



**HAL**  
open science

# Projet SIGFRIED 1: SIG Facteurs de Risques Environnementaux et Décès par cancer Integration de bases de données environnementales dans un SIG pour servir à l'analyse des disparités géographiques de cancer

Julien Caudeville, Jean-Baptiste Masson

## ► To cite this version:

Julien Caudeville, Jean-Baptiste Masson. Projet SIGFRIED 1: SIG Facteurs de Risques Environnementaux et Décès par cancer Integration de bases de données environnementales dans un SIG pour servir à l'analyse des disparités géographiques de cancer. 5. Edition des Journées Interdisciplinaires de la Qualité de l'Air (JIQA), Feb 2008, Lille, France. pp.NC. ineris-00973282

**HAL Id: ineris-00973282**

**<https://ineris.hal.science/ineris-00973282>**

Submitted on 4 Apr 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Projet SIGFRIED 1 : SIG Facteurs de Risques Environnementaux et Décès par cancer. Intégration de bases de données environnementales dans un SIG pour servir à l'analyse des disparités géographiques de cancer.

**J. Caudeville, JB. Masson**

INERIS : Institut National de l'Environnement  
industriel et des RISques - DRC - ERSA  
Parc Technologique Alata - BP 2  
60550 Verneuil en Halatte

Université de Technologie de Compiègne  
Unité Mixte Recherche C.N.R.S. 6599  
Centre de Recherches de Royallieu  
60200 COMPIEGNE

Mots-clés : SIG, pollution extérieure, carte d'exposition, bases de données, géostatistique, modélisation, évaluation de risque, cancer.

## Résumé

Cet article vise à présenter le contexte et la méthodologie de la thèse SIGFRIED1 (SIG Facteurs de Risques Environnementaux et Décès par cancer). Cette étude a pour objectif de construire des cartes d'exposition construites à partir des bases de données de pollution par milieu pour servir à l'analyse des disparités géographiques de cancer en France par l'utilisation des Systèmes d'Informations Géographiques. Il développe les quatre étapes méthodologiques retenues pour la construction des cartes d'exposition : l'acquisition des données, leur traitement, la scénarisation des expositions et l'évaluation du risque.

## 1. Introduction

### 1.1. Disparités spatiales et temporelles de l'incidence des cancers

Les études récentes sur l'incidence du cancer font apparaître d'une part une évolution temporelle, et d'autre part une hétérogénéité spatiale.

En 2002, le CIRC (Centre International de Recherche sur le Cancer) estimait au niveau mondial l'incidence du cancer à 11 millions de cas et le nombre de décès à 7 millions. Les données les plus récentes pour la France [1] permettent d'estimer qu'un homme sur deux et qu'une femme sur trois seront touchés par cette maladie au cours de leur vie. Un premier facteur explicatif de l'augmentation de l'incidence pourrait être le vieillissement de la population mais, après standardisation de ces données par rapport à l'âge et la structure des populations, ce seul facteur apparaît insuffisant [2]. Selon cette étude, pour une progression de l'incidence des cancers de 63% entre 1978 et 2000, 35% sont liés à une progression intrinsèque et 28% au vieillissement. Les comportements individuels (tabagisme, consommation d'alcool, alimentation) mais aussi l'exposition à des substances cancérigènes, notamment dans un cadre professionnel, sont les principaux facteurs d'environnement suspectés, sans que la part de chaque facteur de risque ne soit précisément quantifiée [3].

## 1.2. Les expositions environnementales comme facteur de risque

Trois types de facteurs de risques sont aujourd'hui attribués au développement du cancer : les facteurs génétiques, les facteurs comportementaux et les facteurs environnementaux.

