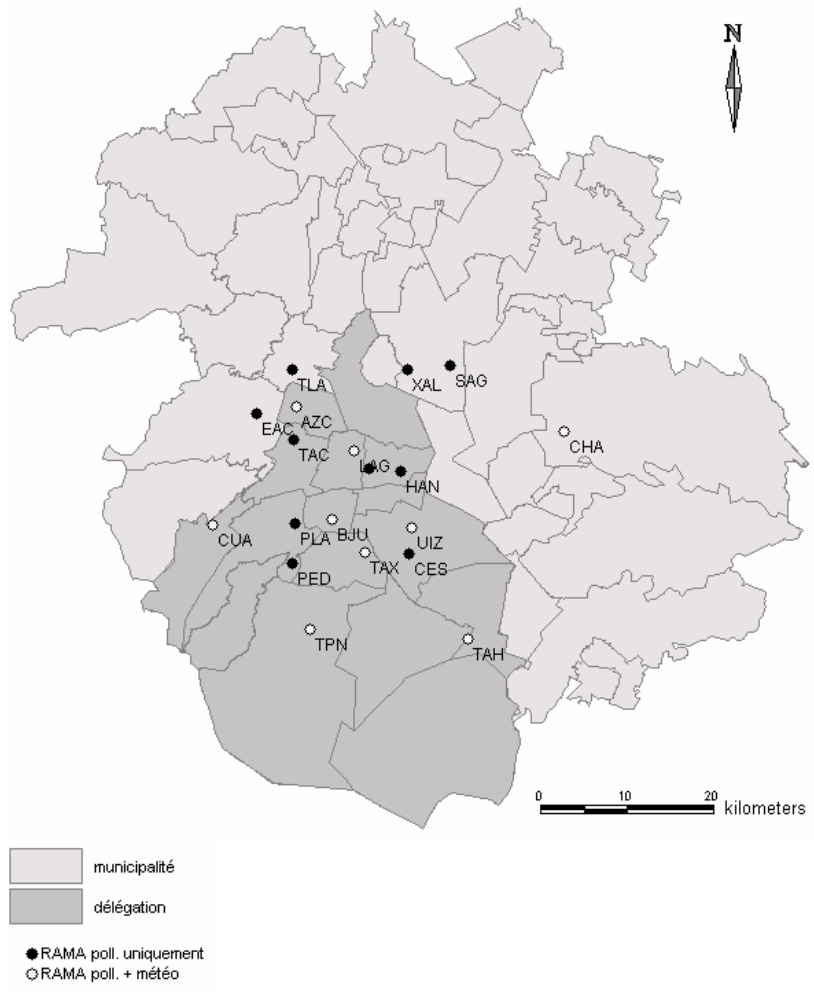


Figure 4.2 : emplacement des stations météorologiques

Ces données sont disponibles pour l'intervalle de temps sélectionné pour l'ozone. Elles pourraient s'avérer utiles lors de l'exploration des mesures de concentration, ou lors de l'interprétation des résultats. Elles restent cependant secondaires pour ce travail et ne seront pas décrites plus en détail.

4.3. Données territoriales

Les données realtives au territoire sont produites par l'INEGI sous la forme de couches vecteur. Les éléments sélectionnés pour ce travail sont :

- délimitations territoriales des délégations et municipalités de la zone métropolitaine
- stations de mesure RAMA
- relief
- réseaux routier et ferroviaire
- principales inustries
- espaces verts urbains
- surfaces agricoles
- plans d'eau

Le système de projection utilisé pour la ZMVM est Univeral Transverse Mercator, Hémisphère Nord, zone 14.

4.4. Système Municipal de Base de données (SIMBAD)

L'Institut National pour la Statistique, la Géographie et l'Informatique permet un libre accès à un recueil de statistiques nationales, établies par municipalité pour chacun des Etats du Mexique[18]. On y trouve principalement des données relatives à la géographie humaine telles que l'indice de développement humain (*IDH*), le recencement de la population selon les différents groupes sociaux (*âge, sexe, etc.*), etc.

Ce type d'information qualitative pourrait être utilisé d'un point de vue contextuel.

4.5. Inventaire des émissions 2000 (*voir chapitre 3*)

[18] http://sc.inegi.gob.mx/simbad/

5. Traitements

5.1 Analyse exploratoire

La description statistique de données ne peut se faire que par l'intermédiaire d'une analyse exploratoire complète d'un point de vue temporel, spatial et multivarié.

L'exploration du jeu de données de concentration en ozone (*décrit en 4. « Données »*) se fait sur les séries brutes télédéchargées (*voir figure 5.1.1*).

2	FECHA	HORA	LAG	TAC	EAC	SAG	AZC	TLA	XAL	MER	PED	CES
3	01.01.1996	1	0.024	0.012	0.036	0.015	0.034	0.034	0.008	0.015	0.007	0.018
4	01.01.1996	2	0.022	0.014	0.029	0.017	0.029	0.026	0.008	0.019	0.005	0.016
5	01.01.1996	3	0.017	0.019	0.031	0.02	0.031	0.029	0.007	0.01	0.006	0.015
6	01.01.1996	4	0.017	0.02	0.026	0.023	0.034	0.019	0.007	0.009	0.005	0.016
7	01.01.1996	5	0.017	0.027	0.026	0.018	0.029	0.018	0.007	0.009	0.005	0.017
8	01.01.1996	6	0.016	0.029	0.026	0.023	0.024	0.017	0.009	0.008	0.006	0.016
9	01.01.1996	7	0.015	0.011	0.02	0.023	0.023	0.014	0.018	0.012	0.018	0.013
10	01.01.1996	8	0.021	0.014	0.017	0.016	0.024	0.019	0.013	0.011	0.011	0.012
11	01.01.1996	9	0.027	0.021	0.024	0.019	0.028	0.016	0.017	0.02	0.018	0.014
12	01.01.1996	10	0.048	0.044	0.044	0.046	0.043	0.042	0.033	0.03	0.039	0.019
13	01.01.1996	11	0.06	0.064	0.047	0.065	0.059	0.061	0.073	0.049	0.058	0.02
14	01.01.1996	12	0.075	0.058	0.046	0.044	0.047	0.055	0.055	0.057	0.064	0.026
15	01.01.1996	13	0.08	0.048	0.046	0.049	0.047	0.048	0.056	0.069	0.064	0.028
16	01.01.1996	14	0.058	0.047	0.044	0.048	0.045	0.049	0.066	0.065	0.05	0.035
17	01.01.1996	15	0.052	0.049	0.047	0.072	0.048	0.052	0.051	0.049	0.044	0.039
18	01.01.1996	16	0.048	0.045	0.042	0.047	0.039	0.045	0.043	0.032	0.042	0.037
19	01.01.1996	17	0.045	0.043	0.038	0.042	0.036	0.043	0.039	0.028	0.039	0.025
20	01.01.1996	18	0.034	0.037	0.035	0.028	0.034	0.037	0.027	0.023	0.035	0.025
21	01.01.1996	19	0.02	0.023	0.03	0.02	0.031	0.03	0.015	0.016	0.028	0.025
22	01.01.1996	20	0.013	0.014	0.027	0.014	0.026	0.027	0.012	0.012	0.019	0.019
23	01.01.1996	21	0.017	0.012	0.028	0.014	0.023	0.015	0.014	0.014	0.012	0.017
24	01.01.1996	22	0.024	0.013	0.017	0.014	0.023	0.017	0.011	0.018	0.011	0.015
25	01.01.1996	23	0.021	0.012	0.011	0.018	0.024	0.013	0.013	0.012	0.005	0.015
26	01.01.1996	24	0.023	0.013	0.028	0.016	0.02	0.01	0.013	0.015	0.004	0.016
27	02.01.1996	1	0.022	0.01	0.016	0.015	0.021	0.01	0.004	0.016	0.004	0.02
28	02.01.1996	2	0.021	0.009	0.014	0.021	0.022	0.007	0.003	0.009	0.004	0.015

Figure 5.1.1: extrait de série horaire de concentrations pour 1996, http://www.sma.df.gob.mx/simat/

5.1.1. Objectifs :

L'analyse exploratoire a pour but l'identification d'éventuelles structures dans le jeu de données de concentrations.

Cette analyse comporte trois sous-chapitres : le premier explore le jeu de données dans le temps. Il y est donc la seule variable capable de décrire la concentration. Le temps est en effet une variable primordiale pour l'explication de n'importe quel phénomène naturel. La prise en compte de cette variable est déterminée par la correspondance entre l'échelle de temps de la mesure, et celle de l'évolution du phénomène. Les deuxième et troisième chapitres décrivent la concentration comme un phénomène dont le comportment est multivarié, et où les variables explicatives sont les stations. En effet, dans le deuxième sous-chapitre il s'agit d'observer les relations de dépendance entre les différentes stations. Le troisième chapitre est une analyse exploratoire spatiale traditionnelle.

Ces étapes se succèdent pour une raison précise : l'exploration spatiale d'un jeu de données ne peut se faire que sur des valeurs « *simultanées* ». La représentation d'un phénomène n'a de sens que lorsque les valeurs représentées sont mesurées dans le même intervalle de temps. Le pas de l'intervalle est défini par la rapidité avec laquelle le phénomène évolue.

Ces différents chapitres comprennent les objectifs suivants :

- Développer une méthodologie permettant une analyse statistique exploratoire complète, temporelle, multivariée et spatiale, afin de mettre en évidence des comportements généraux et plus spécifiques de la concentration en ozone.

- Identifier d'éventuelles valeurs aberrantes.

- Compte tenu du volume de données continuellement produites par le RAMA, ainsi que des caractéristiques de ces données, développer une « *méthodologie exploratoire* », ainsi que des fonctions d'analyse spécifiques aux séries brutes, afin d'effectuer un traitement juste et rapide de telles séries.

- Etudier la corrélation entre les stations, ainsi que l'importance de chacune d'entre elles dans l'explication du phénomène étudié.

- Détecter des structures spatiales dans les données et régionaliser la concentration par l'application de méthodes géostatistiques d'interpolation.

Une autre intention implicite de la démarche exploratoire est la réduction de la quantité de données.

Les résultats escomptés de cette étude sont l'identification dans le temps d'épisodes[19] de pollution, la modélisation géostatistique de ces épisodes, et finalement la représentation cartographique des épisodes modélisés.

[19] *Un épisode est un intervalle de temps durant lequel les concentrations attribuées aux différents points de mesure de l'espace à représenter sont représentatives de cet intervalle.*

5.1.2. Méthodologie :

Ce paragraphe présente la démarche suivie pour répondre aux objectifs.

0. Prise en main des différents logiciels utilisés :

   MATLAB 7.0 et son extension « Statistical Toolbox »
   VarioWin 2.2
   ArcGIS 9.0 et ses extensions « Géostatistical Analyst » et « Spatial Analyst »
   Manifold 6.0

1. Exploration temporelle des concentrations (*voir § 5.1.3.1*):

   Dans quel intervalle horaire se situent les mesures (*hypothèse émise jusqu'à présent : 13-17h*).

   Quel est le type de distribution dans le temps des mesures (*gaussienne, asymétrie*), et par conséquent quel est l'estimateur à choisir pour décrire une journée de mesures, pour dans un second temps évaluer l'évolution annuelle des concentrations.

   Représentation de l'estimateur en fonction du temps : quelle est l'évolution temporelle (*séries annuelles*) des concentrations

   Identification des épisodes de pollution

2. Exploration multivariée des concentrations en fonction des stations (*voir § 5.1.3.2*):

   Etude de la variabilité des mesures par station sur la base de diagrammes boxplot

   Etablir une matrice de corrélations

   Analyse en Composantes Principales : identification des variables qui peuvent potentiellement offrir un intérêt à la description du phénomène

3. Exploration spatiale (*voir § 5.1.3.3*) :

   Description des unités d'observation

   Vérification de l'échantillonnage

   Vérification des conditions de stationnarité pour l'utilisation de méthodes géostatistiques d'interpolation

   Analyse structurale des données : variographie

4. Interpolation

5.1.3. Traitements et résultats :

5.1.3.1. Exploration temporelle

Le phénomène exploré est la concentration y, la variable est le temps t. La mesure y_i au temps t_i s'écrit donc :

$$y_i = f(t_i)$$

où i=(1,…,n) et n est le nombre de mesures dans la série annuelle.

Les différents traitements sont réalisés à l'aide du logiciel de calcul matriciel MATLAB 7®. Quelques fonctions supplémentaires ont été développées sur cette interface afin de traiter plus particulièrement les séries de données présentes (*voir annexe 2*). Les principales spécificités des données brutes sont les suivantes :

- Volume important de mesures : Sept séries annuelles (*1996 à 2002*) de concentrations horaires pour chacune des stations de mesures. Ces séries brutes se présentent donc sous forme de sept matrices de taille (8760×21) chacune.

- Séries comprenant la totalité des heures du jour (*1 à 24*). Les heures intéressantes pour le calcul d'un danger d'exposition se situent dans la journée. Les valeurs nocturnes sont attribuées à « *l'ozone de fond* », une pollution d'origine continentale.

- Séries parfois entachées de valeurs non définies *NaN* (*« Not a Number »*).

L'analyse d'une telle quantité de mesures ne peut que se faire que par le développement de codes. Notona également que la présence d'années bissextiles rendent souvent le développement et l'exécution des fonctions imaginées difficiles.

5.1.3.1.1. Intervalle horaire

Les sept ans de mesures de concentrations ont été cumulés en fonction de l'heure de la journée. Les différentes stations se distinguent ci-dessous par la couleur :

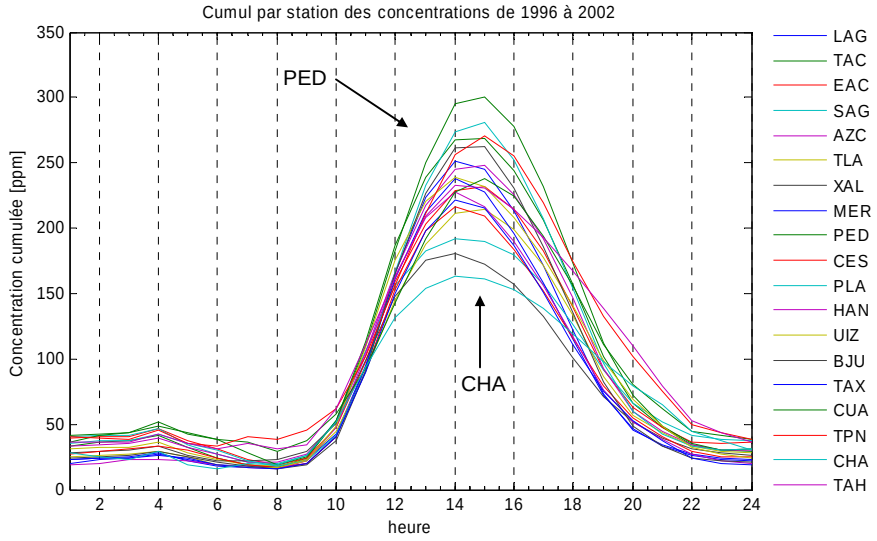


Figure 5.1.2: cumul par station des concentrations pour l'ensemble des séries horaires (1996-2002)

Le cumul permet ici de conserver toutes les mesures afin de déterminer dans quel intervalle horaire se situent la plus grande partie des mesures. Cette fonction est commutative et associative, ce qui signifie que d'effectuer la somme des jours ou la somme des heures représentée ci-dessus revient au même. L'inconvénient de cette représentation est du à la présence de valeurs de type *NaN*. Ce principe de codage est utilisé par le RAMA pour représenter les valeurs qui ne peuvent pas être validées. Ces valeurs sont le résultat d'erreurs de transmission des appareils, opérations de maintenance du réseau, etc. Le tableau 5.1.1 présente le nombre de valeurs *NaN* remplacées pour chaque série annuelle, afin d'identifier les années les plus défavorables. On peut y observer que l'année 2002 est plus particulièrement entachée de valeurs *NaN*. Elles ont pour effet de fausser tout types de calculs et seront éliminées dans la suite du travail (*notamment par les fonctions mentionnées en début de chapitre*).

Tableau 5.1.1 : Proportions de NaN dans les séries annuelles originales

Série annuelle	*Nb deNaN*	*Nb de mesures*	*Proportion*
ozone 96[20]	*51*	*166'896*	*0.3‰*
ozone 97	*-*	*166'440*	*-*
ozone 98	*-*	*166'440*	*-*
ozone 99	*-*	*166'440*	*-*
ozone 00	*-*	*166'896*	*-*
ozone 01	*1'558*	*166'440*	*9.3‰*
ozone 02	*10'266*	*166'440*	*6.2%*

Les courbes présentées à la figure 5.1.2 sont des courbes en cloche de type gaussienne. Toutes les stations présentent la même type évolution journalière, à savoir une production d'ozone qui s'amorce entre 8 et 10h, en correspondance avec l'augmentation des émissions et du rayonnement solaire[21]. Le maximum se situe aux environs de 15h. Cette figure permet également d'observer que la croissance est légèrement plus marquée que la décroissance, ce qui pourrait être expliqué par une nouvelle augmentation des émissions en fin de journée (*les transports sont la principale source conduisant à la formation d'ozone*) associée à une diminution du rayonnement solaire. Finalement, ces courbes montrent que l'évolution journalière du phénomène est très caractéristique et est indépendante de l'espace. Il est donc possible d'extraire les heures les plus intéressantes pour la suite de l'analyse, c'est-à-dire les heures qui décrivent les valeurs de maximum, et qui sont les mêmes pour toutes les stations. L'intervalle **12-18 heures** peut être retenu dans la suite de l'anaylse.

Cette figure permet également de « *classer* » les stations en fonction de l'amplitude du pic pour les sept années de mesures, ceci afin d'avoir une première idée des stations plus ou moins touchées (*voir tableau 5.1.2*).

Tableau 5.1.2 : classification des stations en fonction de l'amplitude du pic à 15h

Station	*Val. cumulée*	*Rang*	*Station*	*Val.cumulée*	*Rang*
PED	*300.21*	*1*	*TAH*	*230.89*	*11*
PLA	*280.84*	*2*	*MER*	*227.72*	*12*
TPN	*270.6*	*3*	*HAN*	*216.42*	*13*
TAC	*268.96*	*4*	*TLA*	*214.06*	*14*
BJU	*262.2*	*5*	*TAX*	*215.81*	*15*
AZC	*248.26*	*6*	*CES*	*208.9*	*16*
LAG	*244.78*	*7*	*SAG*	*190.21*	*17*
CUA	*238.12*	*8*	*XAL*	*172.62*	*18*
EAC	*231.91*	*9*	*CHA*	*161.68*	*19*
UIZ	*231.81*	*10*			

[20] *1996 et 2000 sont des années bissextiles*

[21] *Le rayonnement solaire et les émissions font partie des principaux facteurs de production*

Pedregal (*PED*) est la station qui subit les pics d'ozone les plus importants alors que Chapingo (*CHA*) n'est que peu touchée en comparaison (*les valeurs NaN n'ont que très peu d'influence sur ce calcul*). On remarque également que la valeur cumulée est extrêmement progressive entre les stations.

5.1.3.1.2. Type de distribution et estimateur :

Les deux statistiques simples retenues pour la description d'un jour de mesure sont la moyenne et la médiane. Le choix de l'estimateur doit se faire sur la base de l'étude de la distribution journalière des mesures. Au vu du grand nombre de données (*2557 jours au total*), l'analyse se fait sur 30 jours sélectionnés au hasard parmi les huit ans de mesures, de 12 à 18 heures. L'ensemble des résultats figure dans l'annexe 3.

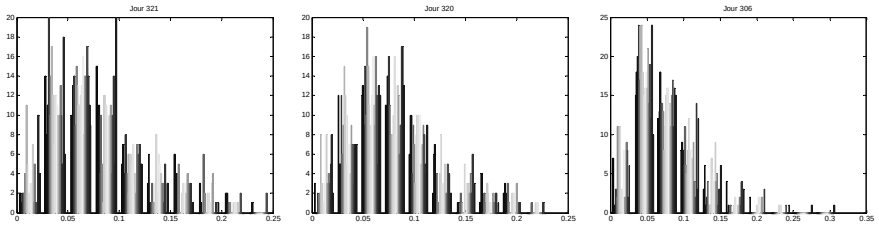


Figure 5.1.3 : extrait de la distribution des concentrations de 1996 à 2002 pour les jours sélectionnés 321, 320 et 306.

Les graphiques ci-dessus illustrent une distribution qui est asymétrique pour la plus grande partie des cas, elle n'est donc pas gaussienne. Une répartition normale devrait être à première vue caractérisée par une courbe symétrique autour de la moyenne, ne présentant pas de sous population. L'évidence fait qu'il n'est ici pas nécessaire de vérifier la corrélation des mesures avec un ajustement gaussien.

Les différences de tons de gris représentent les différentes stations, cette distinction n'apporte ici que peu d'information.

L'estimateur retenu pour décrire l'ensemble des mesures durant un jour est donc la médiane, accompagnée de la distance interquartiles. L'avantage de la médiane et des distances interquartiles réside également dans le fait qu'elles sont des statistiques plus robustes par rapport aux valeurs aberrantes. Rappelons cependant que les mesures diffusées par RAMA sont des mesures validées, ce qui élimine en grande partie ce type de valeurs.

5.1.3.1.3. Evolution temporelle de la médiane :

La figure 5.1.4 illustre l'évolution annuelle de la médiane journalière pour chaque station. Le calcul a été réalisé pour l'ensemble des séries annuelles (*voir annexe 4*).

Les points représentés sont des valeurs par station alors que le trait continu est une tendance pour l'ensemble des stations. Les statistiques utilisées sont la médiane (*bleu*) et la moyenne (*rouge*). La différence est négligable dans ce cas.
On peut y observer à une échelle annuelle que certaines années sont plus structurées que d'autres (*notamment 1998*). Ces séries mettent en évidence une sorte de cycle lors duquel le maximum des valeurs médianes est atteint entre les **90ème et 150ème jours** (*avril/mai, encardré*). Le minimum se situe globalement aux environs du 250ème jour (*août/sept).* Ces variations cycliques de concentrations pourraient certainement être expliquées par une étude approfondie des conditions météorologiques. Dans le contexte du travail, et compte tenu des données météorologiques à disposition, une première explication de la position des pics peut se faire à l'aide de variables telles que la température et l'humidité. En effet, celles-ci influencent fortement la formation d'ozone (*voir chap.1, « facteur géographique »*), et les mois de mars, avril, mai sont pour les années à disposition parmi les plus chauds et secs (*voir annexe 5*).

A une échelle interannuelle, on peut observer un minimum relativement constant, qui cependant semble affecté par des valeurs nulles en 2000, 2001 et 2002, ce qui parait étrange puisque l'intervalle horaire utilisé devrait correspondre clairement à des heures de production d'ozone. Le maximum est variable. Un calcul de la moyenne sur les mois de maximum (*avril et mai*) estime une concentration qui passe de l'ordre de 0.1 à 0.05 ppm (*voir figure 5.1.5*). Cette décroissance est relativement logique puisque des mesures importantes de contrôle des émissions ont été introduites dès le début des années '90. Ce phénomène suggère une distinction entre les différentes années par la suite.

Après observation détaillée des séries brutes, on peut observer que les valeurs nulles mentionnées ci-dessus concernent plus particulièrement les stations ENEP_Acatlán, Tlalnepantla, Cuajimalpa, San_Agùstin, et UAM Iztapalapa, notamment pour les années 2000 et 2002, où des séries de valeurs nulles en colonnes se succèdent pour plusieurs jours. Ce qui signifie à priori que les instruments de mesure devaient être soit hors d'usage, soit en révision. Il est cependant difficile à ce stade d'exclure un nombre aussi important de données parmi les matrices de données brutes (*voir figure 5.1.6*), et encore plus difficile d'évaluer l'impact de ces séries sur la suite de calculs.

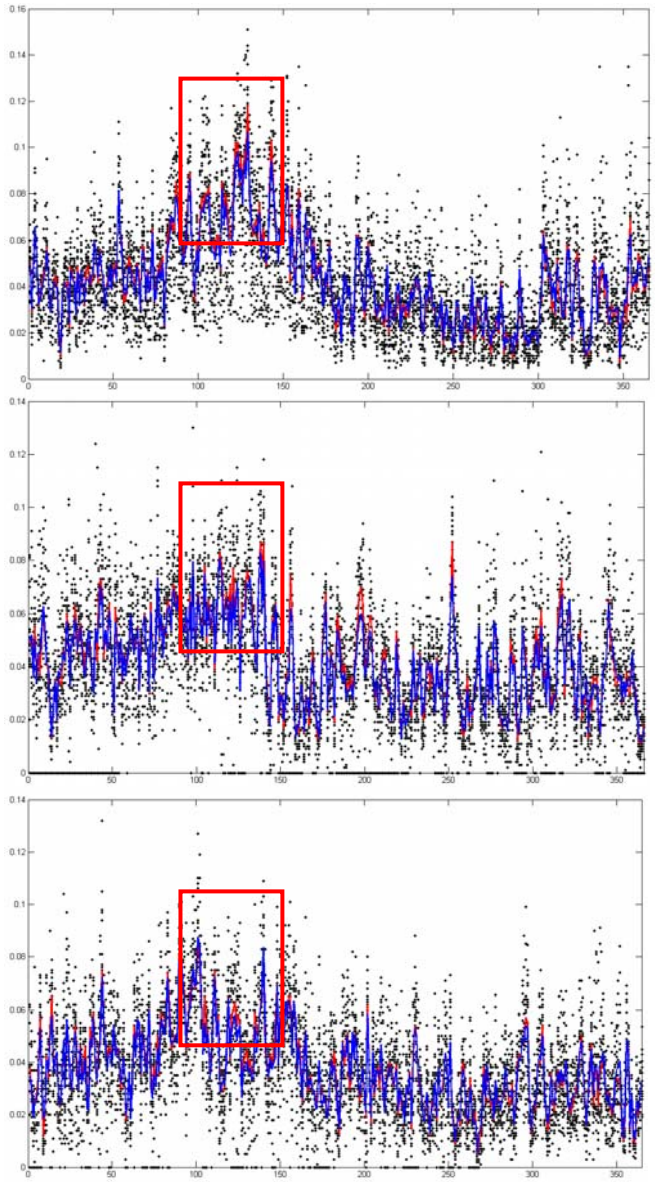


Figure 5.1.4 : extrait pour les années 1998, 2000 et 2001 de l'évolution annuelle de la médiane.

