

Noise Filtering Cancer Mortality Data for a Better Assessment of Health-Environment Relationships: Application to the Picardy Region

Mahdi-Salim Saib^{1,2*}, Julien Caudeville¹, Florence Carre¹, Olivier Ganry³, Alain Trugeon⁴ and Andre Cicolella¹

¹French National Institute for Industrial Environment and Risks; Parc Technologique Alata, BP 2, 60550 Verneuil-en-Halatte, France

²University of Picardie Jules Verne; 33 rue St Leu, Amiens, 80039, France

³University Hospital of Amiens; Place Victor Pauchet Amiens, 80054, France

⁴Regional Observatory of Health and Social Issues in Picardie (OR2S); 3, rue des Louvels, Amiens, 80036, France

Abstract

Cancer is one of the leading causes of mortality. However, it is necessary to analyze this disease from different perspectives. Cancer mortality maps are used by public health officials to identify areas of excess and to guide surveillance and control activities. However, the interpretation of these maps is difficult due to the presence of extremely unreliable rates, which typically occur for sparsely populated areas and/or less frequent cancers. The analysis of the relationships between health data and risk factors is often hindered by the fact that these variables are frequently assessed at different geographical scales. Geostatistical techniques that have enabled the process of filtering noise from the maps of cancer mortality and estimating the risk at different scales were recently developed. This paper presents the application of Poisson kriging for the examination of the spatial distribution of cancer mortality in the "Picardy region, France". The aim of this study is to incorporate the size and shape of administrative units as well as the population density into the filtering of noisy mortality rates and to estimate the corresponding risk at a fine resolution.

Keywords: Poisson kriging; Filtering; Cancer mortality

Introduction

One objective of the French National Plan for Health and the Environment (NPHE) is to prevent diseases caused by environmental factors, particularly cancer. In this context, the Cancer Inequalities Regions, Counties and Environment (CIRCE) project aims to quantify how much the socio-economic and environmental factors account for geographical inequalities, here defined mortality due to cancer.

In France, geographical health inequalities are a recent study topic. Previous studies were based on either individual-level surveys [1,2] or spatially aggregated data (administrative unit) from specific regions of France [3]. On a regional scale, data are often available at a fine level of resolution. This allows for building environmental, socioeconomic and health indicators. The following are two illustrative examples of this: (1) Rey et al. [4] built a FDep deprivation index, which was mapped using the most detailed census administrative level (French census tract IRIS); (2) Advances in computational technologies and development of widely accessible georeferenced databases are permitting the connection of information systems such as Geographic Information Systems (GIS) and risk models. Exposure indicators described in Caudeville et al. [5,6] for quantifying human exposure to chemical substances were mapped at a resolution of a 1 km² grid.

Regarding health data, because there is a protection rule for individual patient data, these data are not publicly available. Only aggregated data are available at the level for which the disclosure or reconstruction of the patient identity is impossible. These corresponding levels of these census units may be regions or counties in France. This aggregation unfortunately results in large uncertainty about rates or risks calculated for small or sparsely populated areas. This effect is known as the "small number problem" [7]. Another challenge for epidemiology is the analysis and synthesis of the relationships between spatial data collected at different spatial scales.

The geostatistical approach, in this context, presents a spatial methodology that allows for filtering the noise caused by the small number problem and enables the estimation of mortality risk and the associated uncertainty at different spatial scales.

The geostatistical analysis of disease data has received increasing attention with kriging becoming more popular. Lai performed ordinary kriging on Chinese cancer mortality data of 63 rural counties [8]. To produce a set of contour maps, the spatial structure of the cancer mortality rates was studied but other possible covariates were not incorporated. A first attempt to take into account the discrete nature of cancer data was the use of binomial cokriging which was employed to produce a map of childhood cancer risk in the West Midlands Health Authority Region (WMHAR) of England [9]. The application of this technique to Long Island (USA) data led to negative variogram estimates. To avoid this problem, binomial cokriging was extended to the case when the variance of observed rates is smaller than expected under the binomial model. One geostatistical filtering approach used is modified binomial cokriging which was applied to estimate breast cancer incidence in Long Island, New York [10]. The modified technique was shown to be more flexible and robust concerning the underlying hypothesis that all counties have the same spatial support, and the simulation studies have demonstrated its more accurate estimates [11].

Another geostatistical technique, Poisson kriging, was recently developed to filter noise from the data by accounting for spatially varying population sizes and spatial patterns. The methodology for estimating a spatial Poisson distribution was first introduced by Kaiser et al. [12]. They developed the spatial "auto-models"

***Corresponding author:** Mahdi-Salim. Saib, French National Institute for Industrial Environment and Risks; Parc Technologique Alata, BP 2, 60550 Verneuil-en-Halatte, France, Tel: +33-(0)344-618-115; Fax: +33-(0)344-556-399; E-mail: MahdiSalim.SAIB@ineris.fr

Received May 14, 2014; **Accepted** July 09, 2014; **Published** July 14, 2014

Citation: Saib MS, Caudeville J, Carre F, Ganry O, Trugeon A, et al. (2014) Noise Filtering Cancer Mortality Data for a Better Assessment of Health-Environment Relationships: Application to the Picardy Region. J Biomet Biostat 5: 200. doi:10.4172/2155-6180.1000200