La prédisposition génétique a été longtemps considérée comme le facteur principal. L'augmentation annuelle de 1% des cas de cancer chez l'enfant [4], les études chez les migrants [5] et celles chez les jumeaux [6] montrent cependant que les facteurs génétiques n'ont qu'un rôle mineur. Le second type de facteurs correspond au comportement individuel (tabagisme, alcoolisme, exercice physique, nutrition...). Mais les cancers progressant le plus (cancer du sein, cancer de la prostate) sont peu liés aux facteurs de risque classiques comme tabac et alcool, et sont beaucoup plus soupçonnés d'être associés aux expositions environnementales, notamment aux perturbateurs endocriniens [3]. Selon les études, la fraction attribuable aux facteurs environnementaux peut varier entre 0,07 [7] et 80% [2], ce qui fait de ce troisième type de facteurs le plus controversé. En France, un réel débat s'est engagé autour de cette question.

Cette variabilité s'explique d'une part par la perception sociale de la définition de la notion d'environnement et d'autre part par la diversité des méthodologies utilisées.

La notion d'environnement est par essence très complexe, car elle est la résultante de facteurs de risque liés à l'environnement extérieur, à l'environnement intérieur (notamment professionnel), au mode de vie, à la nutrition et aux expositions médicales. L'exposition est à évaluer non seulement aujourd'hui, mais au cours des décennies passées, puisque les pathologies cancéreuses ont un temps de latence de 10 à 20 ans en moyenne.

C'est la reconnaissance de cette complexité qui a conduit à considérer que l'étude de l'exposition des populations pouvait être considérée comme une nouvelle discipline scientifique [8], concept traduit en français par expologie [9].

Dans ce contexte, l'angle d'attaque adopté ici est de développer une approche prenant en compte les disparités spatiales pour évaluer l'influence des facteurs environnementaux sur le risque de cancer.

## 1.3 Utilisation des Systèmes d'Information Géographique en expologie et en épidémiologie environnementale.

Un système d'information géographique (SIG) est un outil informatique permettant d'organiser des données numériques spatialement référencées et de les représenter sur des cartes. Il constitue un moyen efficace de conduire des analyses spatiales et d'en exploiter les résultats. Des études d'analyse des disparités géographiques sont menées depuis quelques années aux Etats-Unis. Les grandes agences fédérales américaines se sont dotées de programme SIG dans les procédures d'évaluation de risque (évaluation de l'exposition, Excès de Risque Unitaire,...) (Figure 1) [10]. Un travail de ce type a été conduit par l'US EPA (United States Environmental Protection Agency) sur l'ensemble du territoire des Etats Unis (Figure 2) [11,12]. Cette agence produit et diffuse des cartes d'exposition et de risque sanitaire. Dans d'autres études, les disparités spatiales de cartes environnementales sont étudiées et confrontées à des cartes sanitaires pour mettre en évidence des corrélations et des associations entre cancer et pollution [13,15].

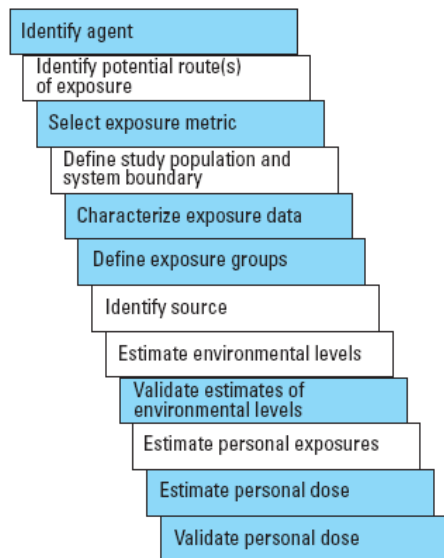


Figure 1 : Intégration des outils des SIGs dans les procédures d'évaluations de risque [10].