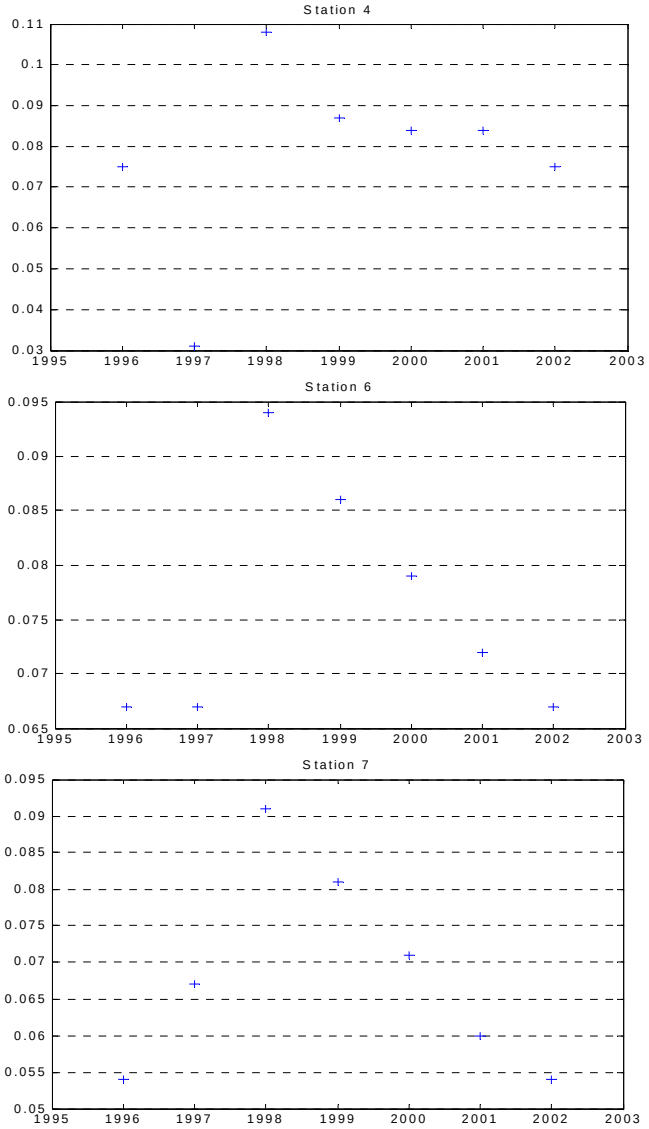


Figure 5.1.5 : extrait pour les stations SAG (4), TLA (6) et XAL (7) de l'évolution interannuelle du maximum

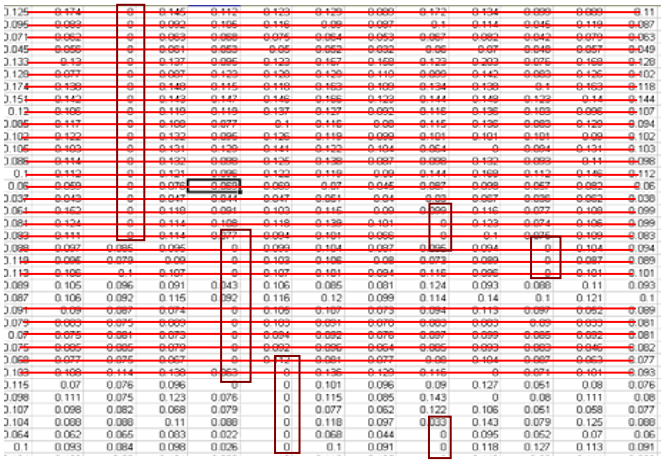


Figure 5.1.6 : présence de valeurs nulles, extrait de la série annuelle 2002

Aussi, la distinction entre les différentes stations lors de la représentation de l'évolution annuelle de la médiane journalière n'apporte ici que peu d'information, c'est pourquoi elle n'est pas évoquée. La figure 5.1.7 en couleur semble toutefois mettre en évidence une prépondérance de valeurs faibles pour la station Chapingo (*bleu cyan*). Cette observation est en accord avec le tableau 5.1.2, à savoir que Chapingo semble n'être que peu affectée par des valeurs de pics.

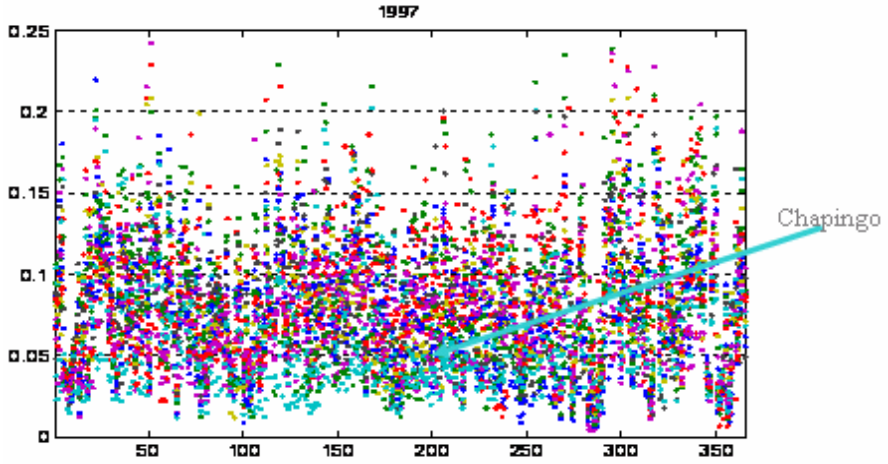


Figure 5.1.7 : évolution annuelle de la médiane pour 1997, distinction entre les stations

5.1.3.1.4. Identification des épisodes de pollution

Une situation de pollution est une série dans l'espace de valeurs relatives à un épisode de pollution. L'obtention d'épisodes et de situations de pollution est indispensable à l'analyse et à la spatialisation des données géographiques, puisque la simultanéité est une condition de la régionalisation. En effet, en admettant dans un cas simple qu'une valeur $Z(x_1)$ mesurée en t_1 et influencée par un champ de paramètres $C(x_1)$ soit comparée à une valeur $Z(x_2)$ mesurée en t_2 et influencée par un champ de paramètres $C(x_2)$, la valeur estimée en $Z(x_i)$ n'a plus de sens puisqu'il est dans la majeur partie des cas déjà impossible de définir chacun des champs.

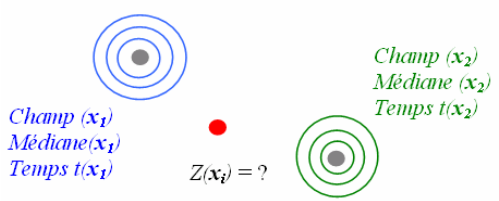


Figure 5.1.8 : illustration du concept de simultanéité (séminaire interdisciplinaire « Interpolation de mesures de qualité de l'air sur la ville de Mexico », Elena Andrey, juin 2004)

A ce stade de l'exploration temporelle des données, nous avons remarqué que chacune des années de mesure présente des valeurs maximales différentes. Il est donc nécessaire de découper le jeu en sept premiers ensembles. A l'intérieur de chacun des sept jeux, il devient aussi plus judicieux de ne s'intéresser qu'aux mois de plus forte concentration, c'est-à-dire aux mois d'avril et mai (*~ jours 90 à 150*), ceci pour plusieurs raisons : la problématique du travail s'intéresse aux concentrations maximales, de plus la sélection de celles-ci permet de réduire une grande partie des traitements numériques. La démarche suivie au chapitre « *Type de distribution et estimateur* » peut ici s'appliquer afin d'identifier quel estimateur permettrait le regroupement des données durant cette période de maximum, a savoir calcul de la médiane journalière (*entre 12 et 18 heures*), représentation de la distribution, choix de l'estimateur (*moyenne / médiane*) pour la période considérée.

Il s'agit donc de représenter la distribution de la médiane par jour sur les mois d'avril et mai. Les résultats obtenus sont des courbes en cloche (*voir annexe 6*). Aucune ne présente de sous populations, ce qui confirme l'hypothèse d'homogénéité de la séquence temporelle choisie. Les courbes sont toutes asymétriques, ce qui préconise l'utilisation de la valeur médiane pour produire les dites « *situations* ».

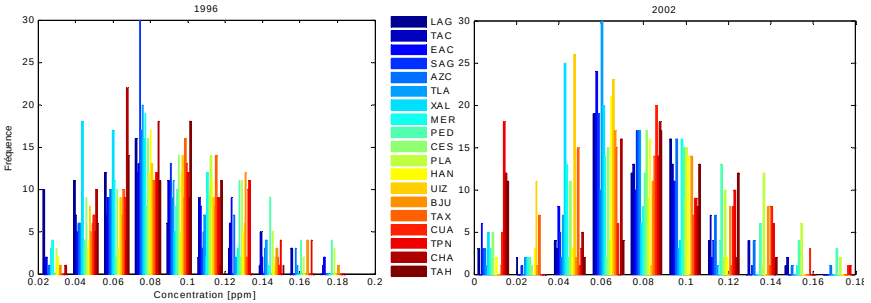


Figure 5.1.9 : extrait pour les années 1996 et 2002 de la distribution de la médiane journalière pour les mois d'avril et mai

Les distributions montrent également que les années 2000, 2001 et 2002 présentent une forte proportion de valeurs nulles, conformément aux observations du chapitre précédent. En 1997 et 1998 les stations SAG et CHA présentent également une quantité anormalement forte de valeurs très faibles.

Rappelons que la définition proposée pour un épisode de pollution est « *un intervalle de temps durant lequel les concentrations attribuées aux différents points de mesure … sont représentatives de cet intervalle* ». Il a donc été choisi de sélectionner **la journée** à l'intérieur de la période de maximum qui se rapproche le plus de la médiane calculée, c'est-à-dire où la somme sur les 19 stations des écarts à la médiane par station est la plus petite. Cette démarche permet de réduire l'intervalle de temps considéré au plus possible, compte tenu de l'évolution très rapide du phénomène. Dans le cas ou deux journées présentaient le même écart minimal, le choix se porterait sur la série dont les écarts aux écarts minimaux étaient les plus faibles. A noter que cette étape présente un inconvénient : la modélisation de phénomènes qui peuvent être tout à fait instantanés (embouteillage, incendie, etc.) et non représentatifs de la famille. Le plus juste serait donc de comparer les cartes produites par la médiane pour avril-mai avec les cartes produites par le calcul de l'écart minimal à la médiane. (permet entre autres d'éliminer des valeurs anormales).

La figure 5.1.10 présente un résumé de la solution proposée pour l'identification des épisodes.

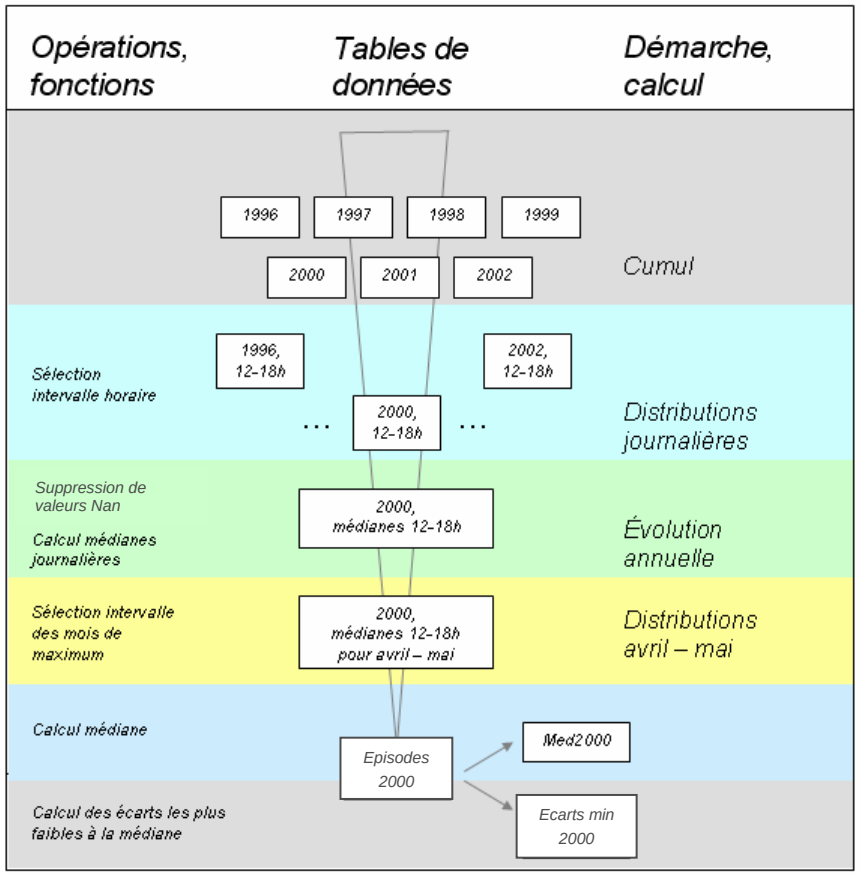


Figure 5.1.10 : procédure d'identification des épisodes

5.1.3.2. Exploration multivariée

Les traitements précédents permettent l'étude de l'influence des variables indépendamment de celle du temps sur la concentration.

Les analyses suivantes sont réalisées dans un premier temps sur chacune des heures du jour (*chapitre « représentation de la variabilité »*), puis sur les valeurs médianes journalières sélectionnées.

Le phénomène exploré est la concentration y, les variables explicatives sont les stations ($sta_1, sta_2, \ldots, sta_{19}$). La mesure y_{ij} en fonction des variables s'écrit donc :

$$y_{ij}=f(sta_{ij})$$

$i=(1,\ldots,n)$ où n = 61, le nombre de jours pour les mois d'avril et mai, et $j=(1,\ldots,p)$ où p = 19, le nombre de stations. La matrice $Y\ (61\times19)$ regroupe l'ensemble des valeurs y_{ij} pour une année donnée. Les analyses suivantes sont réalisées pour la série annuelle 1996 seulement.

5.3.1.2.1. Représentation de la variabilité[22] :

De manière générale la distance interquartiles augmente lorsque les valeurs de concentrations augmentent, ce à quoi nous pouvions nous attendre. En effet, l'augmentation des concentrations dépend de nombreux facteurs (*émissions, météo, etc.*) qui à une heure de la journée donnée peuvent fortement varier. Les dispersions les plus grandes se situent entre 14h et 15h, et sont de l'ordre de 0.07 à 0.08 ppm. Deux stations présentent régulièrement une forte variabilité des concentrations, il s'agit en premier de Pedregal (*PED*) puis de Plateros (*PLA*). Ces stations présentent également de fortes valeurs de concentrations. Les graphiques de l'annexe 10 mettent bien en évidence la relation qui existe entre la variabilité et la concentration.

Remarquons également que la présence de valeurs aberrantes diminue avec l'augmentation de la distance interquartile. L'explication vient du fait qu'une valeur est considérée comme aberrante par l'algorithme de calcul lorsque la distance qui sépare cette valeur du quartile le plus proche est supérieure à 1.5 fois la distance interquartiles. La représentation de ces valeurs est donc relative et ne porte que peu de sens dans cette analyse.

[22] *l'ensemble des diagrammes est en annexe 7*

Pour l'année 1996 le problème ne se pose pas vraiment puisque le nombre de valeurs *NaN* est faible. Dans le cas de la série 2002, l'échantillon aurait fortement réduit par la suppression de ces valeurs, ce qui aurait par conséquent réduit la possibilité pour la médiane de converger vers une même valeur.

De ces représentations nous pouvons enfin déduire que les concentrations maximales se situent donc bien au centre de l'intervalle de temps choisi (*15 heures*).

5.3.1.2.2. Matrice de corrélations

Le calcul de la matrice des corrélations ***R*** (*19×19*) permet de se faire une première idée sur les relations qui existent ou non entre les variables explicatives. L'estimateur de la corrélation pour la $k^{ème}$ et la $v^{ème}$ variable est calculé tel que :

$$r_{kv}=s_{kv}/sqrt(s_{kk}\times s_{vv})$$

$$r_{kv}=(y_{1k}-m_k)(y_{1v}-m_v)+\ldots+(y_{nk}-m_k)(y_{nv}-m_v)]/sqrt[\Sigma(y_{ik}-m_k)^2\ \Sigma(y_{iv}-m_v)^2]$$

Ce calcul est réalisé sur l'ensemble des heures (*12-18*) pour 1996. Les corrélations sont très élevées et toujours positives. Une corrélation positive signifie que pour un couple de stations, lorsque la concentration augmente pour l'une elle devrait également augmenter pour l'autre. Les stations les plus corrélées ($r > 0.8$) sont les suivantes :

Tableau 5.1.3 : principales corrélations

r	***Distance [km]***	***Stations***
0.85	*2.7*	*LAG/MER*
0.88	*5.2*	*TAC/EAC*
0.87	*4.7*	*EAC/AZC*
0.85	*15.3*	*SAG/CHA*
0.84	*4.4*	*AZC/TLA*
0.89	*3.8*	*MER/HAN*
0.90	***4.7***	***PED/PLA***
0.87	*9.6*	*PLA/CUA*
0.81	*15.5*	*TAX/TAH*
0.80	*19.4*	*XAL/CHA*

Cette corrélation devrait être liée à la distance qui sépare les stations. Cette hypothèse est valable pour l'application de n'importe laquelle des méthodes d'interpolation. Le tableau 5.1.3 montre que la pluspart des stations fortement corrélées sont situées à moins de 5 km l'une de l'autre, ce qui est une petite distance pour l'étendue de la surface occupée par le réseau de mesures. On peut également voir quelques valeurs étranges de distances pouvant atteindre 15 voir 19 km ! Aucune valeur intermédiaire n'est observée. Une étude plus approfondie sera faite au chapitre « *Exploration spatiale* ».

5.3.1.2.3. Analyse en composantes principales :

Une Anaylse en Composantes Principales (*ACP*) est une méthode statistique multivariée dont le but est de mettre en évidence des variables d'intérêt. Dans une ACP, chacune des composantes principales est une combinaison linéaire des variables, dont l'importance se mesure en rapport avec la part de variation dans l'échantillon qu'explique chacune de ces composantes. La représentation visuelle des valeurs de variables dans un espace généré par les composantes principales (*2 ou 3*) permet donc d'observer une éventuelle structure dans les données. Cette analyse a été effectuée sur l'épisode 1996.

X (*n×19*) est le jeu de données, où n = 61, les jours de mesures et p = 19, le nombre de variables. Une ACP sur les stations génère ici p = 19 composantes principales. Le pourcentage de variance expliquée en fonction du nombre de composantes utilisées permet d'évaluer l'importance des premières composantes pour la suite de l'analyse.

Dans le cas présent, les deux premières composantes suffisent à expliquer plus de 82% de la variance (*voir figure 5.1.11*). La première composante expliquant à elle seule le 67% de la variance, il est intéressant de regarder plus en détail la part de chacune des stations dans cette composante. Cette démarche indique dans ce cas précis quelles seraient les stations qui ont le plus d'influence sur les valeurs de concentration de l'échantillon.

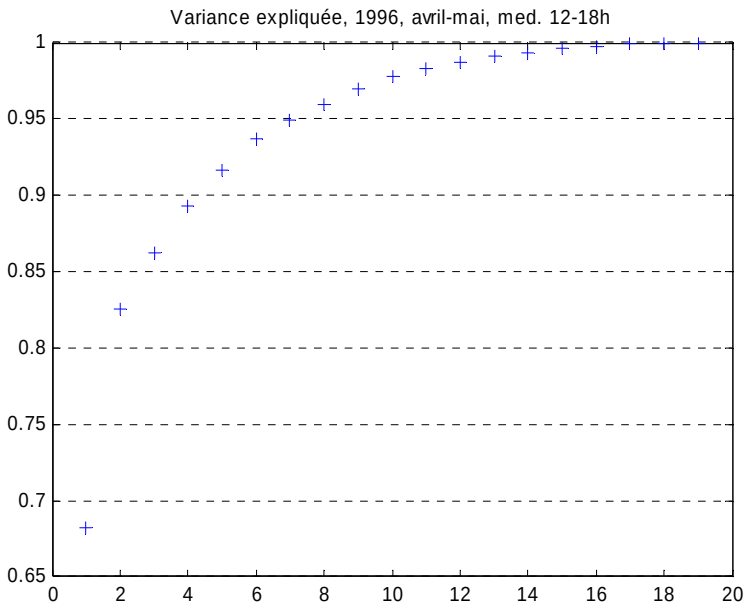


Figure 5.1.11 : variance expliquée en fonction du nombre de composantes principales

Pour ce faire, le tableau 5.1.4 décrit la matrice ***V***(*19×19*) de vecteurs propres. Elle permet d'estimer la part de chacune des variables pour chacune des composantes. Cette matrice est obtenue à l'aide de l'équation suivante :

$$S=V\times L\times V^{T}$$

Où ***L*** (*19×1*) est le vecteur de valeurs propres et ***S*** (*19×19*) la matrice de covariances estimées du jeu de données ***X***.

*Tableau 5.1.4 : description de la matrice **V** des vecteurs propres*

	Comp1	*Comp2*	*…*	*Comp19*
Var1	v_{11}	v_{12}	…	v_{119}
Var2	v_{21}	v_{22}	…	v_{219}
…	…	…	…	…
Var19	v_{191}	v_{192}	…	v_{1919}

La première composante étant la plus intéressante dans ce travail, il a été choisi de calculer la part de variance expliqué par chacune des variables à l'intérieur de cette composante (*les calculs intermédiaires sont en annexe 8*):

Tableau 5.1.5 : part de variance expliquée par chacune des variables pour la composante n°1

Station	***Composante 1***	***Variance expliquée[%]***
LAG	*0.245*	*4.02*
TAC	***0.280***	***5.26***
EAC	*0.243*	*3.96*
SAG	*0.141*	*1.34*
AZC	*0.238*	*3.80*
TLA	*0.214*	*3.05*
XAL	*0.155*	*1.62*
MER	*0.226*	*3.42*
PED	***0.314***	***6.62***
CES	*0.202*	*2.75*
PLA	***0.294***	***5.80***
HAN	*0.220*	*3.24*
UIZ	*0.197*	*2.59*
BJU	***0.269***	***4.85***
TAX	*0.227*	*3.45*
CUA	*0.242*	*3.92*
TPN	*0.235*	*3.69*
CHA	*0.144*	*1.39*
TAH	*0.183*	*2.23*

Les valeurs sont plus ou moins homogènes. Les stations les moins significatives pour la première composante sont Chapingo et San Agustin. Le chpitre « *Représentation de la variabilité* » avait déjà mis en évidence le fait que les stations qui présentent les concentrations les plus faibles sont également celles pour lesquelles la variabilité est la moins importante. C'est le cas pour Chapingo et San Augustin qui figurent aux rangs 19 et 17 dans le tableau 5.1.2.
Plateros, Pedregal, Tacuba et Benito Juarez sont les stations les plus importantes dans l'explication de la première composante. Ces stations se situent aux rangs 1, 2, 4 et 5 du tableau 5.1.2.

On peut aussi remarquer que la présence de valeurs *NaN* rendrait ici impossible la recherche des composantes principales, et les valeurs de type « *Inf* » (*infinite*) ont le même résultat sur ce calcul.
En effet, pour une matrice $\mathbf{X_0}$ de départ (*n×p*), où var_k est la k^{ème} variable, m_k est la moyenne arithmétique pour la k^{ème} variable, s_{jk} la covariance entre les variables

k et j, $\sum$ la matrice de covariance ($p \times p$) et $\mathbf{V}(p \times p)$ la matrice des vecteurs propres, on obtient:

$$s_{jk} = [(y_{1k}-m_k)(y_{1j}-m_j)+\ldots+(y_{nk}-m_k)(y_{nj}-m_j)]/(n-1)$$

$$\boldsymbol{v_1}^T \sum \boldsymbol{v_1} = \text{maximal sous contrainte que } \boldsymbol{v_1}^T \sum \boldsymbol{v_1} = 1$$

$$comp_1 = v_{11}(var_1-m_1)/s_1 - v_{21}(var_2-m_2)/s_2 - \ldots - v_{1p}(var_p-m_p)/s_p$$

Dès lors, le fait de diviser par la covariance est impossible en présence de valeurs *NaN* ou *Inf*.