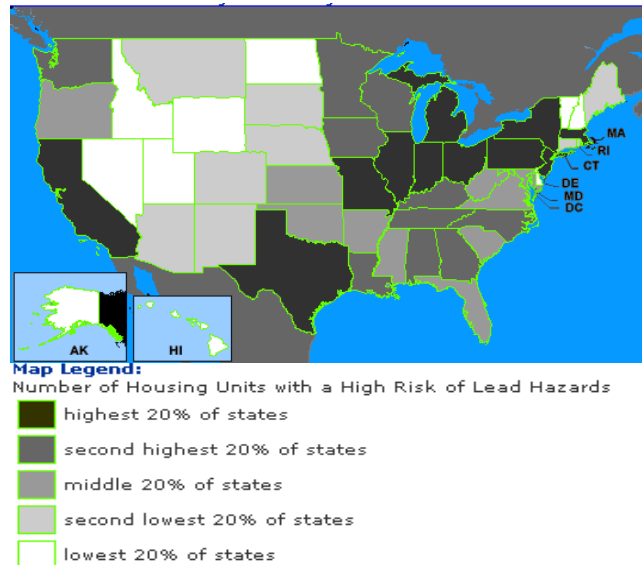


Figure 2 : Carte de risque des Etats-Unis par état diffusée par l'US EPA.

#### 1.4 Le projet CIRCE

Les travaux que nous exposons ici s'intègrent dans le projet CIRCE (Cancer, Inégalités Régionales, Cantonales et Environnement), dont l'objectif est de quantifier l'exposition humaine aux polluants environnementaux cancérigènes, de saisir ses hétérogénéités spatiales et d'évaluer son influence sur le risque de cancer. Ce projet, piloté par l'INERIS et l'unité de recherche 593 de l'INSERM, regroupe de nombreux partenaires scientifiques :

- les Observatoires Régionaux de Santé (ORS) de quatre régions (Picardie, Rhône-Alpes, Ile-de-France, Nord Pas-de-Calais) ;
- les centres de Géostatistiques de l'Université de Technologie de Compiègne et de l'Ecole des Mines de Paris ;
- des partenaires internationaux comme BioMedware et le Silent Spring Institute.

La démarche générale du projet CIRCE s'inspire des travaux réalisés en Amérique du Nord sur des sujets semblables: citons notamment les études de l'US-EPA [11,12], à Cape Cod [13], ou encore à Long Island [14,15].

Dans le cadre de la première phase du projet CIRCE, les inégalités géographiques de mortalité par cancer ont été répertoriées par les quatre ORS associés au projet dans leurs régions respectives : ceux-ci ont réalisé une analyse descriptive, sous forme de cartographies, des variations spatiales de la mortalité par cancers au niveau cantonal. Les comparaisons avec les situations en France et en Europe permettent de dégager les spécificités régionales. Des analyses plus fines, à un niveau cantonal, révèlent par ailleurs des contrastes importants au sein des régions (Figure 2).

L'objectif de SIGFRIED est de construire des bases de données spatialisées, renseignant l'exposition des populations à divers polluants dont le pouvoir cancérigène est prouvé ou suspecté. Ce travail est attribué aux deux projets SIGFRIED, le premier se focalisant sur les polluants des milieux extérieurs (air, eau, sol) et le second sur l'air intérieur.

La seconde phase du projet CIRCE portera sur le croisement de la cartographie sanitaire (réalisée par les ORS) et des cartographies d'expositions environnementales (construites au cours des thèses SIGFRIED).

### 1.5 Les thèses SIGFRIED (Systèmes d'Informations Géographiques, Facteurs de Risques Environnementaux et Décès par cancer)

- SIGFRIED 1 : Construction de bases de données environnementales, pollution extérieure.

Le projet doit permettre de disposer de cartes d'exposition construites à partir des bases de données spatialisées de pollution par milieu (eau, air et sol), d'évaluer les risques cancérigènes liés à certains polluants, de construire des cartes d'exposition et d'identifier les facteurs environnementaux potentiels de risque sur l'incidence des cancers en France par une nouvelle approche méthodologique.

- SIGFRIED 2 : Construction d'indicateurs de la qualité de l'air intérieur et intégration dans un SIG.