De cette analyse nous pouvons tirer, en plus de l'identification des stations qui influencent plus particulièrement la série de données, que l'ensemble des traitements est cohérent.

5.1.3.3. Exploration spatiale

L'analyse spatiale regroupe un ensemble de méthodes et d'opérateurs associées aux SIG, exploités dans le but de modéliser l'espace géographique (*EG*), d'en extraire des informations, d'en dériver des informations synthétiques ainsi que d'identifier des relations fonctionnelles entre entités ou phénomènes. La composante de l'espace géographique qu'il s'agit ici plus précisément de représenter est le phénomène de pollution urbaine par l'ozone troposphérique. La variable d'intérêt (*concentration d'ozone*) est donc de type continue. Elle est mesurée en partie par millions (*ppm)* sur une échelle de type cardinale. Les valeurs mesurées sont maintenant indépendantes du temps, seule la variation spatiale est étudiée. Il s'agit donc encore une fois de données univariées.

Chacun des épisodes est représenté par un échantillon de la forme suivante :

Coordonnées		*Variables*
x_1^1	x_1^2	c_1
x_2^1	x_2^2	c_2
...	...	...
x_{19}^1	x_{19}^2	c_{19}

où x_j^i est la composante i des coordonnées planimétriques pour la station n° j, et c_j est la concentration pour la station j.

5.1.3.3.1. Repères théoriques : géostatistique, théorie des variables aléatoires, stationnarité et variographie

Géostatistique :
La géostatistique est une approche mathématique conduisant à la régionalisation de variables d'intérêt selon des méthodes d'interpolation de type géostatistiques. Elle s'oppose aux méthodes d'interpolation de type déterministes. L'approche géostatistique est celle utilisée dans ce travail.

De manière générale l'ensemble des techniques se base sur l'association des concepts de similarité et de proximité, à savoir les points les plus proches les uns des autres tendent à être plus similaires que les points plus éloignés. Les méthodes mathématiques déterministes (*ou géométriques*) régionalisent la variable d'intérêt

sur la base de fonctions mathématiques. La variable est supposée interpolable par le simple fait qu'elle soit de nature spatialement continue, et son comportement spatial est par conséquent empirique. Les techniques géostatistiques diffèrent des méthodes déterministes au sens où elles introduisent la notion de autocorrélation spatiale entre les valeurs mesurées. L'estimation de la variable d'intérêt se fait sur la base de modèles statistiques qui tiennent compte de relations statistiques entre les points échantillonnés. Ces techniques tiennent donc compte d'une information plus complexe et permettent de quantifier l'incertitude lors de l'interpolation.

Contrairement à la statistique classique (*qui suppose que les variables aléatoires sont indépendantes et identiquement distribuées*), la géostatistique admet des variables spatialement corrélées.
L'approche géostatistique se base sur la théorie des variables aléatoires.

Théorie des variables aléatoires :
Une variable régionalisée est l'expression mathématique de la variation spatiale d'un phénomène physique. La grande complexité de la variation du phénomène réel implique cependant généralement l'impossibilité de considérer une variable régionalisée comme une fonction de l'espace. En effet, il existe certainement un très grand nombre de points pour lesquels un phénomène réel peut se comporter de manière tout à fait irrégulière, malgré une tendance définie sur l'ensemble de l'*EG*. On peut définir la variable régionalisée $z(\mathbf{x})$ comme étant l'ensemble des valeurs définies que l'infinité de points composant l'*EG* sont susceptibles de prendre :

$$\{z(\mathbf{x}),\ \forall \mathbf{x}\ \in\ EG\}$$

Toute valeur mesurée sur l'*EG* est donc une valeur régionalisée. La théorie des variables aléatoires considère toute valeur régionalisée $z(\mathbf{x_i})$ comme étant la réalisation d'une variable aléatoire $Z(\mathbf{x_i})$,où *i* est un indice relatif à la mesure (*i=1,…,19*). Aussi, $Z(\mathbf{x_i})$ peut avoir des propriétés différentes en chaque point de mesure. L'ensemble des variables aléatoires est la fonction aléatoire $Z(\mathbf{x})$:

$$\{Z(\mathbf{x}),\ \forall \mathbf{x}\ \in\ EG\}$$

Le modèle proposé par Wackernagel à la figure 5.1.12 illustre cette théorie.

La théorie des variables aléatoires permet donc d'intégrer une information plus complexe, de nature probabiliste (*par opposition à déterministe*), lors de la régionalisation de phénomènes naturels.

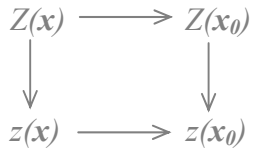


Figure 5.1.12: Interprétation probabiliste de la variable régionalisée (Wackernagel, 1998)

Stationnarité :
L'estimation de la variable à régionaliser se fait sur la base de la théorie des variables aléatoires. L'approche géostatistique admet des mesures corrélées en fonction de la distance, et les modèles statistiques utilisés par les techniques géostatistiques d'interpolation tiennent compte des autocorrélations spatiales entre les points échantillonnés. Ces relations se précisent à l'aide d'une analyse structurale des données, au centre de laquelle se trouve la variographie (*abordée dans le § suivant*).

Cette analyse structurale admet l'hypothèse de stationnarité de la fonction aléatoire $Z(\mathbf{x})$. La stationnarité stricte signifie une invariance en translation du comportement d'un phénomène. Lorsque celui-ci est décrit par une fonction aléatoire ce sont les caractéristiques de la fonction aléatoire qui sont invariantes en translation :

Si $F\mathbf{x_i}$ est la distribution de probabilité de la variable aléatoire $Z(\mathbf{x_i})$ à la station i (i=1,…19), i.e. $P(Z(\mathbf{x_i}) \leq z)$, alors $F\mathbf{x_1,\ldots,x_{19}}(z_1,\ldots,z_{19})$ est la fonction de distribution multiple (pour l'ensemble des stations) pour les 19 variables aléatoires. Dans ce cas la stationnarité stricte est définie comme :

$$F\mathbf{x_1,\ldots,x_{19}}(z_1,\ldots,z_{19}) = F\mathbf{x_1+h,\ldots,x_{19}+h}(z_1,\ldots,z_{19})$$

La stationnarité stricte étant une condition difficile à remplir (*puisqu'elle suppose la connaissance de la fonction de distribution multivariée !*) elle est généralement remplacée par la notion de stationnarité intrinsèque.

La stationnarité intrinsèque d'une fonction aléatoire suggère l'invariance en translation des caractéristiques de la distribution de l'écart entre paires de valeurs échantillonnées ($Z(\boldsymbol{x}+\boldsymbol{h}) - Z(\boldsymbol{x})$). Cette hypothèse ne nécessite donc plus que la connaissance de la fonction de distribution de l'écart entre paires de variables, qui dans le cas de la modélisation de phénomènes naturels tels que la concentration d'ozone est supposée gaussienne. La stationnarité intrinsèque d'une fonction aléatoire est donc vérifiée lorsque les conditions suivantes sont remplies :

$$\mu(\boldsymbol{h}_{kl}) = E[Z(\boldsymbol{x}_i+\boldsymbol{h}_{kl})] = E[Z(\boldsymbol{x}_k)] = cste$$

$$\sigma^2(\boldsymbol{h}_{kl}) = Var[Z(\boldsymbol{x}_k+\boldsymbol{h}_{kl}) - Z(\boldsymbol{x}_k)] = Cov[Z(\boldsymbol{x}_k+\boldsymbol{h}_{kl}),Z(\boldsymbol{x}_k)] = C(\boldsymbol{h}_{kl})$$

Où k et l identifient les points de mesures $\boldsymbol{x}_i$ ($i=1,\ldots,19$).
Les deux premiers moments caractérisent entièrement la distribution des écarts $F_{\boldsymbol{h}_{kl}}(z_k,z_l)$ dans le cas d'une distribution normale. Il s'agit alors de la moyenne arithmétique et de l'écart type. La moyenne doit être constante et la variance ne doit dépendre que de la distance $\boldsymbol{h}$ séparant les points de mesure sur l'ensemble de l'*EG* considéré.

Variographie :
L'analyse variographique est le point central de l'analyse structurale d'un échantillonnage et donc de la procédure d'interpolation géostatistique. Du fait de la prise en compte non seulement de la distance[23] mais aussi des relations de dépendance (*autocorrélations*) qui existent entre les points de mesure, c'est-à-dire de la structure spatiale de l'échantillonnage.
L'analyse variographique a pour but de mesurer « la force » de la autocorrélation spatiale entre les variables aléatoires $Z(\boldsymbol{x}_i)$ par une fonction de la distance $\boldsymbol{h}$:

soit par la covariance : $C(\boldsymbol{x}_k,\boldsymbol{x}_l) = C(\boldsymbol{h}_{kl}) = C(\boldsymbol{x}_k-\boldsymbol{x}_l) = Cov(Z(\boldsymbol{x}_k),Z(\boldsymbol{x}_l)$

soit par le variogramme : $\gamma(\boldsymbol{x}_k,\boldsymbol{x}_l) = \gamma(\boldsymbol{h}_{kl}) = \frac{1}{2}(z_k-z_l)^2 = \frac{1}{2}\, var(Z(\boldsymbol{x}_k)-Z(\boldsymbol{x}_l))$

mais ces fonctions sont équivalentes puisque $\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - (\boldsymbol{h})$

Pour chaque paire de valeurs possible (*19 stations = 19+18+17+…+1=171 paires de valeurs*) le carré de la différence entre les valeurs est calculé. En pratique ce sont généralement les valeurs de $\gamma(\boldsymbol{h}_{kl})$ qui sont reportées en fonction du vecteur de séparation $\boldsymbol{h}$, on peut alors représenter la nuée variographique.

[23] *La distance est le seul paramètre de calcul des techniques déterministes d'interpolation telles que la pondération inverse à la distance, la triangulation, etc.*

Une éventuelle anisotropie est un phénomène important à modéliser dans le cas de régionalisation de mesures atmosphériques. Dans le cas de mesures d'ozone elle pourrait être causée par des facteurs tels que la direction préférentielle des vents, la topographie, etc. La modélisation d'une tendance directionnelle se fait par un variogramme surfacique.

Dans le cas d'un phénomène anisotropique le variogramme expérimental est une sorte de coupe dans le variogramme surfacique. Il est construit sur la base des directions pricipales d'anisotropie identifiées (*ou non*), et regroupe un ensemble de paires séparées par le vecteur $\boldsymbol{h}$, orienté dans la direction choisie.

L'ajustement des points du variogramme expérimental sur une fonction théorique de covariance définit les paramètres de l'interpolation. Le variogramme théorique est noté $\gamma^*(\boldsymbol{h})$. Cet ajustement est un processus itératif et manuel. Les paramètres à fixer sont la fonction de covariance, la pépite[24], le seuil, l'existence d'une anisotropie en fonction de la direction et le voisinage à considérer.

Finalement, il n'existe pas vraiment de règles pour la modélisation variographique. Celle-ci repose sur le bon sens et l'expérience de l'opérateur. L'évaluation de la qualité d'une modélisation se fait habituellement à l'aide de la procédure de validation croisée, ou plus simple encore par l'observation.

Méthodes géostatistiques d'interpolation :
Les méthodes géostatistiques d'interpolation sont pour ainsi dire quasiment toutes regroupées dans la famille des krigeages. Matheron (*1963*) propose la définition suivante « *…un ensemble de techniques revenant à effectuer une pondération des échantillons, …les poids étant calculés de façon à rendre minimale la variance d'estimation résultante, compte tenu des caractéristiques géométriques du problème…* ».

Il existe de nombreux différents types de krigeages. Tous produisent un estimateur sans biais, dont la description mathématique peut se faire de la manière

[24] *L'effet de pépite est une discontinuité initiale attribuée soit au type de continuité de la variable, soit aux erreurs systématiques de mesure, soit aux variations à très petite échelle du phénomène. Cet effet est donc sans doute faible, voire négligeable pour les données à disposition. Aussi, l'identification des causes de la pépite n'est pas possible uniquement sur la base des données.*

suivante (*les λ_i (i=1,…, N) sont les poids du krigeage, N est le nombre de stations de mesure*) :

Estimateur :	$Z(\mathbf{x_0}) = \Sigma \lambda_i Z(\mathbf{x_i})$	*i = 1,.., N et N le nombre de points d'appuis (stations)*
Conditions :	$\Sigma \lambda_i = 1$	*Garantie d'un estimateur sans biais*
Principe :	$[Z(\mathbf{x_0}) - \Sigma \lambda_i Z(\mathbf{x_i})]^2 = \sigma^2 = min$ *(pour i = 1,.., N)*	*variance d'estimation minimale compte tenu de la condition sur la somme des poids*
Equations du krigeage :	*système matriciel permettent de choisir les poids λ_i t.q. l'erreur quadratique d'estimation soit minimale ; utilise la fonction de corrélation C(**h**) modélisée sur la base du variogramme expérimental :* $\boldsymbol{\Gamma} \times \boldsymbol{\lambda} = \boldsymbol{g}$ *, ou alors*	

$$\begin{bmatrix} \gamma(x_1 - x_1) & \dots & \gamma(x_1 - x_n) & 1 \\ \dots & \dots & \dots & \dots \\ \gamma(x_n - x_1) & \dots & \gamma(x_n - x_n) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \dots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(x_1 - x_0) \\ \dots \\ \gamma(x_1 - x_n) \\ 1 \end{bmatrix}$$

Les différents types de krigeage reposent tous sur les hypothèses de stationnarité intrinsèque et de distribution univariée[25] normale de chaque variable aléatoire pour obtenir un estimateur « *optimal* ».

La méthode de krigeage adoptée ici est le Krigeage Ordinaire. Cette méthode est l'une des plus utilisées, et s'avère être la plus apropriée à l'interpolation des données présentes (*voir Séminaire interdisciplinaire « Interpolation de données de qualité de l'air sur la ville de Mexico », Elena Andrey, juin 2004*). Elle est une méthode d'interpolation locale et fait intervenir la notion de voisinage: la valeur en un point donné est estimée à l'aide des points de mesure d'une fenêtre d'interpolation dont les dimensions sont définies par l'opérateur. La stationnarité intrinsèque n'y est donc définie que localement.

Les différents traitements sont en majeurs partie réalisés à l'aide des logiciels spécifiques : VarioWin 2.2® pour l'analyse structurale, et ArcGIS 9™ (*extension Geostatistical Analyst*®) pour la représentation.

5.1.3.3.2. Unités d'observation

Les unités d'observation sont les stations de mesure. Chacun des points est caractérisée par ses coordonnées géographiques (*UTM H.N. 14*), par des photos de l'endroit où la station est implantée, par le type de zone (*urbaine, suburbaine, rurale*), par l'altitude de l'échantillonneur au dessus du sol, par le secteur géographique (*NO, NE, SO, SE, etc.*), etc. Ces informations se trouvent à l'adresse suivante :
http://www.sma.df.gob.mx/simat/pnrednueva.htm.

5.1.3.3.3. Echantillonnage

La représentativité de l'échantillon de l'ensemble de la zone d'étude peut généralement se calculer à l'aide d'une statistique ***R*** de qualité de l'échantillonnage[26] (*« Analyse spatiale », R. Caloz, 2003*), tel que :

$$R = \frac{\bar{d}}{\bar{d}_{al}} = \frac{\frac{1}{n}\sum_i d_i}{\frac{1}{2}\sqrt{s/n}}$$ = *le rapport entre la distance moyenne entre points d'appui et la distance moyenne dans une distribution aléatoire*

[25] *Une distribution univariée normale se rapporte à une distribution normale pour chacun des attributs de la VA dans le cas d'une analyse multivariée*
[26] *cet indice ne dit rien sur l'adéquation du pas d'échantillonnage*

d_i est la distance du point *i* au plus proche voisin, *s* la surface de la zone d'étude et *n* le nombre de points d'appui.

Lorsque ***R*** = 1 l'échantillonnage est aléatoire. Pour ***R*** < 1 la répartition des points se rapproche d'un échantillonnage de type groupé, non représentatif de la distribution spatiale d'un phénomène. Lorsque ***R*** > 1, l'échantillonnage est supposé représentatif (*voir figure 5.1.13*).

Dans le cas présent, *n* = 19, *s* = 5'071 km^2 (*total de la surface de la ZMVM*) et ***R*** = 0.68. Cette valeur pour ***R*** est peu satisfaisante. Dans le cas où la ZMVM était réduite à une surface plus rapprochée des stations (*définition d'un polygone de s = ~ 1'800 km^2, voir figure 5.1.14*) ***R*** = 1.15, ce qui se rapprocherait plus d'une distribution de type aléatoire.

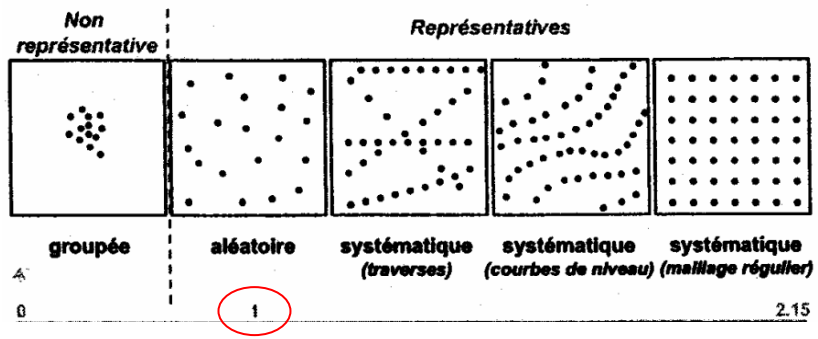


Figure 5.1.13 : typologie des modes d'échantillonnage, cours « Analyse spatiale, R. Caloz, SSIE 2003 »

A ce stade il est important de nuancer les conclusions tirées quant à la qualité de l'échantillonnage. En effet, la statistique ***R*** s'applique généralement à un phénomène de nature « *invisible* », c'est-à-dire lorsque qu'il n'existe aucune connaissance préalable sur le comportement spatial de la variable d'intérêt. Dans ce cas précis le meilleur échantillonnage est de type aléatoire. La densité de points doit être constante et représentative du phénomène à régionaliser.

Dans le cas du RAMA les stations sont distribuées stratégiquement pour les raisons suivantes :

- les connaissances préalables ainsi que les données en rapport avec la pollution atmosphérique urbaine sont nombreuses, ce qui privilégie la localisation des stations

- les stations mesurent des variables autres que l'ozone (particules, SO_2, NO_x, humidité relative, température, etc.) pour lesquelles les points choisis peuvent être justifiés

- les stations de mesure ont été implantées dans les années '80. La zone urbaine s'est encore développée depuis.

R ne rend donc que partiellement compte de la qualité de l'échantillonnage pour le RAMA.

*Figure 5.1.14 : délimitation d'un polygone rapporché pour le calcul de la statistique **R***

5.1.3.3.4. Représentation préliminaire des données

Une première representation des données permet d'observer les tendances générales du comportement spatial de la variable, le degré d'aggrégation spatiale des observations, ainsi que les éventuelles anomalies. La technique d'interpolation adoptée est celle qui a été jusqu'à présent utilisée, pour la représentation spatiales de valeurs de concentration en ozone troposphérique,

dans les exemples proposés par les différents articles consultés. Il s'agit de la pondération inverse à la distance.

Les données représentées sont les épisodes issus du chapitre « *Exploration temporelle* ». L'année 2002 ne peut cependant pas être utilisée puisqu'elle contient une trop grande quantité de valeurs nulles. En effet, les journées de mesures utilisables reviennent au final dans cette série à 3%, sur un total de 61 jours (*avril-mai*). Lors des analyses précédentes les matrices étaient beaucoup plus grandes et cette proportion était par conséquent beaucoup plus grande.

L'unité de mesure généralement utilisée pour les représentations et l'interprétation des données est le « *part per billion* » (*1ppm = 1000 ppb*). La validation des mesures ainsi que les explorations temporelle et multivariée des données garantissent qu'aucune valeur aberrante ne devrait entacher les mesures. Les points représentent les stations de mesure. La statistique ***R*** de qualité de l'échantillonnage calculée au chapitre précédent montre que pour une surface la plus raprochée possible des stations (*zone d'interpolation par IDW*) la distribution spatiale est relativement aléatoire.

Il a été choisi dans un premier temps de vérifier la correspondance entre la médiane calculée sur les 61 jours de maximum (*« Médiane »*) et l'épisode à cartographier (*« Ecarts min. »*): la journée selectionnée est-elle représentative du groupe temporel identifié ? La démarche adoptée est une simple superposition des situations en transparence.

Dans un second temps les épisodes ont étés représentés pour observer les tendances générales du comportement de l'ozone.

Deux échelles de classification ont donc été mises en place. La première « *Total* » considère les deux situations. La seconde (« *Ecarts min* ») ne tient compte que des épisodes à représenter. Ces échelles segmentent les données en 10 classes de valeurs, dont les limites sont fixées par les quantiles des séries de données considérées (*q10%, q20%, etc.*). Les classes obtenues sont les suivantes :

Total	Ecarts min.
conc. [ppb]	conc. [ppb]
33.1 - 62	61.0 - 67
62.1 - 71	67.1 - 77
71.1 - 79	77.1 - 81
79.1 - 86	81.1 - 85
86.1 - 90	85.1 - 87
90.1 - 95	87.1 - 91
95.1 - 97	91.1 - 97
97.1 - 105	97.1 - 101
105.1 - 116	101.1 - 111
116.1 - 141	111.1 - 133

Le résultat de la superposition (*voir figure 5.1.15*) montre que les deux situations sont visuellement très proches. Les épisodes identifiés sont en quelque sorte « *validés* » de ce point de vue.

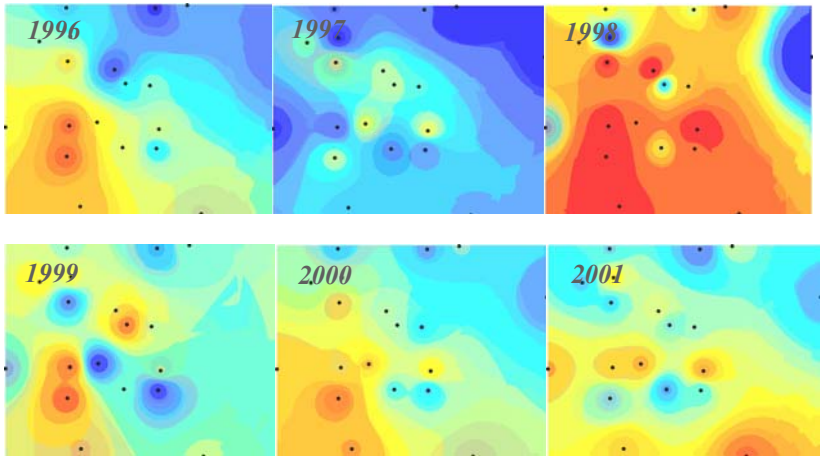


Figure 5.1.15 : interpolation par pondéraion inverse à la ditance des episodes 1996 à 2001, superposition en transparence de la médiane calculée et de la journée la plus représentative de la période de maximum

La figure 5.1.16 montre pour les épisodes 1996, 1998, 1999 et 2000 que les concentrations les plus fortes se situent au sud-ouest (*Pedregal, Tlalpan et Plateros*) alors que les concentrations les plus faibles sont au nord-ouest (*Chapingo et San Augstin*). Ces observations ont déjà fait l'objet de remarques aux différents chapitres *« La qualité de l'air pour la Zone Métropolitaine de la Vallée de Mexico, facteur géographique »*, « *Exploration temporelle* » et « *Exploration multivariée* ». Aussi, cette représentation montre à priori une importante variabilité entre les années, ce qui est peut paraître étonnant puisque chacun des épisodes est composé de la même période de pic ainsi que du même estimateur. Ils devraient donc logiquement êtres semblables.

Ces représentations permettent également de conclure sur la méthode d'interpolation utilisée (*IDW*) qui n'est pas appropriée à une observation à plus petite échelle du phénomène régionalisé, les paramètres étant ajustés de façon optimale (*plus particulièrement rayon de recherche*). En effet, les cartes montrent des effets de puits et de sources centrés sur les stations de mesure, ce qui n'est pas logique : les stations ne produisent ni ne détruisent l'ozone.

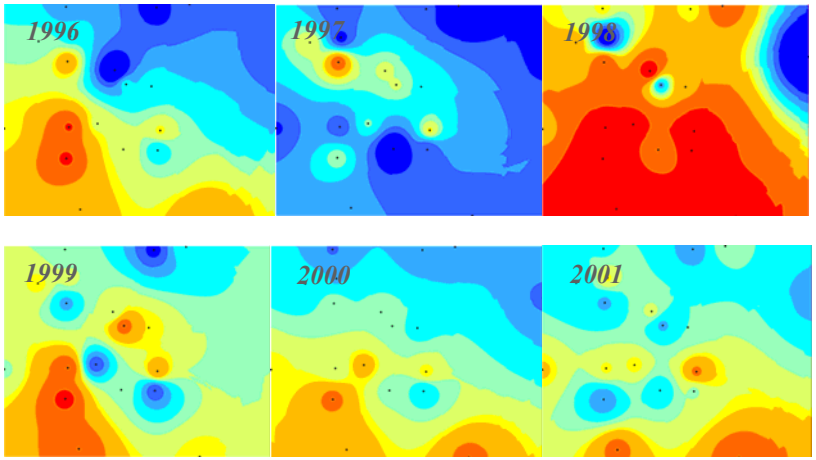
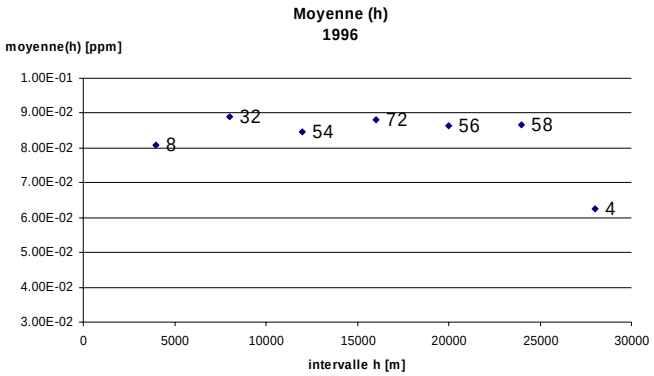


Figure 5.1.16 : interpolation par pondération inverse à la distance des episodes 1996 à 2001, représentation des journées les plus représentatives de la période de maximum (épisode identifiés)

5.1.3.3.5. Vérification de la stationnarité intrinsèque

La stationnarité intrinsèque conduit à la notion de variogramme. L'hypothèse à vérifier est l'invariance en translation de la moyenne et de l'écart-type de la distance séparant une paire de points (*il est admis que la distribution des concentrations est normale*). Cette vérification se fait en représentant $\mu_{conc}(h)$ *et* $\sigma_{conc}(h)$. Les graphiques ci-dessous présentent cette relation pour l'épisode 1996.



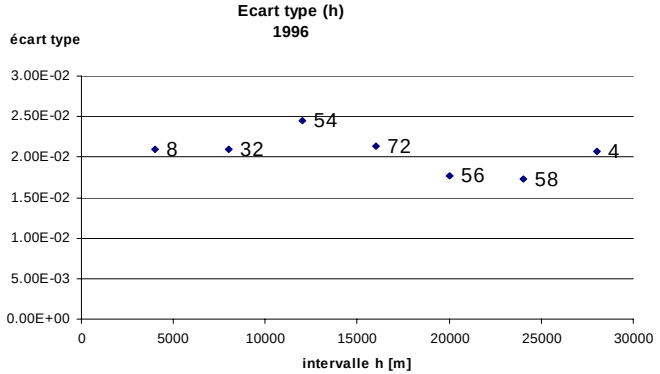


Figure 5.1.17 : vérification de la stationnarité intrinsèque pour l'épisode 1996

Les mesures sont regroupées par paires puis par intervalle de distance h, afin d'obtenir un nombre suffisant de paires pour le calcul des statistiques (*libellé*). Ce nombre est évalué arbitrairement.

5.1.3.3.6. Analyse structurale des données : variographie

La démarche analytique est systématique pour chacun des épisodes. Elle comprend les étapes successives explicitées par les sous chapitres suivants. La majorité des traitements sont réalisés à l'aide du logiciel spécifique VarioWin 2.2®. La modélisation variographique est effectuée dans ArcGIS 9™, par son extension Geostatistical Analyst®.

Cartographie des stations
Une représentation préliminaire permet de vérifier l'emplacement des points échantillonnés (*voir figure 5.1.18*). Il est en effet possible d'introduire des erreurs lors de la création des fichiers de format **.**dat permettant la lecture de la variable à explorer dans VarioWin®.

Après vérification, la saisie des coordonnées semble correcte.

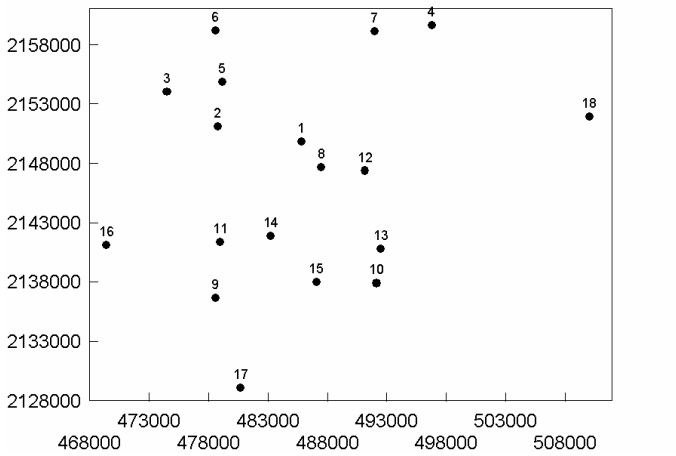


Figure 5.1.18 : vérification de l'emplacement des stations

Nuée variographique:
La nuée variographique est une première analyse des données. Elle permet le repérage des erreurs de mesures (*qu'il ne devrait pas y avoir*) et des anomalies (*valeurs extrêmes*). Elle essaie aussi d'identifier une autocorrélation spatiale de la variable.

La nuée variographique est la représentation de $\gamma(\boldsymbol{h}_{kl})$ en fonction de la distance entre les paires $\boldsymbol{h}_{kl}$:

$$\gamma(\boldsymbol{h}_{kl}) = \tfrac{1}{2}(z(\boldsymbol{x}_k) - z(\boldsymbol{x}_l))^2$$

Pratiquement, la distance maximale $\boldsymbol{h}_{klmax}$ est généralement choisie comme égale à la moitié de la distance maximale séparant une paire de points sur le domaine d'étude. Dans le cas présent, l'espace est délimité par les valeurs suivantes :

Tableau 5.1.6 : délimitation de l'espace géographique des stations

	X [m]	***Y [m]***
Min	*469'387*	*2'127'890*
Max	*510'078*	*2'159'593*
Δ max	*40'691*	*31'703*

D'où $\boldsymbol{h}_{klmax} = \tfrac{1}{2}((\Delta maxX^2 + \Delta maxY^2)^{\frac{1}{2}}) = 25'500$ *m*.

Les paramètres à ajuster sont la direction de recherche, la tolérance angulaire ainsi que la largeur de bande (*voir figure 5.1.19*).

Lors d'une première analyse des données, la recherche des paires est omnidirectionnelle. La direction est donc nulle, la tolérance équivaut à 90° et la largeur de bande est infinie.

La présence de valeurs anormales aux points de mesures sépare le nuage dans sa hauteur et trouble la représentation d'une éventuelle structure dans les données. En effet, certains points présentent systématiquement une variation $\gamma(\boldsymbol{h}_{kl})$ importante avec les autres stations. Une fois ces valeurs écartées, il est possible de vérifier l'existence d'une éventuelle autocorrélation spatiale.

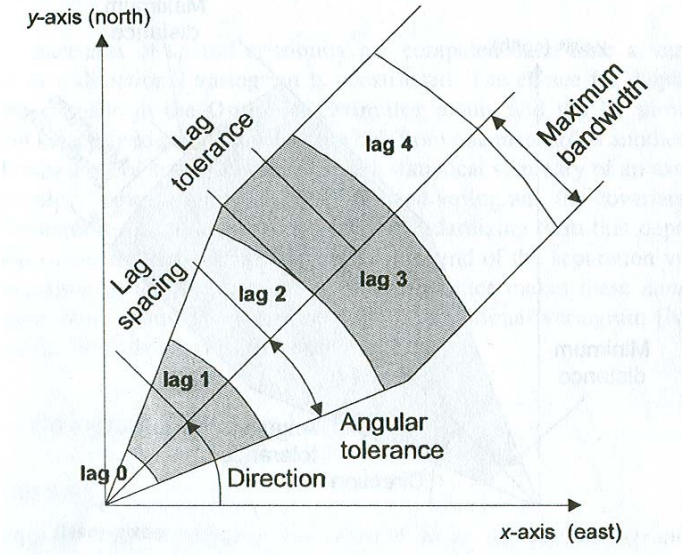


Figure 5.1.19 : direction de recherche, tolérance angulaire et largeur de bande, Pannatier, 1996

Lors de la représentation de la nuée pour chacun des épisodes aucune valeur extrême n'a été détectée. En effet, malgré une structure des données qui n'est pas toujours évidente (*1996*), le nuage de points ne se sépare pas réellement dans sa hauteur (*voir figure 5.1.20*).

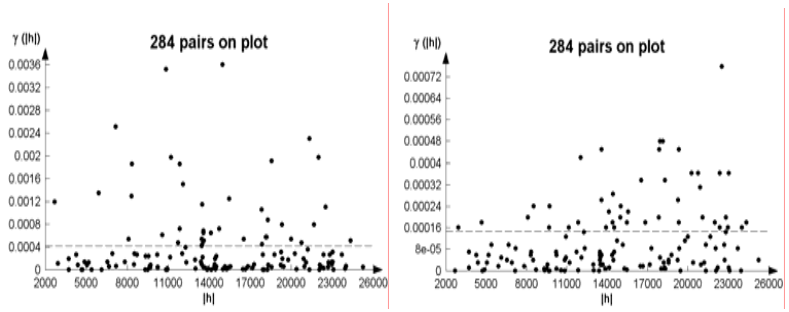


Figure 5.1.20 : nuée variographique pour les épisodes 1996 et 2000

Une représentation cartographique des paires dont la variance est la plus importante permet de localiser les stations pouvant présenter une valeur extrême. Pour l'épisode 1996 par exemple la station la plus concernée est Lagunilla (*voir figure 5.1.21*).

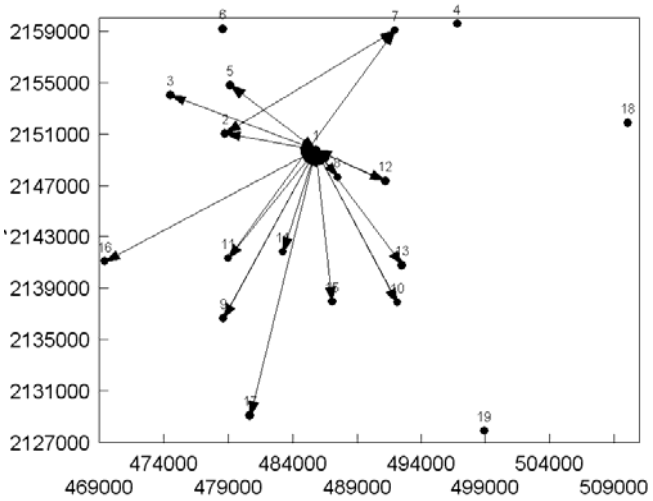


*Figure 5.1.20 : localisation des stations concernées par des variations importantes de γ(**h**)*

Cependant la valeur qui caractérise Lagunilla est de 0.033 ppm, ce qui est une concentration ordinaire dans l'absolu pour la ZMVM. Les autres épisodes ne présentent pas non plus de concentration extrême (*voir tableau 5.1.7*). C'est aussi pour cette raison qu'aucune des valeurs n'est éliminée. Ce tableau permet également d'observer que les valeurs maximum sont situées ici pour la majeure partie à Pedregal, suivi de Tlalpan. Les valeurs minimum se trouvent à Chapingo. Cette observation concorde encore une fois avec la classification effectuée au tableau 5.1.2.

Tableau 5.1.7 : valeurs limites pour les différents épisodes

Episode	***Valeurs limites min/max [ppm]***	***Station(s)***
1996	*0.033 / 0.118*	*LAG / PED*
1997	*0.033 / 0.112*	*SAG / TAC*
1998	*0.029 / 0.141*	*CHA / TPN*
1999	*0.059 / 0.121*	*XAL / PED*
2000	*0.070 / 0.109*	*TLA et CHA / PED*
2001	*0.071 / 0.115*	*PED / TAH*

Finalement certains épisodes présentent une structure (*1998, 2000 et 2001, voir annexe 9*) alors que pour d'autres elle est moins évidente (*1996 et 1997*). Cette démarche ne permet cependant pas de conclure qu'il n'existe absolument pas de corrélation spatiale entre les points. Il se pourrait par exemple qu'une structure soit révélée lors d'une étude de la directionnalité du comportement du phénomène. Il s'agit pour en être sur d'aller plus en avant dans l'exploration spatiale des données.

Cependant, pour 1999 la structure semble presque absente (*voir figure 5.1.22*). Le nombre de points étant trop restreint il est impossible d'éliminer certaines des valeurs qui pourraient masquer la structure spatiale des données (*puisqu'il est certain qu'elle existe*). Il est donc préférable d'écarter cet épsiode dans la suite des traitements.

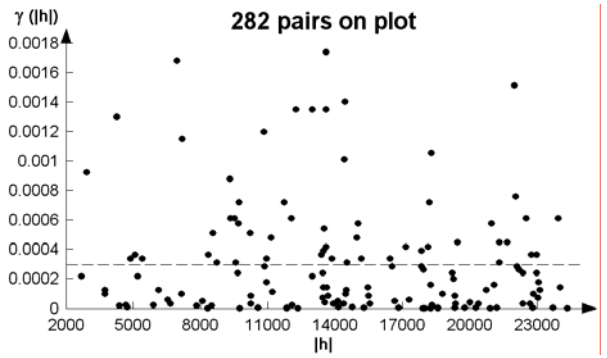


Figure 5.1.22 : nuée variographique pour l'épisode 1999

La représentation de la nuée variographique sous forme d'un histogramme est également un moyen de faire une première estimation d'un ordre de grandeur du pas de l'intervalle dans lequel seront regroupés les paires lors du calcul du variogramme expérimental (*lag spacing*). Il s'agit de trouver le plus petit intervalle comprenant un nombre suffisant de paires pour le calcul des statistiques nécessaires à la construction du variogramme expérimental.

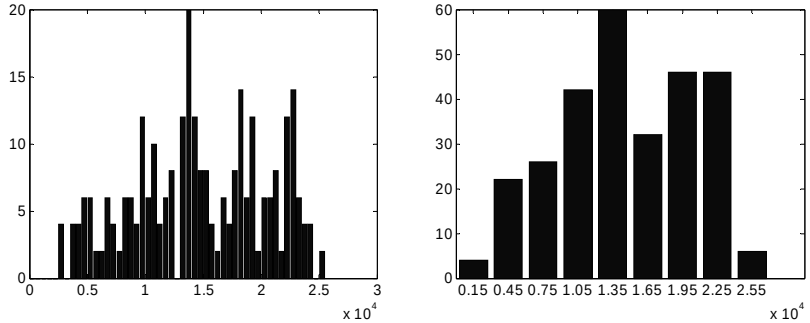


Figure 5.1.23 : histogrammes de la nuée variographique pour 1996 (a) lag spacing = 500 m, (b) lag spacing = 3000

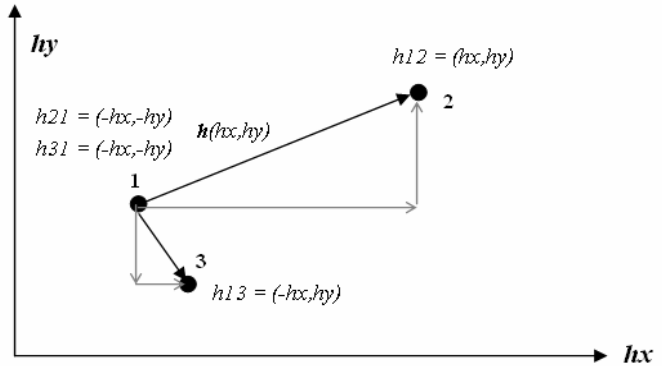
Le premier histogramme correspond au regroupement des paires de valeurs de la nuée variographique pour 1996 dans des intervalles dont le pas est de 500m. Les fréquences sont trop faibles, et le pas est trop petit. Le second regroupe les mêmes paires dans des pas de 3'000m. Dans chacune des catégories on retrouve quelques dizaines de valeurs. Il s'agit à priori d'une bonne fréquence pour les calculs.
On peut également observer sur ces graphes que le nombre de paires devient insuffisant à partir de ~25'000m, ce qui fait dire que l'autocorrélation spatiale ne devrait pas être considérée au-delà de cette distance.

Variogramme surfacique :
Un variogramme surfacique permet l'identification d'un comportement anisotrope[27] de la variable étudiée, c'est-à-dire lorsque la structure spatiale change non seulement en fonction de la distance mais également de la direction. Cette analyse est importante dans le cas de la ZMVM qui subit l'influence complexe de la convergence de masses d'air, ayant par ce biais une influence directe sur la répartition spatiale des concentrations. En effet, il en résulterait un mélange atmosphérique provoquant l'uniformisation des concentrations dans cette direction, et par conséquent une forte corrélation spatiale entre les valeurs quelle que soit la distance entre les paires de stations.

Le variogramme surfacique (*ou bivarié*) est une représentation en plan de $\gamma(\boldsymbol{h_{kl}})$. La différence avec un variogramme univarié est la distinction entre les composantes *x* et *y* du vecteur $\boldsymbol{h_{kl}}$. Cette distinction permet justement d'identifier la direction de la variation :

$$\boldsymbol{h}_{kl} = (hx_{kl} ; hy_{kl})$$
$$|\boldsymbol{h}_{kl}| = (hx_{kl}^2 + hy_{kl}^2)^{1/2}$$

La représentation de ces composantes peut se faire tel que :



On peut aussi déduire de cette représentation que la surface résultante est symétrique.
Les distances maximales empiriquement choisies pour l'évaluation de la autocorrélation sont :

| hx max | = ¾ ΔX max = ¾ × 40'691m ≅ 30'000 m
| hy max | = ¾ ΔY max = ¾ × 30'703 m ≅ 24'000 m

La représentation de la surface nécessite la segmentation de l'espace dans chacune des composantes *hx* et *hy* en un nombre d'intervalles donné. Le pas de chaque intervalle est donné par *lag spacing* et le nombre d'intervalles par *nb.lags*.

hx max = nb. lags.x × lag spacing.x
hy max = nb. lags.y × lag spacing.y

[27] *Il s'agit dans ce cas d'une anisotropie géométrique : la variance maximale (seuil) est identique dans toutes les directions alors que la distance d'extinction (portée) de la autocorrélation diffère*

Le résultat est la variation $\gamma(\boldsymbol{h})$ dans un diagramme bivarié, et dont les limites sont [*-hx max : +hx max ; -hy max : +hy max*].

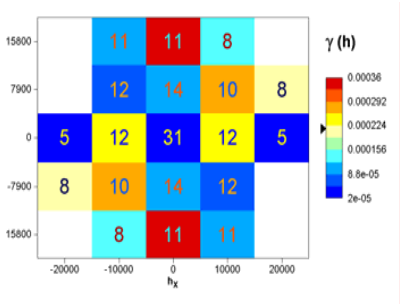


Figure 5.1.24: exemple de variogramme surfacique

Chacune des cases contient un nombre de paires, séparées par une distance $\lvert \boldsymbol{h} \rvert$. Par exemple la case (*7'900 ;10'000*) contient 10 paires de points dont l'écartement peut varier entre (*7'900±7'900/2 ;10'000±10'000/2*) = (*3950:11'850 ; 5'000 :15'000*). Sur le même diagramme, on peut observer des valeurs de concentration plus particulièrement corrélées dans la direction ~130°.

Le choix du *lag spacing* doit être relatif à la distance qui sépare les stations. Si l'intervalle est trop grand alors la présence d'un trop grand nombre de paires par intervalle bivarié provoque un lissage de $\gamma(\boldsymbol{h})$ (*voir figure 5.1.25*). Plus aucune tendance ne peut donc s'observer. Si au contraire le pas est trop petit, les cases ne peuvent plus contenir de paires. La surface n'a plus de sens dans ce cas non plus.

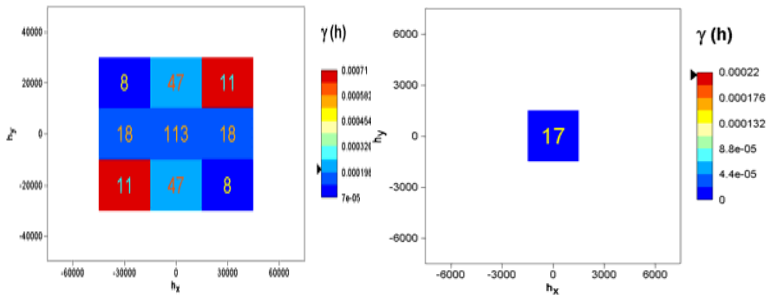


Figure 5.1.25 : illustration d'un mauvais choix du lag spacing

Aussi, dans le choix du *lag spacing,* on essaie d'obtenir une répartition la plus homogène possible du nombre de paires de valeurs.

En général deux ou trois jeux de paramètres sont utilisés pour essayer d'observer une tendance. Dans ce cas précis :

lag spacing.x = 8'000-10'000 m
lag spacing.y = 6'000-8'000 m

Le choix de cet intervalle est très important puisque l'observation d'une anisotropie peut être relative au choix du pas d'intervalle, et donc d'origine artificielle.

On remarque aussi que lorsqu'un certain nombre de valeurs est supprimé à la suite d'observation de valeurs anormales (*lors de la représentation de la nuée variographique*) le variogramme surfacique résultant peut paraître « *troué* » (*voir figure 5.1.26*). Il devient donc plus difficile d'observer une quelconque tendance. Ces trous s'observent également lorsque le pas d'intervalle est trop petit.

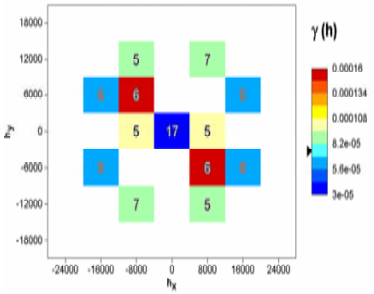


Figure 5.1.26 : variogramme surfacique « troué » : épisode 1999 avec les concentrations de XAL, CES, BJU, PLA et PED supprimées

Pour tous les épisodes une tendance directionnelle est observée. L'orientation de la direction principale d'anisotropie décrivant la continuité maximale de la variable[28] est grossièrement mais systématiquement orientée Est – Ouest (*voir figure 5.1.27*). Cette direction est à l'observation la plus évidente pour les variogrammes 1998, 2000 et 2001. Il s'agit des mêmes épisodes qui présentaient une structuration visuellement claire de la nuée variographique.

[28] *La continuité maximale est la direction selon laquelle les variations sont minimales*

Les épisodes 1996 et 1997 sont graphiquement moins caractéristiques d'une situation anisotrope. La recherche et le calcul de variogrammes directionnels du chapitre suivant permettent toutefois d'affirmer qu'une structuration directionnelle existe vraiment pour ces séries. Ceci permet de mettre en évidence une caractéristique importante de l'analyse variographique : il s'agit d'une démarche exploratoire qui peut parfois être itérative.

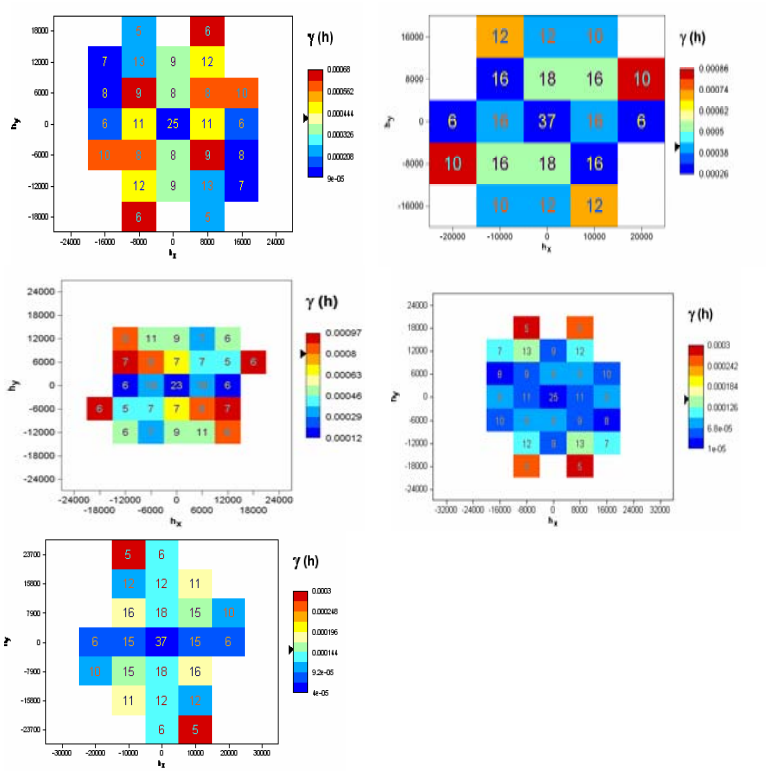


Figure 5.1.27 : variogrammes surfaciques (a) 1996, (b) 1997, (c) 1998, (d) 2000, (e) 2001

Variogramme (omni)directionnel

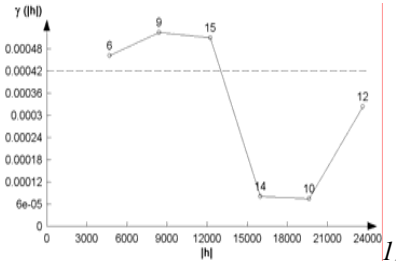
Un variogramme expérimental étudie la variation spatiale dans une direction donnée lorsque le phénomène présente un comportement anisotrope (*directionnel*), ou dans toutes les directions le cas échéant (*omnidirectionnel*). Il se construit généralement en fonction des conclusions tirées de la représentation bivariée et constitue la base de l'ajustement du variogramme modélisé (*ou théorique*).