Pour compléter l'évaluation de l'exposition des populations, l'exposition spécifique dans les milieux intérieurs (formaldéhyde, benzène, etc.) sera étudiée de manière indépendante. En l'absence de données spatialisées à ce jour en France sur la Qualité de l'Air Intérieur dans l'habitat, les résultats de la campagne nationale Logements de l'OQAI (Observatoire de la Qualité de l'Air Intérieur) seront exploités pour établir, dans un premier temps, les facteurs déterminants (bâti, équipement, etc.) des niveaux de concentration des polluants étudiés et de construire des indicateurs de la qualité de l'air intérieur. Dans un second temps, l'objectif du projet, présenté sous forme de poster au cours des JIQA (Journées Interdisciplinaires de la Qualité de l'Air) sera d'intégrer ces indicateurs dans un SIG, de manière analogue à SIGFRIED 1.

## 2 Méthodologie

La construction de cartes de risque dans cette étude nécessite quatre étapes préparatoires : l'acquisition des données, leur traitement, la scénarisation des expositions et l'évaluation du risque (Figure 3).

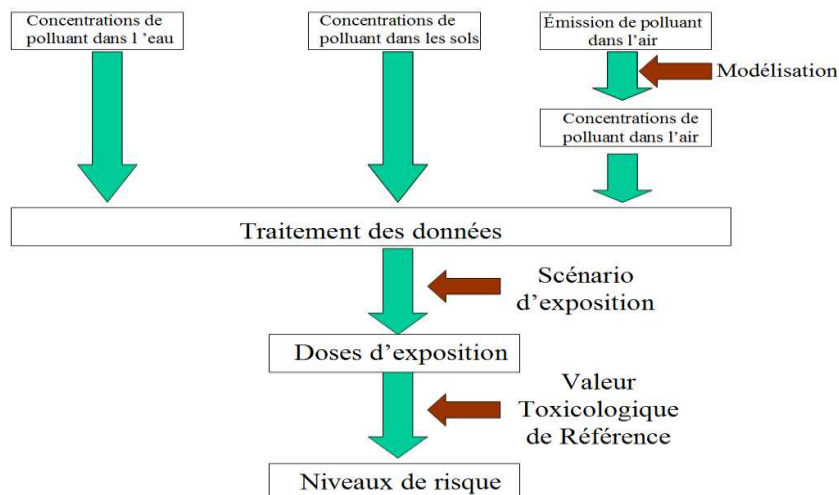


Figure 3 : Description de la méthodologie adoptée pour l'étude.

## 2.1 Inventaire des bases de données environnementales disponibles

Pour permettre une lisibilité sur le type de sources d'information spatialisé disponible et intégrable dans un système d'information géographique, un inventaire des bases de données environnementales sur le territoire français a été effectué.

Cet inventaire recense les bases de données environnementales pérennes, renseigne sur les informations de modalités de production, de gestion, d'organisation et d'accessibilité des données. Il répertorie 27 bases de données environnementales dont 20 sont spatialisées. Sept d'entre elles sont a priori retenues pour les différents médias (eau, air, sol) dans cette étude (Tableau 1).