Les paramètres à ajuster se font sur la base du calcul du variogramme omnidirectionnel. Il s'agit de la distance maximale *hmax*, à peu près identique à celle fixée pour la nuée variographique (≅ *25'500 m*), le pas d'intervalle (*lag spacing*) et le nombre d'intervalles (*nb.lag*), sur la base des mêmes critères que pour le variogramme surfacique (*pas minimum, lissage, absence et répartition des paires à l'interieur des intervalles*). On considère généralement qu'un nombre de 20 à 50 paires par intervalle est suffisant. Il faut finalement ajuster encore la direction d'anisotropie, identifiée par le variogramme surfacique, la tolérance angulaire (*permet une certaine rigueur dans la direction de recherche*) et la largeur de bande (*limite l'amplitude du vecteur* ***h*** *entre les paires dans la direction choisie*).

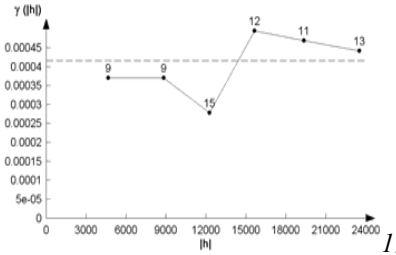
Pour tous les épisodes, la tolérance angulaire choisie est de ±35°, et la largeur de bande est infinie. Dans l'absolu il serait préférable de prendre un angle de tolérance le plus faible possible, mais le faible nombre d'observations de l'échantillon (*19 stations seulement*) indique une valeur plus grande pour laquelle le nombre de paires par intervalle reste acceptable (*dans le cas présent 10 à 20 paires*). La quantité limitée d'observations provoque également de fortes différences entre deux variogrammes dont les paramètres seraient légèrement différents. Il ne s'agit donc pas de se fier à une seule direction mais réellement à un intervalle de direction sur lequel le variogramme serait stable. Le pas d'intervalle est compris entre 3'800 et 4'000m. Le nombre d'intervalles est égal à 6-7.

Les variogrammes expérimentaux résultants (*voir figure 5.1.28*) présentent tous une évolution croissante de la variation, ce qui est logique puisque les valeurs sont corrélées en fonction de la distance.

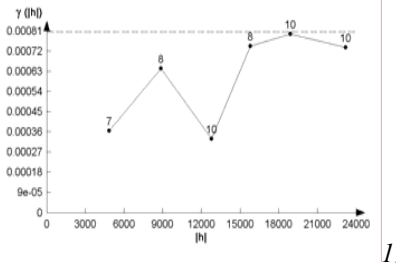
Le variogramme expérimental calculé pour 1998 (*175°*) est particulièrement stable à ±5°, malgré le faible nombre de valeurs. Idem pour le variogramme de l'épisode 2000, malgré la petite inflexion pour les paires de valeurs séparées de 17.5 à 21 km. Cette inflexion est causée par l'influence de la station SAG (= *0.074 ppm*).



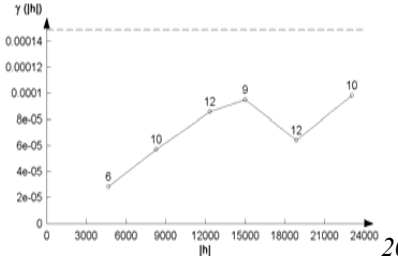
1996, 140°



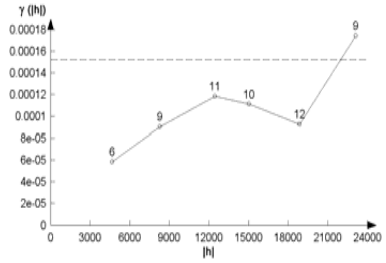
1997, 155°



1998, 175°



2000, 160°



2001, 5°

Figure 5.1.28 : variogrammes expérimentaux directionnels (a) 1996 – 140°, (b) 1997 – 155°, (c) 1998 – 175°, (d) 2000 – 160°, (e) 2001 – 5°

Seul l'épisode 1996 présente une évolution différente. Une variation très importante pour des distances entre paires très petites est repérée. Dans ce cas précis, il s'agit comme pour l'épisode 1999 d'un valeur ou deux qui seraient très différentes (*tout en étant acceptables dans l'absolu*) du reste de la série, plus particulièrement pour des paires de mesures proches. Cette affirmation est prouvée par la suite de l'évolution du variogramme, à savoir des variances qui chutent fortement à nouveau, se stabilisent puis remontent. La recherche de cette valeur ou paires de valeurs qui entachent le calcul du variogramme a été réalisée à l'aide des histogrammes bivariés (*h-scatterplot, voir figure 5.1.29*) ainsi que de la cartographie des stations (*voir figure 5.1.30*). La station plus particulièrement impliquée est Lagunilla, qui mesure la plus petite valeur de la série (*0.033 ppm*). Ce biais sera donc ignoré lors de la modélisation du variogramme.

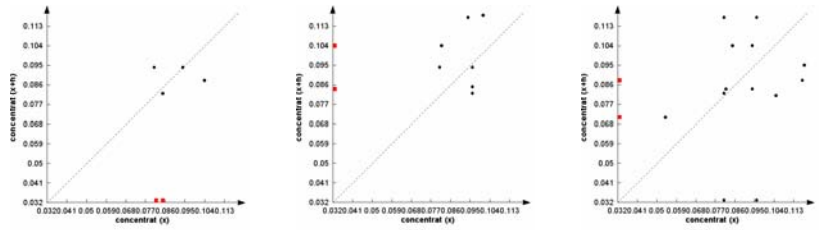


Figure 5.1.29 : h-scatterplot pour les intervalles 1, 2 et 3 du variogramme directionnel 140° pour l'épisode 1996

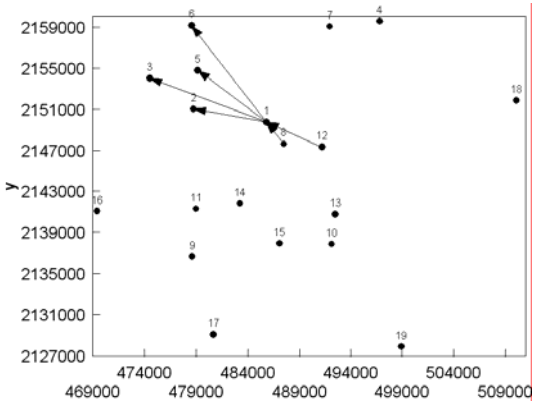
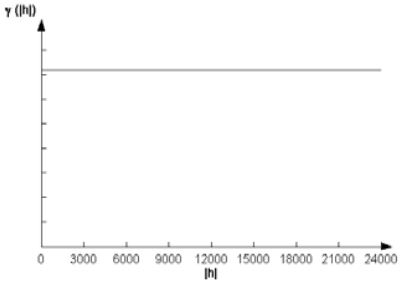


Figure 5.1.30 : repérage de la station Lagunilla

Le premier intervalle (*0 à ~4000 m*) ne contient jamais de paires, ce qui rend l'estimation de la pépite difficile. Comme mentionné précédemment une erreur initiale due à des erreurs systématiques de mesure ou aux variations à très petite échelle du phénomène est sans doute négligeable dans le cas de l'ozone, compte tenu de la grande fiabilité des données diffusées par l'INE.

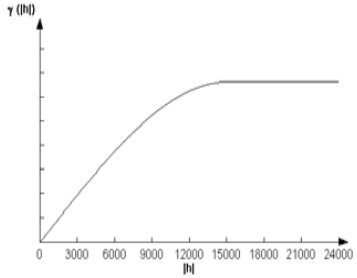
Modélisation du variogramme
L'ajustement d'une fonction théorique de covariance produit un variogramme dit modélisé $\gamma^*(\boldsymbol{h})$ dont les paramètres conduisent à l'interpolation de la variable d'intérêt. Les principales fonctions d'ajustement utilisées figurent ci-dessous, où *c* est le seuil et *a* la portée.

« nugget » (pépite) : $\gamma(h) = \gamma(0) = const.$

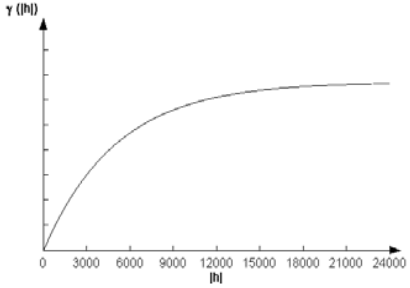


Sphérique : $\gamma(h) = c \times [B]$

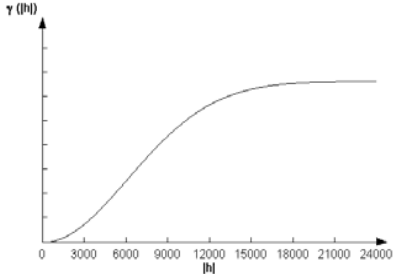
où $B=1$ si $h \geq a$
et $B = 2/3(h/a)-1/2(h/a)3$ si $h \leq a$



Exponnentiel : $\gamma(h) = c \times [1-exp(-3h/a)]$



Gaussien :$\gamma(h) = c \times [1-exp(-(3h)^2/a^2)]$



Le modèle gaussien est particulièrement bien adapté à la modélisation de phénomènes extrêmement continus, il semble donc particulièrement bien adapté à la modélisation des concentrations d'ozone. Ce modèle est donc utilisé pour l'ajustement de l'ensemble des variogrammes expérimentaux.

Lors de la modélisation, il est également possible de combiner linéairement plusieurs fonctions d'ajustement (« *nested model* »), par exemple lorsque les données présentent un changement brusque de pente, indiquant le passage à une nouvelle structuration des valeurs dans l'espace. Ce type de modélisation n'a pas trouvé raison ici. D'une part la faible quantité de paires à disposition fait qu'il semble imprudent de faire des ajustements trop sophistiqués, et d'autre part de tels changements n'ont pas été observés parmi les différents épisodes. Il aurait également été possible de combiner le modèle gaussien avec un effet de pépite, mais l'estimation de celle-ci (*erreurs de mesures + variation à micro-échelle du phénomène*) paraît trop risquée.

De manière plus générale, les superpositions de modèles permettent une modélisation plus juste. Il est cependant généralement recommandé de ne pas « *surmodéliser* » un variogramme expérimental. En effet, la signification physique des structures corrélées perdent une partie de leur sens lorsque ces structures sont le résultat de la superposition de modèles.

La modélisation a été réalisée à l'aide de Geostatistical Analyst® (extension ArcGIS 9™) afin de pouvoir introduire les paramètres de VarioWin 2.2® de la manière la plus complète possible dans un SIG pour l'interpolation.

Validation croisée

Les différents modèles sont évalués à l'aide de la démarche de validation croisée. Le modèle d'autocorrélation spatiale est construit sur la base de l'ensemble des observations. Pour valider le modèle, chacune des observations est retirée puis calculée une à une à l'aide du modèle estimé. La nouvelle estimation est ensuite comparée à la mesure actuelle. Le résultat devrait s'approcher le plus possible d'une droite de pente 1:1, compte tenu du fait que le calcul de la nouvelle estimation se sert de tous les points présents (*ce que la procédure d'interpolation restreint en général pour n'impliquer que les paires de points séparés d'une distance maximale, au-delà de laquelle les mesures ne devraient plus s'influencer*), ainsi que du fait que l'interpolation par krigeage a généralement tendance à sousestimer les valeurs importantes et surestimer les valeurs faibles, ce qui implique une pente légèrement plus faible (*voir figure 5.1.31*). La validation croisée est donc un bon outil qui élimine les résultats faux, mais qui ne permet pas de choisir entre deux « *bons modèles* ». Il pourrait en effet être préférable de se référer à des connaissances plus poussées sur le phénomène, ainsi que mettre en relation les résultats obtenus avec d'autres paramètres d'influence de la distribution afin de valider un modèle. Pour l'ozone troposphérique il s'agirait notamment des courbes de niveau, la couverture du sol, les vents de basse altitude, etc.

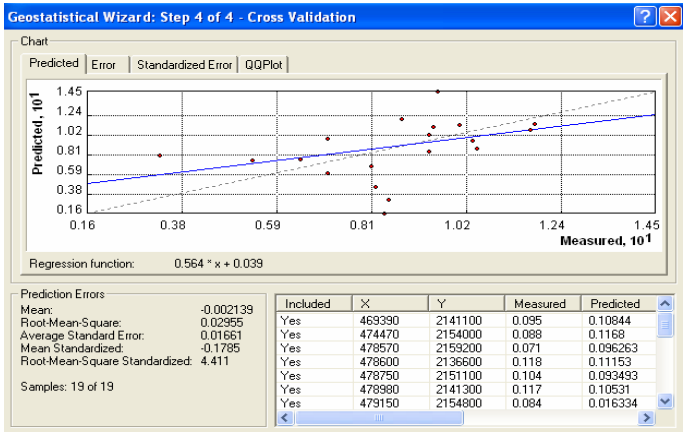


Figure 5.1.31 : exemple de validation croisée, cas d'application au modèle de l'épisode 1996

De manière générale quatre parmi les cinq modèles sont jugés « *valides* ». Il s'agit des modèles des épisodes 1996, 1997, 2000 et 2001. La validité est quantifiée par le calcul que l'angle de la droite 1:1 (*modèle parfait*) fait avec la droite de régression sur les nouvelles valeurs estimées (*voir tableau 5.1.8*).

Tableau 5.1.8 : mesure de validité des modèles

modèle	***mesure de validité***
1996	*18°*
1997	*24°*
1998	*35°*
2000	*25°*
2001	*28°*

Dans le cas présent cette validation tient compte de paires situées à plus de 50 km d'écart pour calculer une autocorrélation qui devrait s'éteindre à ~20-25 km !
Malgré cette remarque, le modèle d'autocorrélation pour 1998 reste plus ou moins incertain. Le manque de précision de la modélisation de l'épisode 1998 proviendrait éventuellement de la présence d'une ou deux valeurs provoquant des variations extraordinaires (*grandes ou faibles*) dont il a été question au tableau 5.1.7.

5.1.3.3.7. Résultats :

Vérification de la validité des épisodes
A cette étape du travail la question qui se pose est celle de la pertinence de chacune des situations représentée parmi les deux mois de maximum desquels les épisodes ont été sélectionnés (*rappel : un épisode est l'estimation d'une journée de mesures horaires par la médiane*). Cette démarche est analogue à celle du chapitre « *Représentation préliminaire des données* ». Afin de répondre les calculs suivants ont été appliqués aux épisodes 1996, 1997, 1998, 2000 et 2001 :

1. calcul d'une matrice des écarts entre médianes journalières (des deux mois de mesures) et médiane journalière sélectionnée (*celle dont les écarts à la médiane sur l'ensemble des 60 jours sont minimaux*)

2. observation pour quelques cas de la distribution des écarts et calcul de la médiane des écarts (*la distribution présente un grand nombre de petits écarts, les grands écarts sont de manière générale beaucoup moins fréquents*).

3. représentation des concentrations de la journée sélectionnée par rapport à la médiane des écarts, afin d'évaluer quelle en est leur proportion :

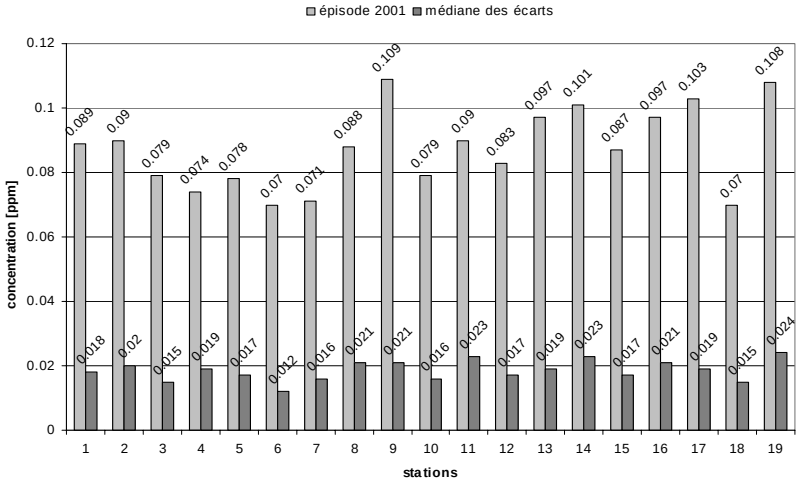


Figure 5.1.32 : représentation de la validité des épisodes

Ce graphique signifie par exemple pour la station n°1 (*Lagunilla*) que la concentration journalière sélectionnée pour la représentation de l'épisode 2001 est de 89 ppb et que cette valeur oscille approximativement de plus ou moins 18 ppb sur l'ensemble de la période de maximum considérée.

4. normalisation des deux séries par l'amplitude, ce qui permet de mieux nuancer la correspondance :

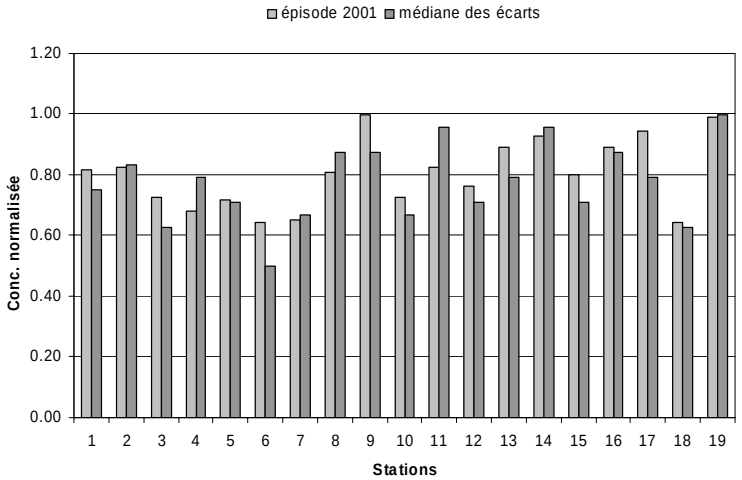


Figure 5.1.33 : représentation normalisée de la validité des épisodes

Les écarts entre les deux séries varient de 0 à 0.15 unités.

5. cumul des écarts et des concentrations sur les stations et calcul de la proportion de l'écart médian cumulé par rapport à la valeur de concentration cumulée de la journée sélectionnée. Cet indice synthétique porte peu de sens. Il permet cependant de comparer les épisodes entre eux. Pour les séries 1996, 1998, 2000 et 2001 cette proportion est de 19 à 24%. Seul l'épisode 1997 présente une proportion d'écarts de 31%. Cette valeur signifie que la représentation de cet épisode est relativement la moins pertinente.

Eléments d'interprétation

La représentation spatiale du comportement de variables de pollution atmosphérique se fait généralement à l'aide de modèles eulériens tridimensionnels transport-chimie. Le but de ce type de modélisation est la résolution des équations du transport, de la transformation chimique et de la diffusion des polluants. Le résultat est la représentation pour un intervalle de temps défini des évolutions temporelle (*pas de temps horaire*) et spatiale du comportment des polluants. Cet intervalle sélectionne une succession de 24 à 48 heures de valeurs maximum lors d'une campagne de mesures de terrain réalisée durant les mois de forte concentration. Le choix de cet intervalle est relatif à certains des objectifs poursuivis par ce type de modélisation, à savoir :

- la compréhension des différents processus physico-chimiques responsables de la formation de la pollution
- la détection dans l'espace et le temps de seuils de valeurs d'alarme, permettant la mise en place de mesures d'urgence
- la prédiction de ces valeurs

Le temps est, dans ce travail, la variable qui différencie les résultats de la modélisation photochimique des cartes interpolées géostatistiquement. D'une part, la journée représentée (*sélectionée parmi les mois de forte concentration*) agrège des mesures horaires sous la forme d'une médiane, d'autre part la procédure de sélection et d'agrégation est réalisée sur chacune des années de mesure à disposition. Les objectifs sont également différents : la cartographie d'une situation représentative d'une période de maximum, ceci à l'aide d'une seule variable (*la concentration en ozone*), permet un suivi temporel à plus grande échelle, d'année en année, afin d'évaluer la chronicité des valeurs atteintes ainsi que des surfaces touchées, sous résèrve de mesures comparables d'une année à une autre.

Malgré cette différence importante, l'utilisation des résultats de la modélistaion déterministe permet ici une meilleure interprétation ainsi qu'une sorte de « *validation* » de la cartographie obtenue pour les différents épisodes, celle-ci étant le résultat d'une analyse purement statistique.

Le comportement de la pollution atmosphérique de la ZMVM est étudié par le Laboratoire de Pollution Atmosphérique et des Sols (*LPAS*) à l'aide du modèle *TAPOM* (*Junier M., Kirchner, F., Clappier A. et van den Bergh*). Dans un premier temps le modèle à meso échelle (*FVM, Martilli A., Clappier A. et M. W. Rotach, 2002*), dont l'ajustement des paramètres est spécifique aux conditions urbaines présentes dans la ZMVM, simule les différents champs météorologiques de la troposphère (*notamment températures, vitesses et directions des vents*). Les valeurs résultantes sont ensuite introduites dans le photochimique *TAPOM*. Les paragraphes suivants présentent donc les résultats du modèle pour le 2 mars 1997.

D'un point de vue météorologique :

L'ozone étant un polluant secondaire, et par conséquent formé à la suite multiples réactions chimiques, les paramètres météorologiques (*notamment les vent de basse altitude*) sont pour une grande partie responsables de la distribution des concentrations de ce polluant. L'étude de champs de vents proches de la surface révèle une situation complexe, fortement influencée par le relief (*sytèmes mécaniques*) ainsi que par la couverture et la pente du sol (*systèmes thermiques*). Une ligne transversale Est-Ouest de convergence des vents est mise en évidence au Sud de la vallée par le modèle à meso échelle et confirmée par des mesures de terrain. Cette démarcation est en partie expliquée par la combinaison de deux types de vents :

1. un vent de vallée typique, pénétrant la vallée par une ouverture dans le relief au SSE, située à ~ 2'500-2'700 m.
2. un vent géostrophique orienté NS, provoqué par la situation géographique du Mexique

Le panache de pollution devrait à priori être fortement influencé par cette situation complexe (*voir figure 5.1.34*). De plus, cette configuration des vents ne présente pas de caractère saisonnier et devrait par conséquent être valable aussi pour les mois d'avril et mai.

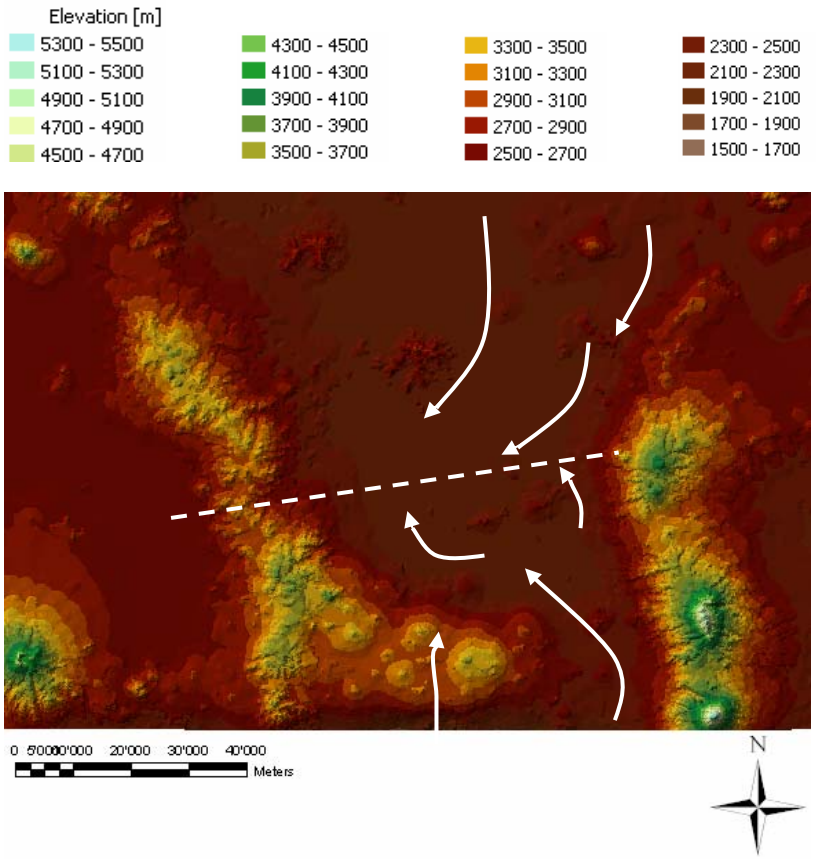


Figure 5.1.34 : relief Vallée de Mexico, triangulation de courbes de niveau équidistantes de 100m, représentation des vents et de la ligne de convergence

D'un point de vue comportement des polluants:

Rappelons que l'ozone est un polluant secondaire dont la formation peut être expliquée par la réaction suivante :

$$VOC + NO_x \xrightarrow{h\nu} O_3$$

Les NOx (*NO, N_2O_5, NO_2*) sont un ensemble de polluants dits primaires. Leur présence est provoquée par émission directe. Ils sont les précurseurs de la formation d'ozone. Le secteur des transports est la principale source d'émission pour la vallée de Mexico (*à plus de 82% en 2000*). Cependant lorsque l'ozone est à proximité directe des NOx, de jour on assiste à une décomposition chimique des molécules selon la réaction suivante :

$$NO + O_3 \leftrightarrow NO_2 + O_2$$

En effet lorsque la concentration en monoxyde d'azote augmente fortement, O_3 est très vite dissocié. En revanche, lorsqu'il est en concentration moyenne le VOC sert de carburant dans le mécanisme suivant :

$$NO_2 + O_2 \leftrightarrow O_3 + NO$$
$$NO + VOC \leftrightarrow NO_2 + VOC'$$

Le NO recycle donc le NO_2.