	NOM	Gestionnaire	Sigle	Champ concerné	Informations recueillis dans la base	Niveau de détail géographique	Année de début
<b>SOLS</b>	Base d'Anciens Sites Industriels et Activités de Service	BRGM	BASIAS	Sites et sols relevant des installations classées potentiellement pollués	Historique des activités (produits utilisés ou générés) ; Utilisation et projets de réaménagement ; Accidents ou pollutions connues ; Critères environnementaux	France métropolitaine et Départements d'Outre Mer	1994
	Base de données sur les sites et sols pollués ou potentiellement pollués	MEDAD	BASOL	Sites et sols pollués ou potentiellement pollués par des activités industrielles	Identification et localisation du site, caractérisation du site, description du site, situation technique du site et caractérisation de surveillance du site et restriction d'usage, traitements effectués	Point de mesure	1992
<b>EAUX</b>	Système d'information en Santé-Environnement sur les Eaux	DGS	SISE-Eaux	Qualité de l'eau : eau d'alimentation, eau de baignade, eau conditionnée, eau minérale naturelle	Descriptif du parc d'installations de production-traitement-distribution des eaux ; Descriptif du réseau de points de surveillance ; Résultats d'analyses de contrôle sanitaire commentés ; Données de la surveillance	Point de mesure	1994
	Banque d'Accès aux Données sur les Eaux Souterraines	BRGM	ADES	Eaux souterraines	Données quantitatives (niveau des nappes) et qualitatives (concentration de nombreux paramètres) relatives aux eaux souterraines	Point de mesure	Variable
<b>AIR</b>	Registre Français des Emissions Polluantes	MEDAD	iREP	Émissions polluantes et déchets des installations classées soumises à autorisation (établissements industriels et élevages)	Résultats de la déclaration annuelle des émissions polluantes	Etablissement	2003
	Base de données de la qualité de l'air	Ademe	BDQA	Qualité de l'air : pollution atmosphérique	Concentration dans l'air de différents polluants	Point de mesure	1985 Variable selon polluant
	Base de données Climatologiques	Météo-France	Climathèque	Climat	Paramètre à renseigner dans les modèles de dispersion atmosphériques : Températures, pluviométrie, humidité, vent, insolation, rayonnement, pression....	Séries chronologiques et statistiques ponctuelles aux stations météorologiques	En fonction des paramètres et de la station de mesure.

Tableau 1 : Bases de données environnementales spatialisées utilisées pour l'étude.

Dans le courant de l'année 2008, un Inventaire National Spatialisé des émissions atmosphériques polluantes en France (INS) sera disponible et éventuellement intégré dans l'étude. L'objectif de l'INS est le recensement exhaustif des émissions atmosphériques d'une cinquantaine de composés avec une résolution spatio-temporelle fine (le km<sup>2</sup> et l'heure) pour tous les secteurs d'activité anthropiques et biotiques.

Un réseau de sites, dit « Réseau de Mesures de la Qualité des Sols » (RMQS) a été très récemment mis en place sur le territoire français pour l'observation de l'évolution de la qualité des sols au niveau national. Cette base de données constitue un bilan sur l'état des sols français sur une maille de 16 km<sup>2</sup> et renseigne sur les teneurs de 9 éléments traces métalliques et 7 éléments majeurs. Certains polluants organiques persistants devraient compléter cette base de données avec l'avancement du projet. Dès sa mise à disposition, cette base de données pourra être elle aussi incorporée à l'étude.

## 2.2 Acquisition et traitement des données

L'acquisition des bases de données provenant de différentes sources implique une certaine hétérogénéité spatio-temporelle des formats de ces données. Un travail de traitement préalable est nécessaire pour homogénéiser la présentation des résultats de manière à obtenir, au final, une feuille de calcul Excel attribuant les coordonnées géographiques aux valeurs de concentration de polluant. Vu la quantité de données numériques traitées, des procédures d'automatisation de gestion de ces banques seront utilisées notamment par la programmation de macros.

Les valeurs dans le sol et dans l'eau issues du traitement de données sont présentées sous forme de concentrations. Seules les données de pollution atmosphérique se présentent sous forme de flux de polluant. Une étape de modélisation supplémentaire est nécessaire pour convertir les données d'émission en données de concentration.

Le modèle CHIMERE est utilisé pour convertir les valeurs d'émissions dans l'air en concentration. CHIMERE est un modèle déterministe de transport/transformation atmosphérique : l'évolution au cours du temps des concentrations de polluants est calculée en reliant la variation temporelle des concentrations de polluants sur le domaine à des processus physico-chimiques, qui augmentent (processus de production) ou diminuent (processus de perte) la concentration d'une espèce chimique dans l'atmosphère. A partir de données météorologiques et d'émissions, le modèle est capable de simuler des concentrations sur l'ensemble du territoire de l'Europe de l'ouest sur des mailles très fines. Ces concentrations sont directement importables dans un SIG.

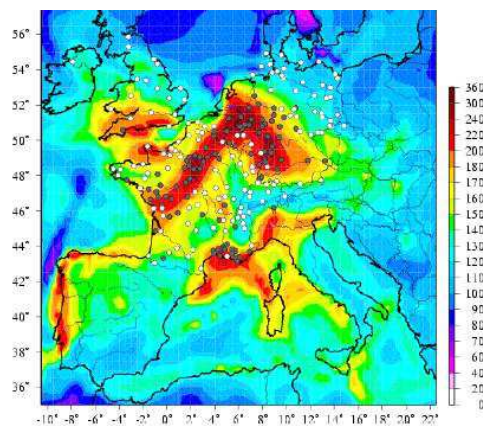


Figure 4 : Simulation du modèle CHIMERE des concentrations d'ozone durant les épisodes de canicules de 2003.

## 2.3 Détermination de scénarios d'expositions

L'une des étapes fondamentales des évaluations des risques est de décrire les situations pouvant conduire à une exposition des personnes vivant dans une zone donnée. Les voies d'exposition dépendent des modes de transfert des polluants.

Des scénarios d'exposition seront construits de manière à caractériser des groupes de référent (âge, sexe, densité de population, localisation) et de quantifier leur exposition aux divers polluants. Les contributions de chaque voie d'exposition sont la synthèse de sous-contributions :

- inhalation : air extérieur, air intérieur ;
- ingestion : consommation alimentaire, absorption d'eau de consommation.

La contamination par voie cutanée est considérée ici comme négligeable par rapport aux autres voies d'expositions.

Pour chacun des scénarios, les paramètres suivants seront définis :

- les densités de population et leur structure (âge, sexe) ;
- l'extension et la résolution spatiale de l'étude ;
- le polluant, ou la famille de polluants, suspecté(e) d'avoir un impact sanitaire (recherche des substances suspectées d'avoir un impact biologique sur le développement du cancer) ;
- l'impact sanitaire attendu ;
- l'ensemble des voies d'expositions ;
- la durée d'exposition.

Pour un scénario donné, l'étude devra évaluer la quantité de substance (ou famille chimique) absorbée par l'organisme à travers les différentes voies d'expositions. Un modèle probabiliste a été développé par l'INERIS permettant le calcul des doses d'exposition liées à une émission atmosphérique de polluant chimique. Il est basé sur des équations de transfert telles que celles présentées dans divers documents de l'US EPA. Après développement, il permettra d'évaluer aussi les risques liés aux pollutions des eaux de consommation et des sols contaminés.

## 2.4 Evaluation du risque

Les données d'expositions sont traduites en terme de risque par l'utilisation des Excès de Risque Unitaire (ERU). Les ERU traduisent la relation quantitative qui existe entre la dose d'exposition et la probabilité d'apparition d'un cancer liée à une exposition chronique. Ces ERU sont extraits de bases de données internationales de référence, transparentes du point de vue de l'élaboration des valeurs proposées et librement accessibles sur Internet. En cas d'existence de plusieurs ERU, le choix de l'un d'entre eux est effectué conformément aux pratiques INERIS publiées en mai 2006. L'évaluation de risque pour les substances cancérigènes utilise l'Excès de Risque Individuel (ERI) pour estimer la probabilité supplémentaire, par rapport à un sujet non exposé, qu'un individu développe au cours de sa vie entière l'effet associé à une exposition à un agent dangereux. Il est calculé selon la formule :

$$ERI = ERU \times \frac{\sum_j E_j \times T_j}{70}$$

avec ERI : Excès de risque individuel (-),

ERU : Excès de Risque Unitaire (mg/kg.j)<sup>-1</sup> ou (mg/m<sup>3</sup>)<sup>-1</sup>,

E<sub>j</sub> : Dose d'exposition calculée pour la classe d'âge j (mg/kg.j) ou (mg/m<sup>3</sup>),

T<sub>j</sub> : Durée d'exposition associée à la classe d'âge j (années).