Ces réactions mènent à penser qu'une source importante d'émissions de NOx pourrait éventuellement expliquer l'apparition de surfaces où l'ozone est en faible concentration. Or les sources pour la ZMVM sont linéaires et ramifiées. Le réseau routier y est très dense. Il paraît donc à priori plus difficile d'identifier ce type de phénomènes sur des cartes de cette échelle (*par opposition avec des sources ponctuelles par exemple*).

La concentration obtenue de dioxide d'azote au 2 mars 1997 à l'aide du modèle photochimique *TAPOM* est présentée à la figure 5.1.35. Les coordonnées sont définies localement (*X=[0 ;200], Y=[0 ;280]*)[29]. Ces images permettent de localiser les principales sources d'émission ainsi que l'heure et la durée du pic. Les heures représentées sont donc choisies en conséquence. Les deux maxima apparaissent à 10h et 20h, au centre Ouest de la vallée.

[29] *Les représentations issues de modèles photochimiques seront géoréférencées par la suite afin de pouvoir être comparées aux modélisations géostatistiques.*

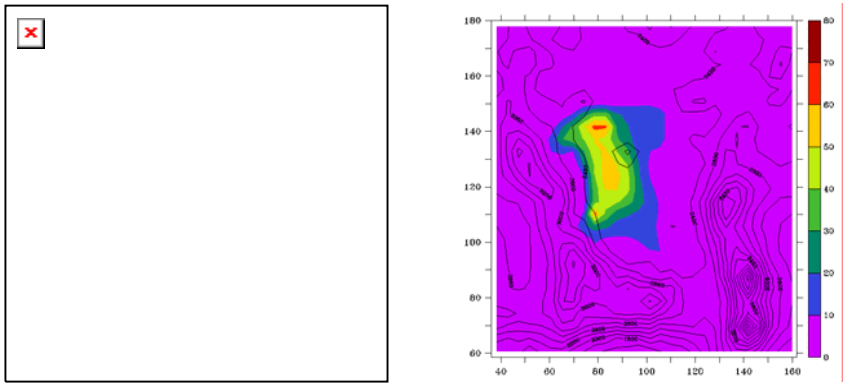


Figure 5.1.35 : concentrations horaires de NO_2 au 2 mars 1997 pour 10 et 20 heures

Une évolution plus détaillée de la concentration d'ozone au 2 mars 1997 est présentée à la figure 5.1.36. Les concentrations sont également obtenues par l'utilisation de TAPOM. L'intervalle représenté correspond à l'intervalle de temps sur lequel les données ont été agrégées pour ce travail (*12 à 18 heures*)

Sur la figure 5.1.36 Les concentrations maximales atteintes se situent entre 140 et 150 ppb. La zone de pic semble se former entre 13 et 14 h en $(X_0,Y_0) = (100,120)$. Le panache se déplace ensuite transversalement d'Est en Ouest, selon l'axe de convergence présenté à la figure 5.1.34, pour venir se plaquer contre les versants ~ *NE* de la vallé en $(X,Y) = (60,120)$.

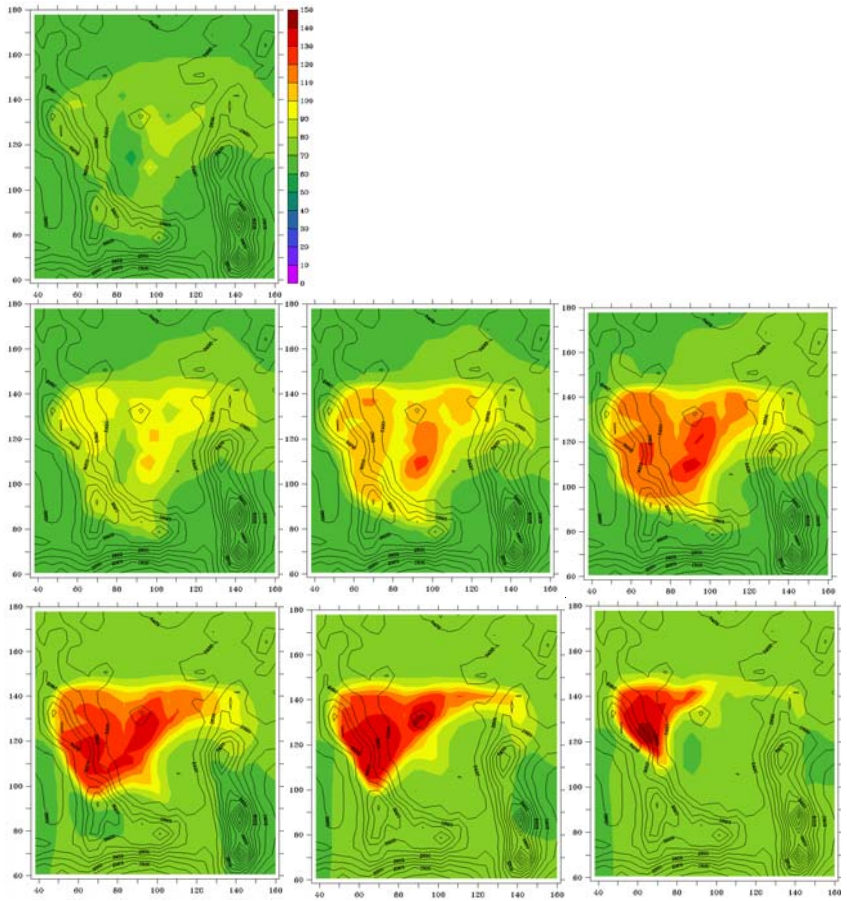


Figure 5.1.36 : évolution horaire des concentrations d'ozone au 2 mars 1997 de 12 à 18 hueres

Description des cartes

La classification des concentrations est la même pour tous les épisodes. Elle permet par conséquent la comparaison des cartes résultantes. La méthode retenue découpe l'ensemble des valeurs (*tous épisodes confondus*) en 10 classes dont les bornes sont définies par les quantiles *q10%* à *q100%*. Ce choix est justifié à la suite de multiples essais par une meilleure répartition du nombre de valeurs dans chacune des catégories. La classification des incertitudes est également réalisée sur l'ensemble des valeurs, toujours afin de pouvoir comparer les cartes entre elles, mais elle découpe les épisodes par intervalles équidistants. Cette échelle de classification paraît plus appropriée à l'interprétation des incertitudes.

concentration [ppb]
29 - 62
62 - 71
71 - 79
79 - 86
86 - 90
90 - 95
95 - 97
97 - 105
105 - 116
116 - 141

erreur [ppb]
0 - 4
4 - 8
8 - 12
12 - 15
15 - 19
19 - 23
23 - 27
27 - 31
31 - 35
35 - 38

La zone interpolée se situe au centre-Sud de la ZMVM. Elle est représentée à la figure 5.1.37.

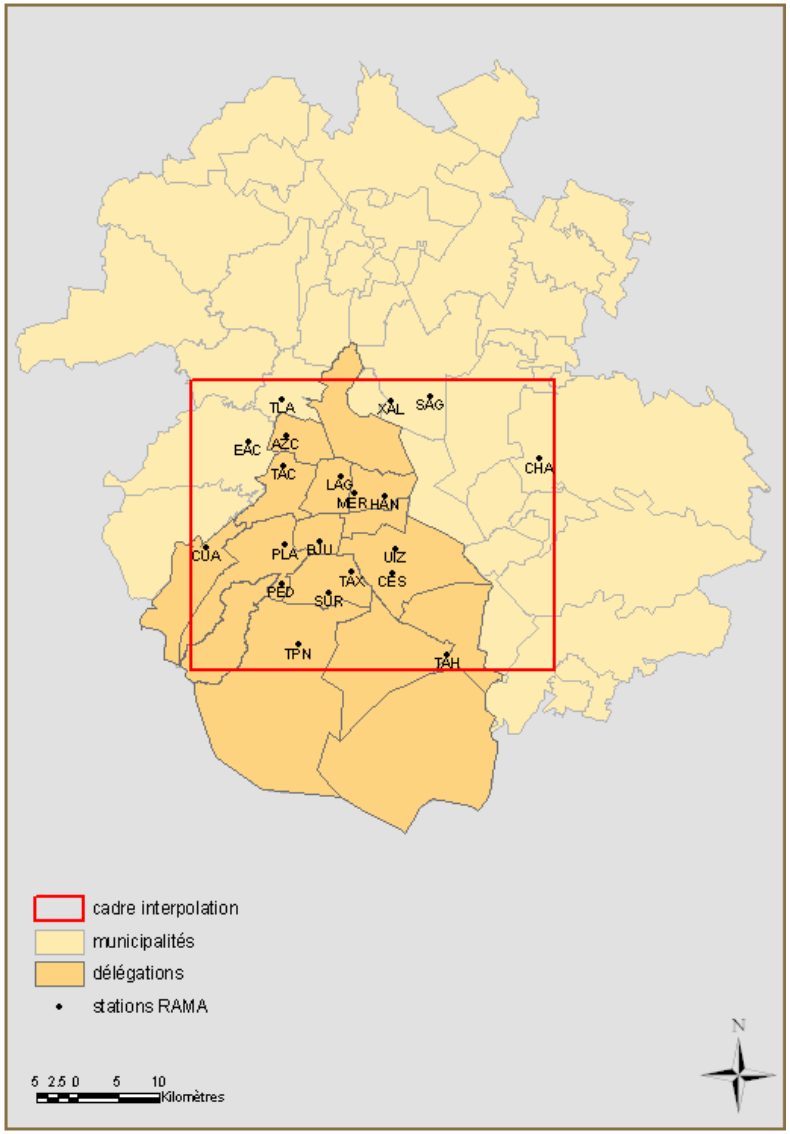


Figure 5.1.37 : ZMVM, localisation de la zone interpolée

Une vue d'ensemble des interpolations est présentée ci-dessous :

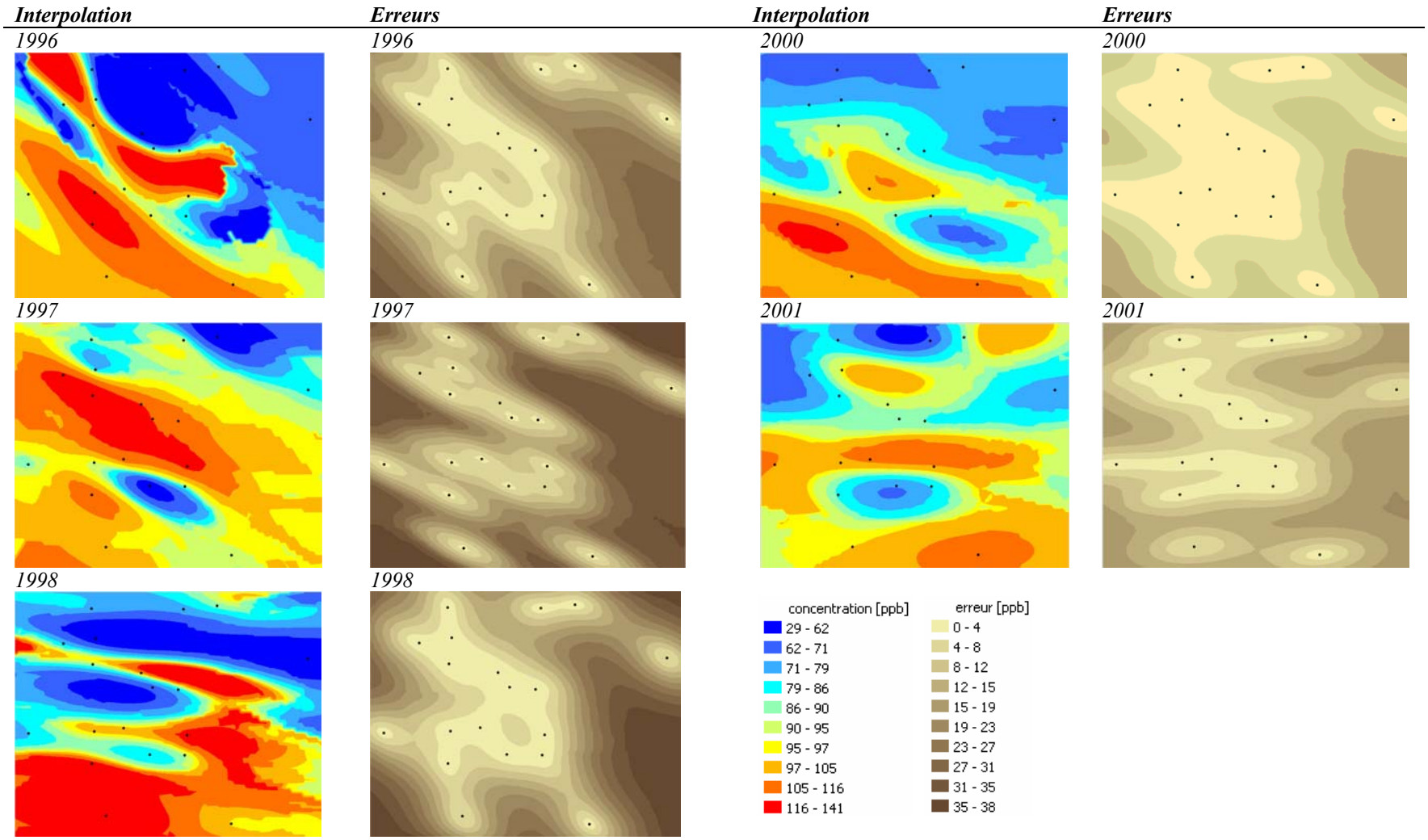


Figure 5.1.38 : vue d'ensemble des épisodes interpolés

Cette vision globale permet quelques observations générales sur les concentrations interpolées. Premièrement la variabilité spatiale interannuelle est importante. Ensuite quelques observations systématiques peuvent toutefois être mises en évidence :

1. des structures spatiales longitudinales, orientées *EO,* par exemple pour 1998:

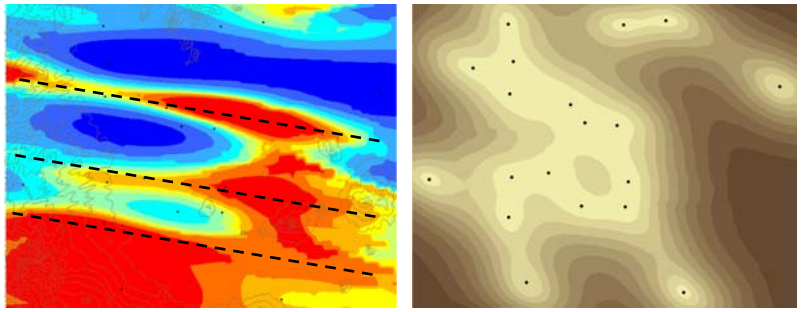


Figure 5.1.39 : interpolation de l'épisode 1998 et carte des incertitudes, illustration des structures longitudinales

2. les concentrations les plus importantes sont localisées au *SO* de la zone interpolée (*Pedregal et Tlalpan*), et les concentrations les plus faibles sont au *NE* (*Chapingo et San Augustin*). Cette situation est bien illustrée par l'épisode 2000. L'épisode 2001 est de ce point de vue un peu atypique.

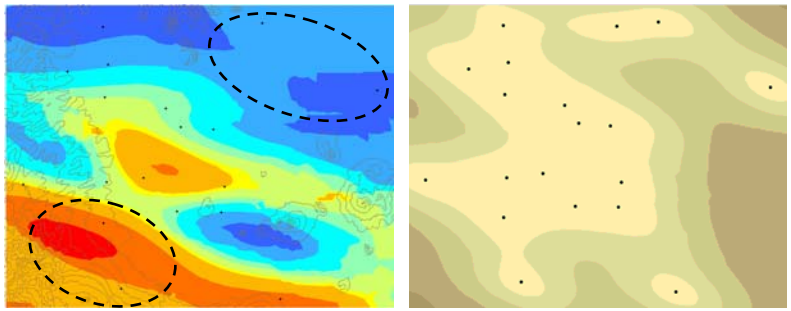


Figure 5.1.40 : interpolation de l'épisode 2000 et carte des incertitudes, illustration des zones de concentrations min et max

3. la présence d'un double panache séparé d'une zone circulaire de faibles concentrations, exemple des épisodes 1997 et 2001.

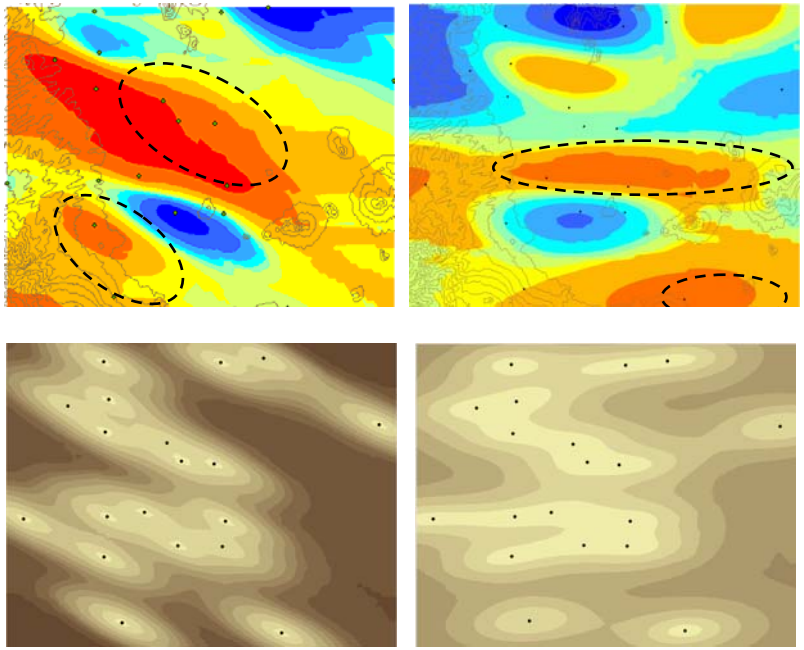


Figure 5.1.41 : interpolation des épisodes 1997 et 2000 et carte des incertitudes, illustration de la présence d'un double panache

Les deux stations TAX (*Taxqueña*) et CES (*Cerro de la Estrella*) se retrouvent systématiquement dans la zone de faible concentrations.

4. pas/peu d'effet de puits ni de source.

5. Quelques remarques générales concernant les cartes des erreurs peuvent également être faites. En effet, les épisodes 1996 et 1997 présentent une directionnalité importante. Les incertitudes pour ces deux années sont également les plus importantes puisqu'elles atteignent ~30-40 ppb. Les épisodes 2000 et 2001 présentent en revanche des erreurs maximales relativement faibles (*10 à 20 ppb*).

6. on remarque aussi quelques artefacts de l'interpolation, certainements dus au trop faible nombre de stations de mesure.

Interprétation

L'étape d'interprétation des cartes essaie d'une part d'évaluer le « *réalisme* » de ces modèles géostatistiques par comparaisons avec les résultats de modèles photochimiques, d'autre part de fournir des explications visuelles complémentaires (*par superposition de couches géographiques*) quant au comportement du phénomène sur la base des observations ci-dessus.
Aussi une incertitude de plus ou moins 20% sur la concentration est admise. Il n'est donc pas recommandé d'interpréter au-delà de cette valeur.

Dans un premier temps les structures spatiales longitudinales d'Est en Ouest sont certainement liées à la ligne de convergence des vents décrite plus haut. Un parallèle peut être également être fait avec le déplacement transversal de la zone de maximum lors de la simulation d'un épisode type par le modèle *TAPOM* (*voir figure XXX*).

Ensuite, les simulations de modèles photochimiques concordent généralement bien avec les zones de maximas (*Sud-Ouest : Pedregal – PED et Tlalpan – TPN*) et minimas (*Nord-Est : Chapingo – CHA et San Augustin – SAG*), sauf pour l'épisode 2001 qui semble être un peu particulièr de manière générale. Ces zones présentent des estimations de l'imprecision faibles (*~0-10 ppb*) puisqu'elles se trouvent à proximité directe de stations de mesure.

La distribution spatiale des concentrations semble être fortement influencée par le relief (*voir figures 5.1.42 et 5.1.43*), ce qui peut être expliqué au final par le transport des polluants. On peut observer une zone circulaire (*encadrés rouges*) de faibles concentrations généralement localisée sur un haut point du relief (*200-500 m*). La forte incertitude (*~ 30 ppb sur des valeurs de ~ 80 ppb*) sur la partie Est de cette zone pour l'épisode 1996 met toutefois en doute la pertinence de l'encadré.

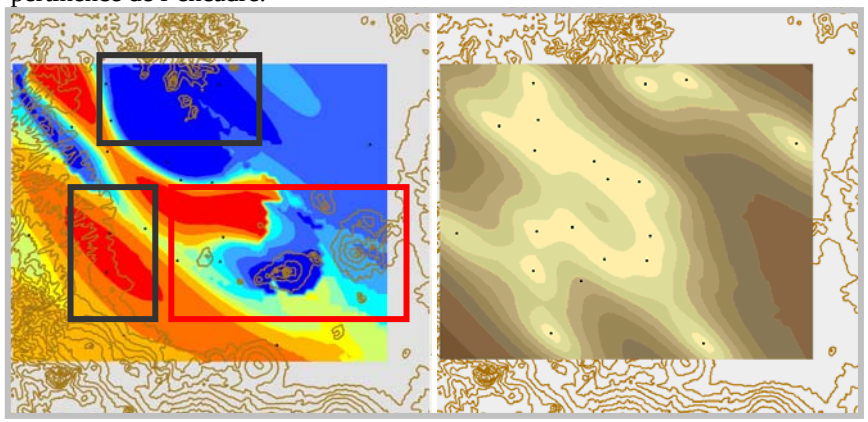


Figure 5.1.42 : épiode 1996, interpolation et carte des incertitudes, influence du relief sur la distribution des concentrations, courbes de niveau 100 m

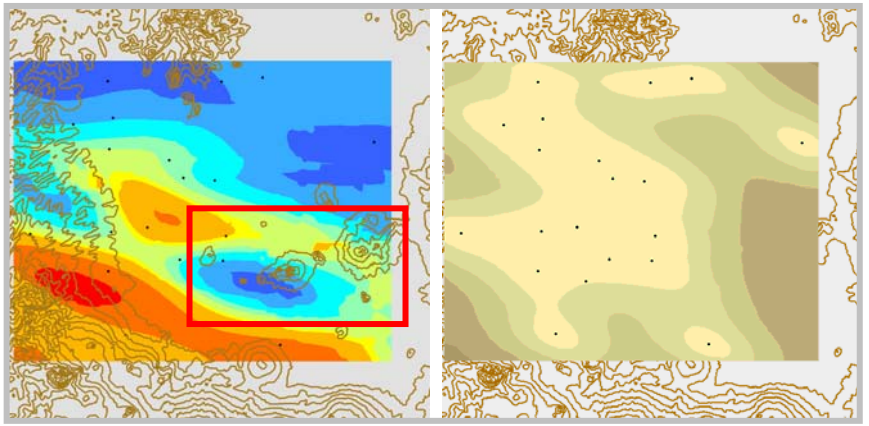


Figure 5.1.43 : épisode 2000, incertitudes et carte des incertitudes, influence du relief sur la distribution des concentrations, courbes de niveau 100 m

La zone de faibles concentrations étant plus ou moins circulaire, on pourrait également supposer la présence d'un piège à NO qui détruirait de manière instantanée toute l'ozone présente. Pour vérifier cette hypothèse une superposition des épisodes 1996 à 2001 avec les cartes de la figure 5.1.35 a été réalisée. Les deux objets (*pic de NO_2 et zone circulaire*) apparaissent séparés de plus de 30 km.

La localisation de la partie la plus importante des industries moyennes et grandes (*NNO, voir figure 5.1.44*) peut vouloir dire deux choses:

- soit un déplacement de la pollution primaire en fonction du déplacement des masses d'air vers le Sud où l'ozone peut se former
- soit une forte émission de polluants primaires provoquant la destruction de l'ozone

Une combinaison des deux hypothèses est possible.

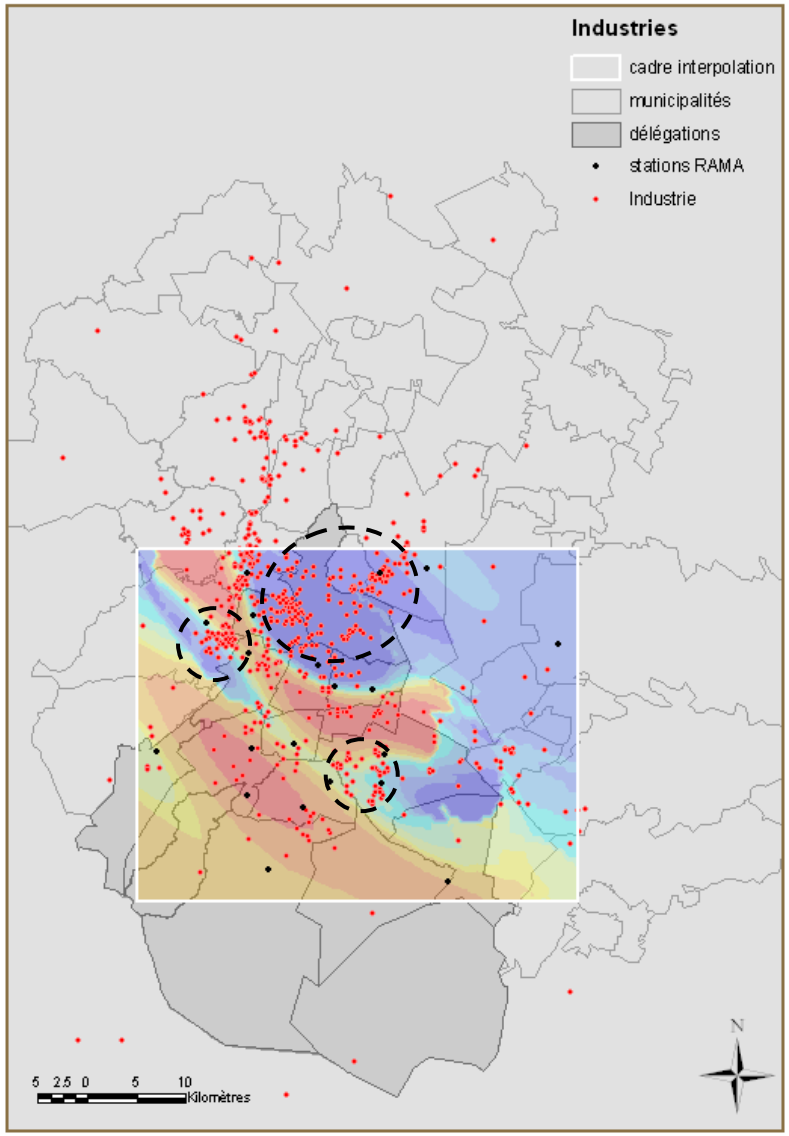


Figure 5.1.44 : localisation des grandes et moyennes industries, superposition en transparence de l'épisode 1996

.

La présence d'un double panache pose la question suivante. En admettant qu'un déplacement significatif des masses d'air se fasse à un échelle horaire (*ou moindre*) les concentrations dans l'espace de la vallée seraient certainement conduites à suivre la même allure. L'agrégation d'une journée de mesure sous la forme de la médiane poserait-elle donc un problème de « *décalage temporel* » ? La représentation d'un double panache ne serait alors qu'une question de dynamique d'un seul panache, situation généralement simulée par les modèles photochimiques. La vérification de la présence d'une éventuelle dynamique pourrait se faire visuellement par la cartographie des mesures horaires de cette même journée. Compte tenu du temps restant à disposition il a été choisi de procéder à une analyse quantitative un peu plus grossière. Il s'agit donc de sélectionner trois groupes de stations situées dans les trois zones à étudier, à savoir dans le premier panache (*A*), dans la zone de valeurs minimales entre les deux panaches (*B*), et dans le second panache (*C*) :

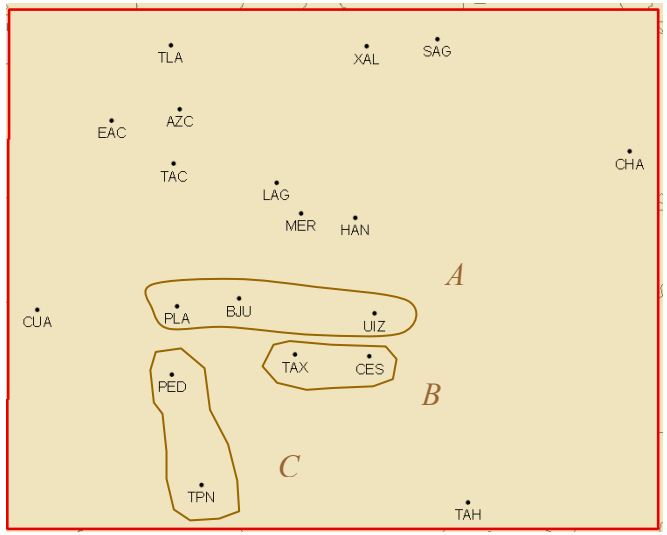


Figure 5.1.45: localisation des groupes A, B et C

La vérification se fait en calculant pour chacune des stations à l'intérieur de chaque groupe un delta d'un heure à l'autre (*de 12 à 18h*). Tous les éléments d'un groupe devraient présenter la même évolution des valeurs de delta dans le temps. Les résultats escomptés dans le cas où il y aurait un décalage temporel important devraient se présenter tels que :

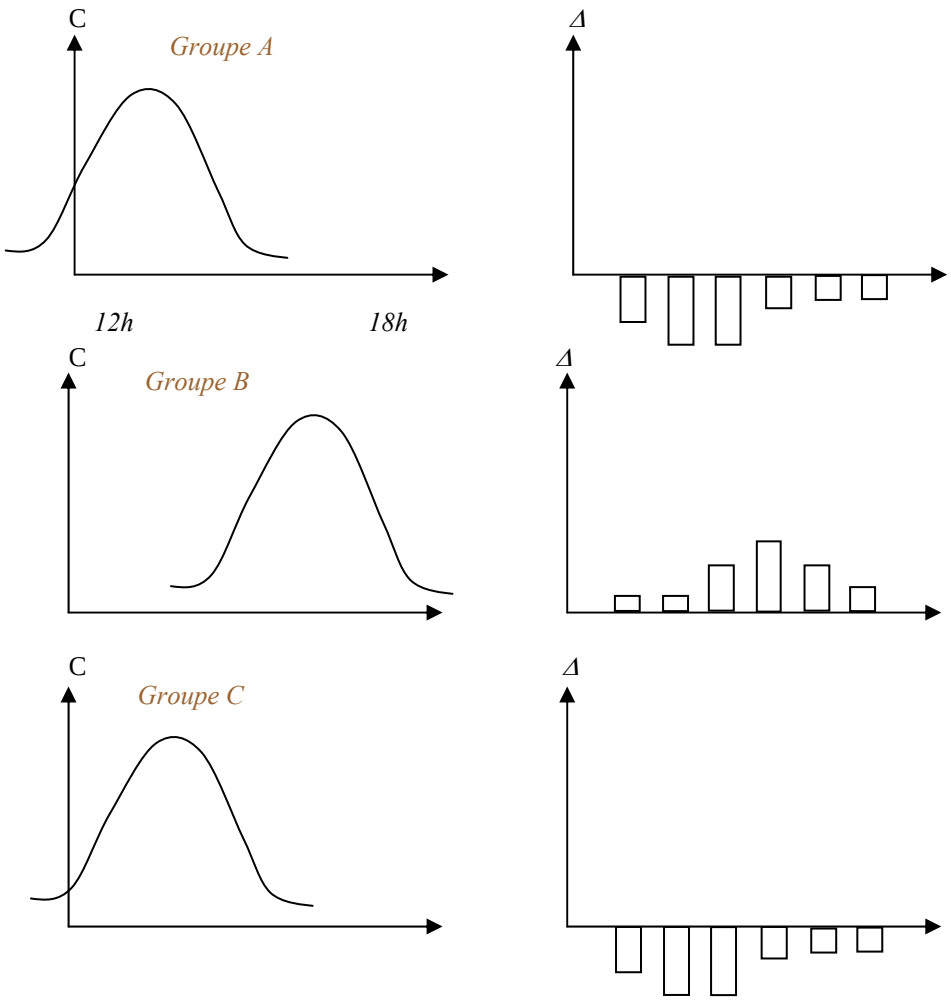


Figure 5.1.46 : résultats escomptés dans le cas de l'existence d'une dynamique temporelle sous-jascente aux épisodes représentés

Les résultats les plus caractéristiques pour les deux premiers groupes se présentent aux figures 5.1.47 à 5.1.49. Les courbes pour les groupes A et C ressemblent un peu aux résultats escomptés. Pour le groupe B la relation est moins évidente. Cette analyse ne permet pas réellement de conclure. Elle renforce cependant plus ou moins une hypothèse de décalage temporel et incite à aller chercher plus avant.

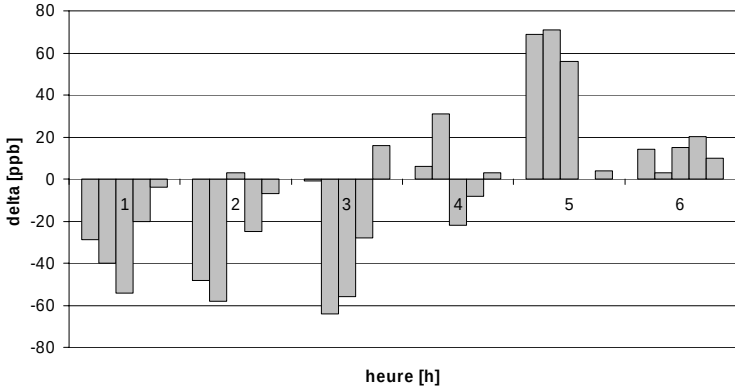


Figure 5.1.47 : station PLA, extraite du groupe A, évolution du delta en fonction de l'intervalle horaire

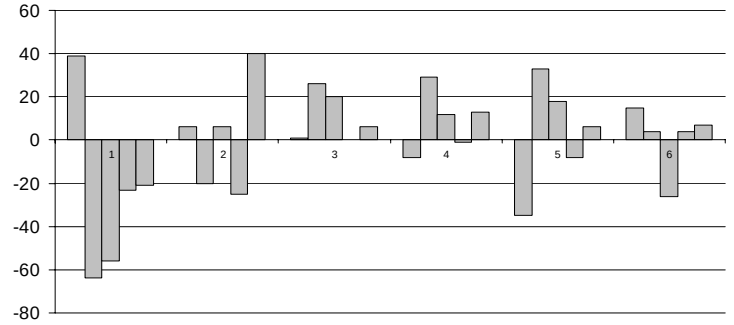


Figure 5.1.48: station CES, extraite du groupe B, évolution du delta en fonction de l'intervalle horaire

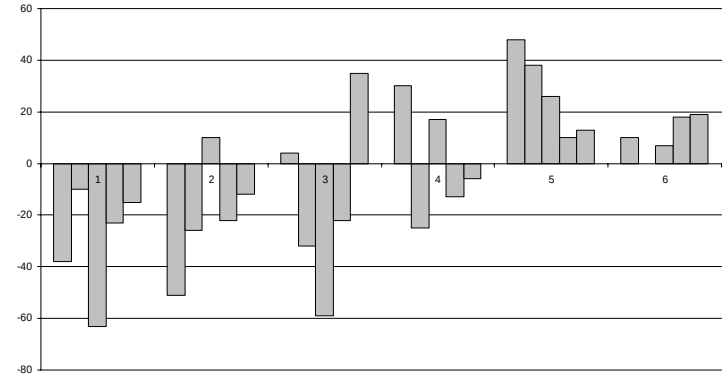


Figure 5.1.49 : station PED, extraite du groupe C, évolution du delta en fonction de l'intervalle horaire

5.2. Contexte environnemental

L'objet de ce chapitre est la prise en compte des résultats précédents dans un contexte environnemental plus général. Le but de la démarche étant de réunir l'ensemble des facteurs et enjeux impliqués par le « *danger d'ozone* », afin de pouvoir en mesurer les conséquences.

Les résultats précédents mettent également en évidence la difficulté que la modélisation géostatistique peut avoir à expliquer le comportemement de l'ozone troposphérique. En effet, sans connaissances préalables (*littérature scientifique, représentations issues des modèles transport-chimie, etc.*) il n'aurait pas été possible d'évaluer la vraisemblance des cartes obtenues autrement que « *géostatistiquement* », ce qui n'est pas envisageable compte tenu de la complexité du comportement de ce phénomène. Rappelons encore que l'ozone est un polluant secondaire, dont la concentration est fortement dépendante d'un point de vue spatial et temporel des mécanismes physico-chimiques de formation, ainsi que des conditions atmosphèriques (*notamment mouvements de masses d'air*). Cette complexité explique pourquoi il n'est pas question dans ce travail d'utiliser la modélisation géostatistique à des fins prédictionnelles.

L'idée ici est d'insérer les épisodes de « *danger d'ozone* » dans un modèle spatial descriptif de la ZMVM. Le but de ce modèle serait la gestion du risque lié à la pollution atmosphérique par l'ozone troposphérique.

5.2.1 Modélisation conceptuelle

La modélisation conceptuelle est une description conceptuelle des structures de données relatives à une application. L'application dans ce cas précis est la gestion du risque environnemental lié à la pollution de l'air par l'ozone troposphérique dans la ZMVM.

La conception d'un outil d'aide à la décision passe inévitablement par une modélisation conceptuelle de la base de données spatiale projetée. En effet, le nombre trop important des éléments liés au risque de pollution par l'ozone ainsi que la complexité des interdépendances entre ces éléments justifie cette étape. Elle permet donc d'aborder ainsi que de décrire formellement les besoins de l'utilisateur.

5.2.1.1. Description du modèle

Le but est pour l'utilisateur (*le décideur*) de « *mesurer* » le risque d'ozone sur les différents objets exposés (*population, surfaces agricoles, surfaces vertes, etc.*), compte tenu des différents facteurs d'influence (*émissions, relief, zones habitables, etc.*), et donc de prendre des mesures concrètes lorsqu'elles sont possibles. Il définit l'orientation donnée au modèle.

Le modèle présenté ci-dessous est réalisé à l'aide du logiciel Perceptory. Le langage formel utilisé est UML.

Ce modèle ne tient compte que des données disponibles collectées sur place, et n'est par conséquent pas complet. Quelques attributs supplémentaires ont toutefois été imaginés dans l'optique du modèle. Les généralisations non exhaustives sont notées « … ».

5.2.1.2. Schéma conceptuel

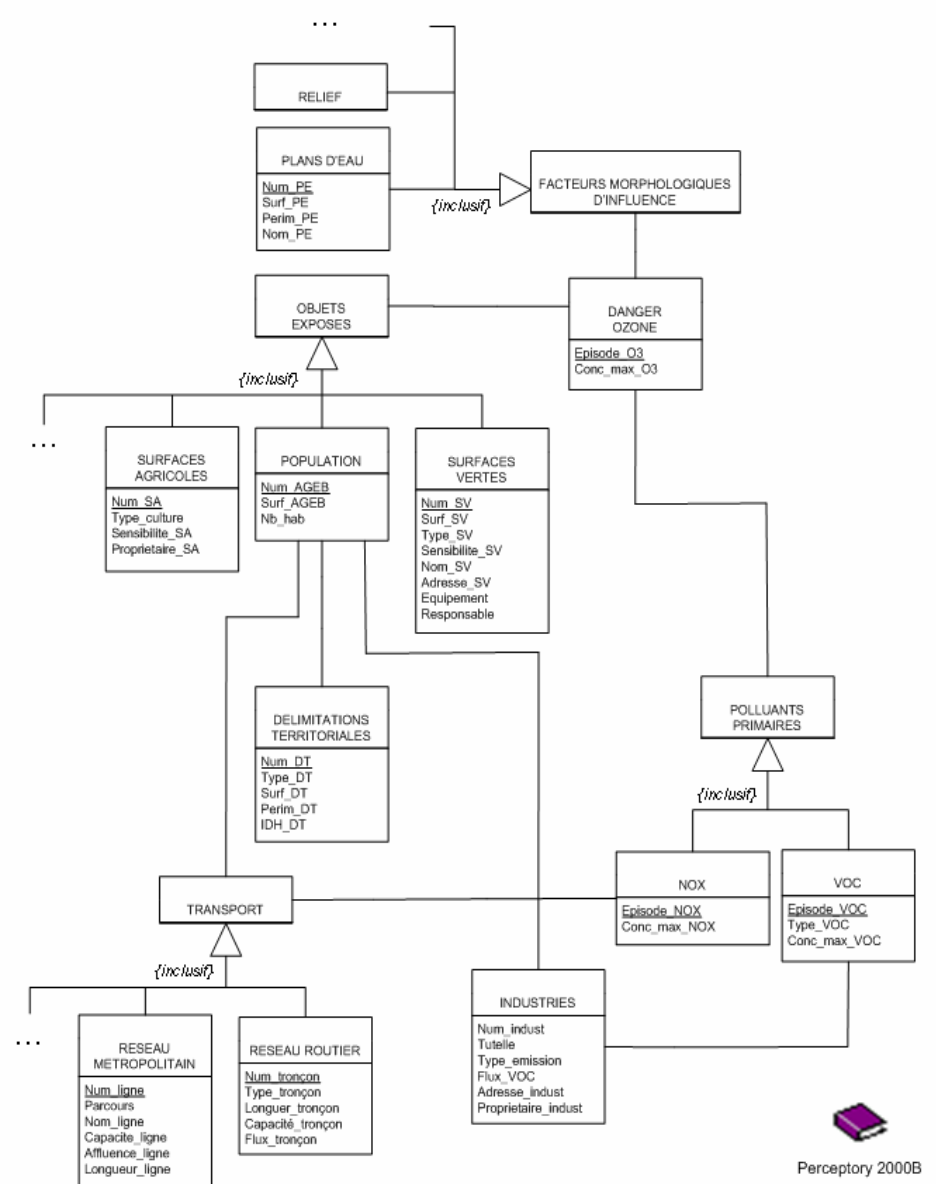


Figure 5.2.1 : schéma conceptuel

5.2.1.3. Description des objets du modèle

Les généralisations sont en gras souligné, les classes d'entité sont en gras, les attributs sont en italique, les attributs imaginés sont en italique suivis de « *(i)* » et l'identifiant en italique souligné. Les relations sont en gras et sont suivies du symbole « **(↔)** ».

Danger_O_3 et **Objets_exposes** sont les classes d'entités les plus importantes du modèle.
Chaque tuple est identifié par un *Episode_O_3*. Le second attribut est *Conc_max_O_3* , la concentration maximale atteinte pour un épisode.

Objets_exposes est le deuxième élément central du modèle. Il est une généralisation des différents objets ci-dessous. La liste est non eshaustive. Une contrainte {inclusive} s'applique à cette généralisation, ce qui signifie que **Danger_O_3** peut affecter plusieurs **Objets_exposes** à la fois.

Surfaces_agricoles : les surfaces agricoles sont affectées par une diminution du rendement des cultures, notamment par l'action oxydante de l'ozone sur les feuilles des plantes, ne pouvant par conséquent réaliser une photosynthèse que partielle. L'identifiant est *Num_SA*, un numéro de « *parcelle* ». Les autres attributs sont *type_culture*, c'est-à-dire cultures annuelles ou pluviale par exemple, *propriétaire_SA (i)*, une valeur caractérisant la sensibilité de la culture à l'ozone et *sensibilité_SA (i)*. En effet, l'ozone est le polluant atmosphérique de loin le plus nocif par son action directe sur la végétation (*BUWAL, 1999*). Pour les espèces sensibles, une brève exposition à de fortes concentrations suffit pour provoquer des altérations visibles des feuilles, tout particulièrement chez les légumineuses (*soja, haricot, etc.*). Les pollutions persistantes comme c'est le cas pour la ZMVM provoquent des troubles chroniques, une croissance diminuée, une production réduite de substances de réserve ainsi qu'une vulnérabilité accrue aux maladies. Les baisses de rendements agricoles occasionnées peuvent atteindre jusqu'à ~15% selon les cultures, la région et l'année. Il est important de dire que ces surfaces sont cependant localisées pour la pluspart au Sud de la vallée, et que le réseau de mesures ne permet actuellement pas d'interpoler de manière fiable la concentration d'ozone sur les surfaces les plus éloignées du centre.

Surfaces_vertes : les surfaces vertes sont généralement des jardins publics, parcs, bandes de végétation urbaine, etc. (*type_SV*). Elles sont identifiées par *Num_SV*. Les autres attributs permettent la dénomination et la localisation (*Nom_SV, Adresse_SV*), où encore de s'informer sur le *Responsable* ou encore l'état d'*Equipement* de la surface.

Population : cette classe est un peu complexe au sens où la population est définie par AGEB (*Area Geostadìstica Bàsica*), d'où l'identifiant *Num_AGEB*. Ces unités géostatistiques font partie de l'ensemble des démarcations géostatiques nationales. Elles ont été définies lors du dernier recensement de la populaion en 2000 par l'Institut National de Statistique, Géographie et Informatique (*INEGI*). Chaque AGEB contient 1 à 50 pâtés de maisons ou autres bâtiments ainsi qu'une population égale ou supérieure à 2'500 hab. La délimitaion est fonction des rues, avenues ou autre objet clairement défini dans le territoire. Les autres attributs sont *Surf_AGEB* et *Nb_hab*, ce qui donne au final une densité de population par AGEB.
Population est une classe importante. Elle est l'objet exposé qui a le plus d'influence sur les émissions de **Polluants primaires**, par l'intermédiaire de l'utilisation des **Transports** ainsi que par la **consommation (↔)** de produits industriels. Il s'agit également d'un objet dont la situation géographique, par l'intermédiaire de **Delimitation_territoriale**, pourrait éventuellement être contrôlée.

Delimitations territoriales : cette classe est identifiée par *Num_DT*. Les attributs sont *Nom_DT*, *Surf_DT*, *Périm_DT*, *Type_DT* et *IDH_DT*. *Type_DT* peut prendre les valeurs « *Municipalité* » ou « *Délégation* ». En effet, lorsque la zone est à l'intérieur du Distrito Federal les délimitations territoriales forment des délégations. Hors du DF, dans l'Etat de Mexico, les délimitations forment des municipalités. On pourrait également imaginer dans la base de données un découpage des zones en fonction de l'occupation du sol (*zones habitables, zones commerciales, zones industrielles, zone de services, etc.*). Une sorte de plan d'affectation existe en réalité, mais est trop récent pour être mis en pratique. La croissance fulgurante de la ville est également un facteur qui empêche l'application de ce type de loi.

Transports : le secteur des transports, notamment la circulation routière, est la principale source d'émission des précurseurs de l'ozone. Les attributs relatifs aux sous-classes de **Transports** pourraient entre autres permettre une estimation des émissions. Cette superclasse généralise de manière {inclusive} mais non exhaustive le **Reseau routier** et le **Reseau métropolitain**. En effet, ces deux classes d'objets peuvent soit **emettre (↔)** des **NOx**, soit **etre utilise (↔)** par la population. Cette super-classe pourrait également tenir compte de l'aéroport qui est une source importante d'émissions, mais pour lequel les données ne sont pas disponibles.

Reseau routier : l'identifiant est *Num_tronçon*, les autres attributs sont *Type_voie* (*voie « principale » ou « secondaire »*), *Longueur_tronçon*, *Capacité_tronçon (i)* et *Flux_tronçon (i)*. *Capacité* serait le nombre de véhicules que l'infrastructure pourrait supporter, et *Affluence* serait un fllux journalier de véhicules.

Reseau metropolitain : l'identifiant est *Num_ligne*, les attributs sont *Nom_ligne*, *Longueur_ligne*, *Capacité_ligne*, *Affluence_ligne* et *Parcours*. *Capacité_ligne* et *Affluence_ligne* sont dans le même esprit que pour le **Reseau routier**.

Industries : le secteur industriel est la principale source d'émission des **VOC**. Une grande partie des industries (*moyennes et grandes*) sont répertoriées, cependant la croissance démesurée de la ville fait qu'une grande partie encore de petites industries ne sont pas recensées. Les attributs sont *Num_ind*, *Tutelle*, *Propriétaire_ind, Adresse_ind*, *Type_émission* et *Flux_VOC.*

NOx : les **NOx** sont principalement émis (**est emis par (↔)**) par le secteur des transports. Ils font partie de la super-classe {inclusive} **Polluant_primaire**. L'objet est de type surfacique. Les tuples sont identifiés par un *Episode_NOx*. *Conc_max_NOx* est la concentration maximale atteinte au cours d'un épisode.