## 2.5 Interpolation spatiale et traitement statistique

Les valeurs d'expositions spatialisées seront interpolées par une méthode géostatistique de krigeage pour obtenir une estimation de ces valeurs sur un maillage homogène du territoire étudié. Par ailleurs, un travail d'évaluation de



l'incertitude est prévu de manière à définir la significativité des valeurs calculées. Une analyse spatiale des cartes permettra d'évaluer la variabilité géographique des expositions et de les comparer avec celle des cartes de cancer.

### 3. Conclusion et perspective

Le travail présenté ici permettra de quantifier et de spatialiser l'exposition et le risque aux polluants environnementaux cancérigènes par l'utilisation des Systèmes d'Informations Géographiques. Dans une étape ultérieure, les cartes de risques estimées seront comparées aux cartes de risques mesurées comme cela a été fait au cours d'études similaires [13,14].

## Références

- [1] Remontet L., Buemi A., Velten M., Jouglu E., Estève J. Cancer incidence and mortality in France over the period 1978-2000. *Santé Publique*. 2003 ; 51 (1) : 3-30.
- [2] Sasco AJ. Cancer and globalization. *Biomed Pharmacother*. 2007 ; doi/10.1016/j.biopha.2007.10.015.
- [3] Sasco AJ., Kaaks R., Little RE. Breast cancer : occurrence, risk factors and hormone metabolism. *Expert Rev Anticancer Ther*. 2003 ; 3 (4) : 546-62.
- [4] Steliarova-Foucher E., Stiller C., Kaatsch P. et al. Geographical patterns and time trends of cancer incidence and survival among children and adolescents in Europe since the 1970s (the ACCIS project) : an epidemiological study. *The Lancet*. 1997 ; 364 (9451) : 2097-2105.
- [5] Sasco AJ. Migrations et cancers. *Rev Med Interne*. 1989 ; 10 : 341-8.
- [6] Lichtenstein P., Holm NV., Verkasalo PK. et al. Environmental and heritable factors in the causation of cancer : analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000 ; 343 (2) : 78-85.
- [7] International Agency for Research on Cancer. Attributable causes of cancer in France in the year 2000. In: IARC working group reports, vol. 3. Geneva: *WHO Press* ; 2007.
- [8] Ott WR. Human exposure assessment: the birth of a new science. *Expo Anal Environ Epidemiol*. 1995 ; 5 (4) : 449-72.
- [9] Cicolella A. L'expologie. Note INERIS DRC ERSA-02-25419-276.
- [10] Nuckols JR., Ward M. Using Geographic Information Systems for Exposure Assessment in Environmental Epidemiology Studies. *Environ Health Perspect*. 2004 ; 112 : 1007-1015.
- [11] Woodruff T. J., Caldwell J., Morello-Frosch R., Rosenbaum A. Public Health Implications of 1990 Air Toxics Concentrations across the United States. *Environmental Health Perspectives*. 1998 ; 106 (5) : 245-251.
- [12] Woodruff T. J., Caldwell J., Cogliano VJ., Axelradet DA. Estimating Cancer Risk from Outdoor Concentrations of Hazardous Air Pollutants in 1990. *Environmental Research*. 2000 ; 82 (3) : 194-206.
- [13] Brody JG., McKelvey W., Swartz CH., Kennedy T., Rudel RA. Breast cancer risk and drinking water contaminated by wastewater: a case control study. *Environ Health*. 2006 ; 5 : 28.
- [14] Jacquez G., Greiling D. Geographic boundaries in breast, lung and colorectal cancers in relation to exposure to air toxics in Long Island, New York. *International Journal of Health Geographics*. 2003 ; 2 (1) : 4.
- [15] Jacquez G., Greiling D. Local clustering in breast, lung and colorectal cancer in Long Island, New York. *International Journal of Health Geographics*. 2003 ; 2 (1) : 3.