VOC: les **VOC** sont principalement produits par le secteur industriel. La classe possède le même type d'attributs que les **NOx**. Un seul les distingue : *Type*_VOC. Il existe en effet un grand nombre de VOC différents impliqués dans la formation d'ozone. Cette classe fait également partie de la super-classe {inclusive} **Polluant_primaire.**

Polluant_primaire : les polluants primaires **NOx** et **VOC** sont **précurseur de (↔) Danger_O_3**. Polluant primaire regroupe des objets surfaciques identifiés

Facteurs_morphologiques_d'influence : les facteurs d'influence devraient être « *atemporels* », contrairement aux paramètres météorologiques par exemple. En effet, les paramètres météorologiques à disposition (*température et humidité relative*) évoluent sur une même échelle de temps que l'ozone. Or l'intégration de la variable temps étant déjà complexe pour l'ozone, il n'est pas envisagé dans ce travail de procéder de la même façon pour la température et l'humidité relative. Cette généralisation est {inclusive} et non exhaustive.

Relief : le relief a ici une forte influence sur le déplacement des masses d'air proches du sol et par conséquent de l'ozone. Les attributs sont *Num_courbe* et *altitude*.

Plans d'eau : les plans d'eau ont également une influence dur le déplacement des masses d'air. Cette influence est fonction de la dimension du plan d'eau. Les attributs sont *Num_PE*, *Nom_PE*, *Surf_PE* et *Périm_PE*.

5.2.1. Base de données géographique

La base de données géographique correspondante a été implémentée manuellement dans ArcGIS 9™. Quelques modifications ont été réalisées afin de pouvoir bénéficier directement de la structure des couches à disposition. Par exemple **Reseau_routier** s'est scindé en **Reseau_routier_princ** et **Reseau_routier_sec**, tous deux ayant les mêmes attributs. Il en est de même pour **Delimitation_territoriale**.

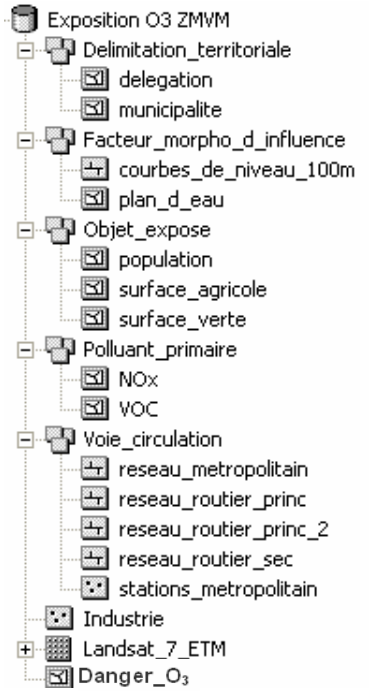


Figure 5.2.2 : base de données géographiques

Cette base de données est utilisée pour illuster quelques uns des aspects du modèle descriptif construit. Une image satellitaire Landsat 7 ETM est dans ce cas un outil à disposition pour l'analyse visuelle.

5.2.2. Représentations cartographiques

Les quelques représentations suivantes illlustrent certains des aspects du modèle conceptuel, notamment l'exposition et les émissions.

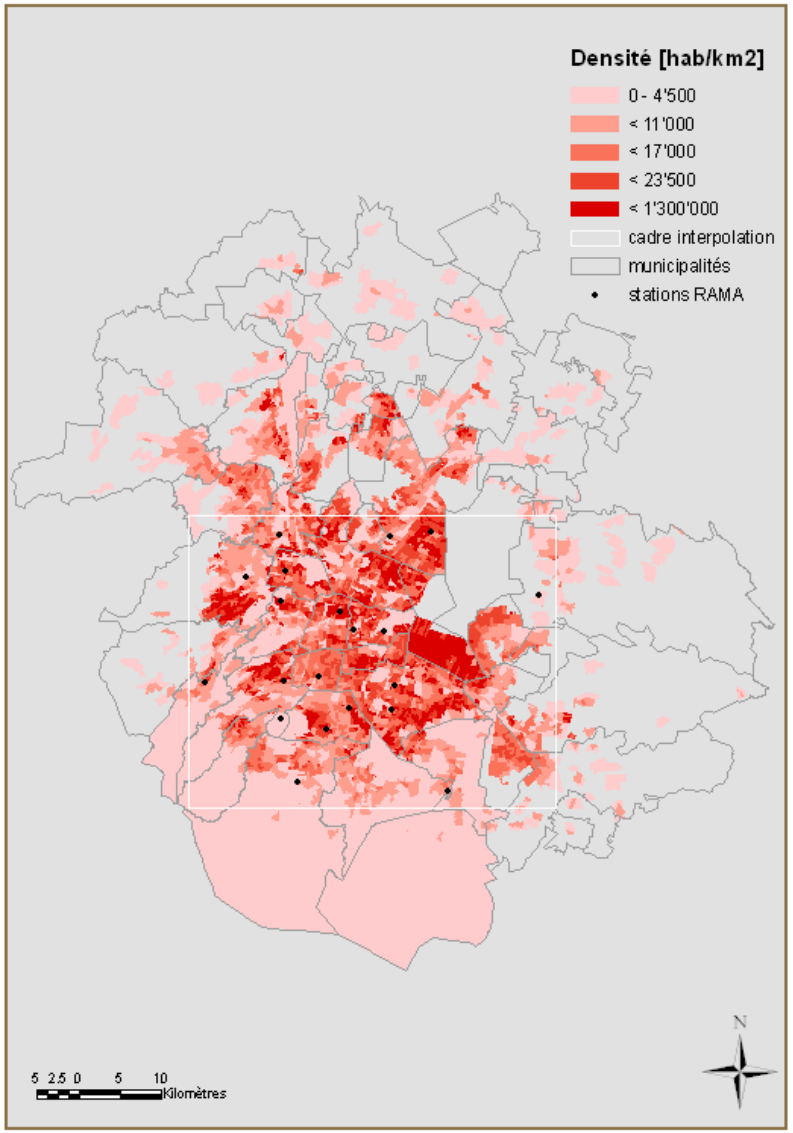


Figure 5.2.3 : densité de population par AGEB pour la ZMVM

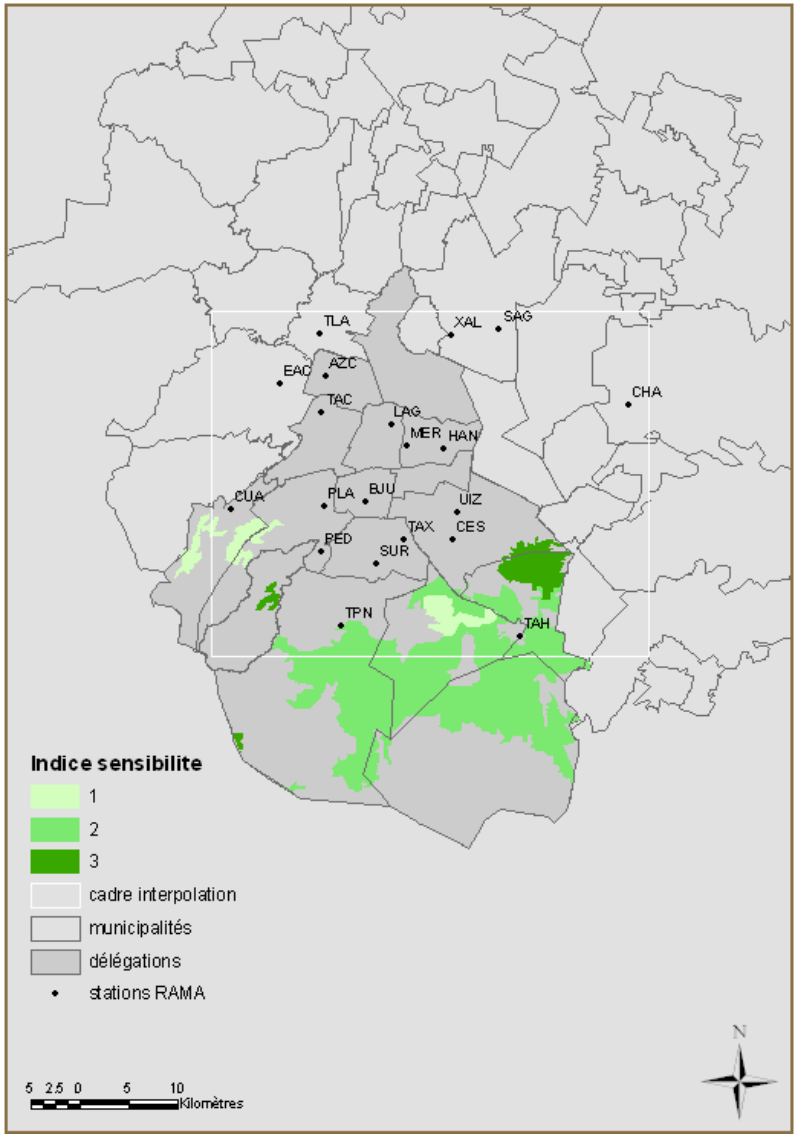


Figure 5.2.4 : indice de sensibilité des cultures à l'ozone des surfaces agricoles de la ZMVM

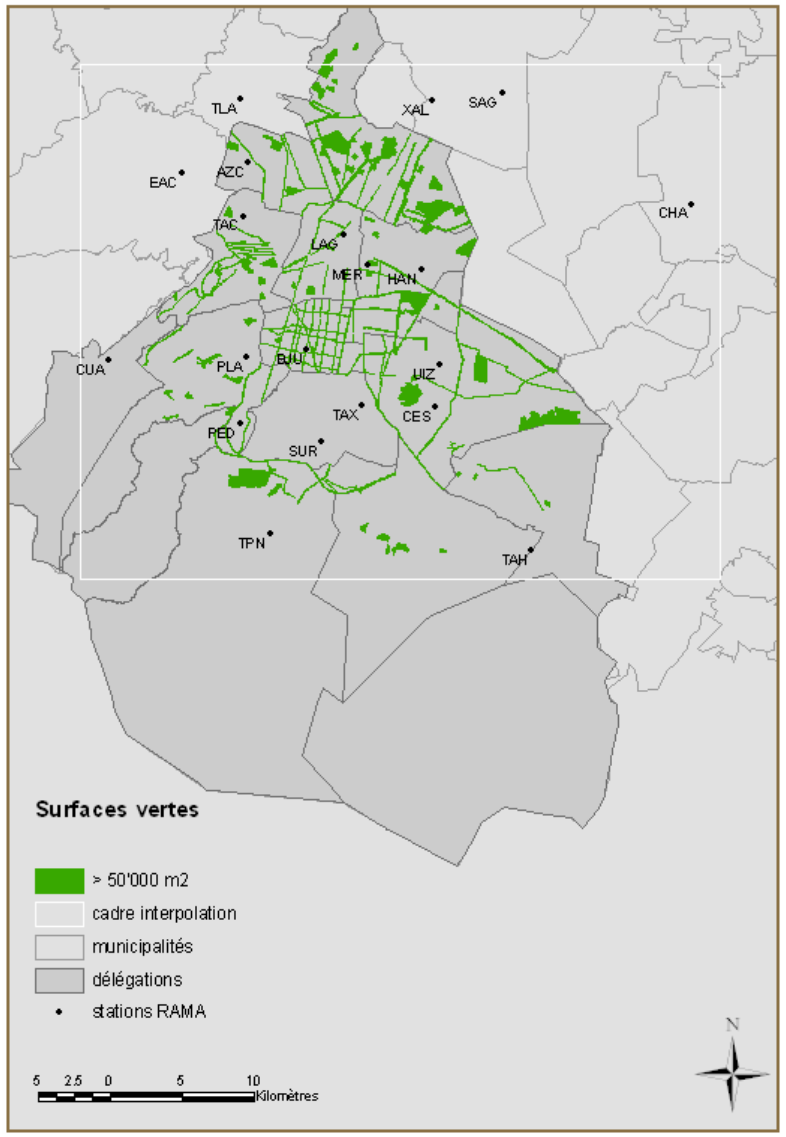


Figure 5.2.5 : surfaces vertes urbaines de la ZMVM

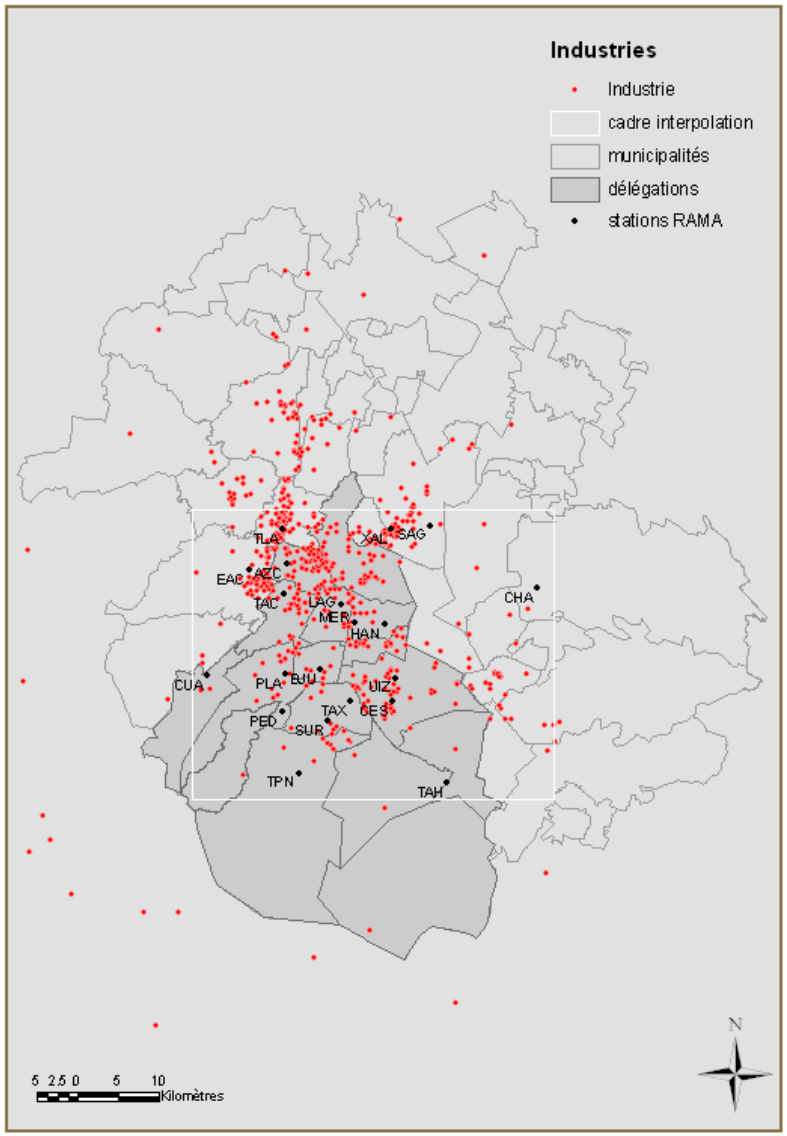


Figure 5.2.6 : principales sources d'émissions industrielles (grandes et moyennes)

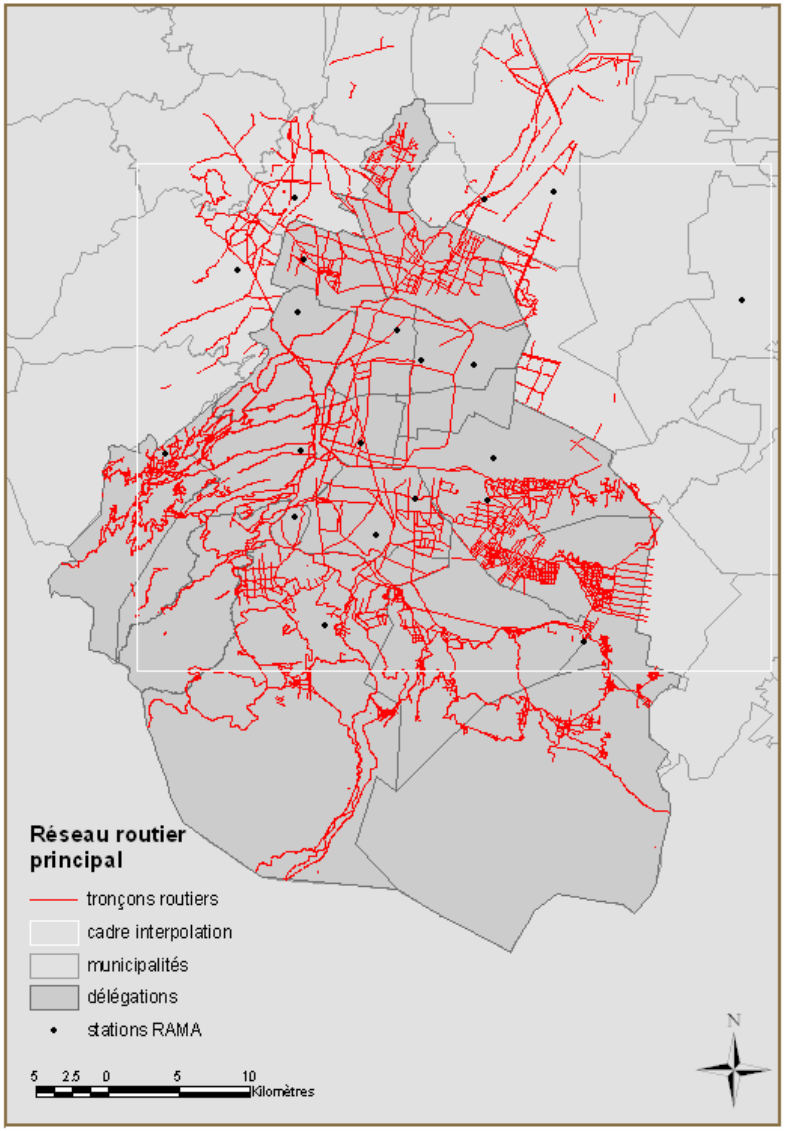


Figure 5.2.7 : réseau routier principal de la ZMVM

6. Synthèse des principaux résultats et conclusions

L'intérêt des principaux résultats de ce travail peut être considéré à deux niveaux, le premier étant relatif au caractère exploratoire de l'étude, le second étant de nature plus qualitative, à savoir comment intégrer de façon structurée un « *danger d'ozone* » à un contexte environnemental comme celui de la ZMVM.

D'un point de vue exploratoire, outre l'obtention d'épisodes à priori temporellement cohérents et comparables, les différents traitements effectués ont permis de l'observation des comportements suivants:

- les valeurs de minimum et de maximum se retrouvent systématiquement aux mêmes endroits : Chapingo et San Agustin pour les valeurs de minimum et Pedregal et Plateros pour le valeurs de pics.

- La variabilité interannuelle entre les mesures et les situations de pollution représentées est importante.

- Il existe ici une relation évidente entre l'amplitude des valeurs mesurées et la variabilité des mesures au même point

- Des stations plus particulièrement impliquées dans l'explication du phénomène ont été identifiées à la suite d'une analyse en composantes principales

- Certains résultats redondants mènent à penser que l'ensemble des traitements est cohérent.

Notons encore que l'identification d'épisodes de pollution est un résultat important puisqu'il permet un suivi graphique dans le temps de l'évolution des conditions atmosphériques. En effet, l'« *évolution historique des concentrations* » a jusqu'alors été l'outil de gestion dont le gouvernement disposait. Il s'agissait plus précisément du calcul du pourcentage de jours durant lesquels la norme est dépassée.

D'un point de vue plus spécifique de l'exploration spatiale des données, la validité des modèles semble vérifiée. L'épisode 1998 est cependant un peu marginal, et le faible nombre de paires de mesures fait partie des points faibles de cette analyse variographique.

En ce qui concerne la comparaison des deux types de modélisation (*transport-chimie et géostatistique*), certains des comportements représentés apparaissent similaires. On dénote également pour les deux types de modèles l'importance jouée par le relief, par son action indirecte sur le mouvement des masses d'air. Cependant l'interprétation des résultats a mis en évidence la difficulté qui existe à intégrer la composante temporelle au problème de la spatialisation de l'ozone par les méthodes géostatistiques. L'existence d'une éventuelle dynamique temporelle sous-jascente aux épisodes identifiés aurait comme effet de remettre en question l'identification de ces épisodes.

Le second niveau de considération des résultats est leur intégration dans un contexte environnemental qui est celui de la ZMVM. Le modèle conceptuel imaginé fait preuve au premier abord d'un certain manque de réalisme. L'intérêt du modèle réside cependant moins dans son contenu en termes d'attributs et de précision des relations que dans la structure du modèle. En effet, celui-ci présente de manière structurée les principaux objets dont il est faut tenir compte pour le suivi du danger d'ozone. Il pourrait en effet être possible de procéder au même type de démarche pour la gestion du danger lié aux émissions de particules par exemple.

7. Remerciements

Au terme de cette étude je remercie MM. Régis Caloz et Alain Clappier pour leur attention et le suivi de mon travail.

Je remercie également MM. Emmanuel Gaillard (*Ecole des Mines de St-Etienne*), Clive Müller (*Laboratoire de Pollution Atmosphérique et des Sols*) et les collaborateurs du Laboratoires des Systèmes d'Information Géographique, dont l'assistance est irréprochable.

Je désire aussi remercier tout particulièrement Mme. Beatriz Cardenas, M. Salvador Blanco et Mlle Magdalena Sanchez pour leur accueil et leur aide à Mexico.

8. Références bibliographiques

CALOZ R., "*Analyse spatiale*", cours en Section des Sciences et Ingénierie de l'Environnement, EPFL, Lausanne, 2003

DE ICAZA DEL RIO G., "*Formation and Transformation Mechanisms of Particulate Matter Under Ten Micrometers (PM10) and Ozone (O3) in the Mexico City Metropolitan Area and the Greater Manchester Area"*, These, Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT), Instituto Nacional de Ecología (INE), Mexico DF, 1999

JUNIER M., KIRCHNER F., CLAPPIER A. et VAN den BERGH H., "*The chemical mechanism generation program CHEMATA, part II: Comparison of four chemical mechanism in a three-dimensional mesoscale simulation*", Atmos. Environ. 39, 1161-1171, 2004

GARCIA-COLIN SCHERER L. & RUBEN VARELA HAM J., *"Contaminacíon Atmosférica IV"*, El Colegio Nacional, Mexico DF, 2003

KANEVSKI M. & MAIGNAN M., "*Analysis and Modelling of Spatial Environmental Data",* Presses Polytechniques Universitaires Romandes, Lausanne, 2004

LEBART L. et al., " *Statistique exploratoire multidimensionnelle "*, Dunod, Paris, 2000

Martilli, A., A. Clappier, and, M. W. Rotach, *"An urban surfaces exchange parameterisation for mesoscale models"*, Boundary Layer Meteorol., 104, 261-304, 2002

MOLINA L. & M. et al, *"Air Quality in the Mexico Megacity : An integrated Assessement "*, Alliance For Global Sustainability Bookseries, Vol.2, Kluwer Academic Publishers, The Netherlands, 2002

MORGENTHALER S., " *Introduction à la statistique "*, Presses Polytechniques Universitaires Romandes, Lausanne, 2001

MOLINA M.J., MOLINA L.T., SOSA G., GASCA J. et WEST J., " *Analysis et Diagnostico del Inventorio de Emisiones de la Zona Metropolitana del Valle de Mexico "*, MIT Integrated Program on Urban, Regional and Global Air Pollution Report No.5, Cambridge MA, 2000

PANNATIER Y., « *VARIOWIN, Software for Spatial Data Analysis in 2D* », Springer, s.l., 1996

ROULET Y-A., "Validation and Application of an Urban Turbulence Parametrisation Scheme for Mesoscale Atmospheric Models", Thèse, Environnement Naturel, Architectural et Construit, EPFL, Lausanne, 2004

WACKERNAGEL H., "*Multivariate Geostatistics"*, Springer, s.l., 1998

WARD P., " *Mexico City "*, John Wiley and Sons, New York, 1998

s.a., " *Inventario de Emisiones a la Atmósfera, Zona Metropolitana de la Valle de México "*, Secretaría del Medio Ambiente (SMA), Mexico DF, 2000

s.a., " *Primer Informe Sobre la Calidad del Aire en Ciudades Mexicanas 1996 "*, Centro Nacional de Investigación y Capacitación Ambiental - Instituto Nacional de Ecología, Mexico DF, 1996

s.a., " *Tercer Informe Sobre la Calidad del Aire en Ciudades Mexicanas 1998 "*, Centro Nacional de Investigación y Capacitación Ambiental - Instituto Nacional de Ecología, Mexico DF, 1998

Autres liens :

http://www.ine.gob.mx
http://sinaica.ine.gob.mx
http://www.sma.df.gob.mx/simat/
http://www.inegi.gob.mx
http://sc.inegi.gob.mx/simbad/
http://www.df.gob.mx
http://www.semarnat.gob.mx
http://148.243.232.103/imecaweb/
http://www.centrogeo.org.mx/Index.htm
http://www.salud.gob.mx/
http://www.umwelt-schweiz.ch/buwal/fr/medien/umwelt/1999_2/unterseite5/, " *Le smog estival n'a pas fini de nous pomper l'air* "
http://www.ine.gob.mx, " *Normas Oficiales Mexicanas en Matiera de Medio Ambiente y Ecologia "*
http://sirs.scg.ulaval.ca/perceptory/

Index des figures

Index des tableaux

Annexes