Copyright: © 2014 Saib MS, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

based on the Poisson distribution to be used to incorporate spatial dependencies among the variables. However, their model is not well suited for irregularly sampled data and interpolation. Monestiez et al. [13,14] introduced Poisson kriging to model spatially heterogeneous observations. The approach applied by Monestiez is similar to binomial cokriging proposed by Oliver except that the count data are assumed to follow a Poisson distribution. Poisson kriging was then generalised to estimate prostate cancer mortality risk in the United States [15], breast and cervix cancer mortality in New England States [16] and cholera and dysentery incidence risk in Bangladesh [17] by incorporating varying population sizes in the processing of cancer data. When the risk values were spatially correlated, simulation studies showed that in most cases, Poisson kriging outperformed other smoothers such as population-weighted estimators and empirical Bayes smoothers [16]. It is not practical to represent each geographic unit by its centroid, especially when geographic units vary greatly in size and shape. The geographical characteristics need to be incorporated for data analysis together with spatially varying population. The framework for Area-to-area (ATA) or Area-to-point (ATP) kriging was first introduced by Kyriakidis for interpolating point values from available areal data [18]. Goovaerts modified the ATP estimator to a Poisson estimator and applied ATP Poisson kriging to lung and cervix cancer mortality in counties of the United States [15].

The aim of this paper is to examine the spatial distribution of cancer mortality in the Picardy region using geostatistical methods, which consists of two steps: (a) filtering of the noise in the data based on Poisson kriging (Area to Area-ATA) and (b) mapping of the corresponding risk at a fine resolution (Area to Point-ATP). The approach is illustrated using age-adjusted lip, oral cavity, pharynx and lung cancer mortality rates recorded from 2000-2009.

Materials and Methods

Study area

The region of Picardy consists of 112 counties (Figure 1), which covers an area of approximately 19,500 km² and is located between North Artois, the Ile-de-France in the south, the Bay of the Somme to the west and east Champagne. It covers the departments of Somme, Oise and Aisne. The urbanization rate in this region is far below the national average (60.4% compared to 74% for the whole country). The agricultural sector provides more than 4% of the French agricultural production. This region also has significant industrial activity. Fine and specialty chemicals account for nearly 15% of the jobs in this region and the automotive industry accounts for 40% of industrial employment (26.5% of assets employed in industry against 19.5% nationally).

Data sources

The health data came from the Regional Health Observatory of Picardy [19], where the age-adjusted mortality rates are calculated for each county from 2000 to 2009. Ten years is likely to be more representative in this case than a simple year in order to reduce temporal rate fluctuation. The average population of counties was computed annually by sex and age group for the years 2000 to 2009. These estimates were based on the census population conducted in 1999 and 2009, infant deaths recorded from 2000 to 2009 and the national mortality rates (metropolitan France). The relative proportion of the population in each cell of 1 km² was derived from the INSEE population data and was downloaded from the INSEE (National Institute of Statistics and Economic Studies) website.

Table 1 shows the cumulative, maximum and minimum number

of mortality and age-adjusted rates/per 100 000 person-years by county from 2000 to 2009.

Geostatistical approach

Spatial prediction (Area-to-area (ATA) and Area-to-point (ATP) Poisson kriging): The cancer count $d(v_\alpha)$ is interpreted as a realization of a random variable $D(v_\alpha)$ that is Poisson distributed with a parameter (expected number of counts), which is the product of the population size $n(v_\alpha)$, by the local risk $R(v_\alpha)$. The local risk $R(v_\alpha)$ can be thought of as a noise-filtered mortality rate for area v_α , which we also refer to as the mortality risk. It is estimated by using a variant of kriging with nonsystematic errors, known as Poisson kriging [13]. The aggregation of data into areal units of different shapes and sizes can cause a visual bias. A particular case of ATA kriging is when the prediction support is so small that it can be assimilated to a single point, in which case ATP kriging [15,18] is used to create high-resolution maps of the estimated mortality risk to reduce this visual bias. To account for the shape of geographical units and their heterogeneous population density, the distance between any two counties is here estimated as a population-weighted average of Euclidian distances between points discretizing the pair of counties [20].

The mortality risk and the associated kriging variance for a unit x are estimated as:

$$\hat{r}(x) = \sum_{i=1}^K \lambda_i z(v_i) \quad (1)$$

Kriging variance is computed as follows:

$$\sigma^2(x) = \bar{C}_R(x, x) - \sum_{i=1}^K \lambda_i \bar{C}_R(v_i, x) - \mu(x) \quad (2)$$

where x represents either an area (v_α) (ATA kriging) or a point u_s within that area (ATP kriging). The kriging weights (λ_i) and the Lagrange parameter $\mu(x)$ are computed by solving the Poisson kriging system of equations:

$$\sum_{j=1}^K \lambda_j [\bar{C}_R(v_i, v_j) + \delta_{ij} \frac{m^*}{n(v_j)}] + \mu(x) = \bar{C}_R(v_i, x) \quad i = 1, \dots, K \quad (3)$$

$$\sum_{j=1}^K \lambda_j = 1$$

where $\delta_{ij} = 1$ if $i=j$ and 0 otherwise. The “error variance” term, $m^*/n(v)$, leads to smaller weights for rates measured over smaller populations. The ATA covariances $\bar{C}_R(v_i, v_j)$ and ATP covariances $C_R(v_p, x = u_s)$ are approximated as the population-weighted average of the point-support covariance $C_R(h)$ computed between any two locations discretizing the areas v_i and v_j , or v_i and u_s . An important property of the ATP kriging estimator is its coherence: the population-weighted average of the risk values estimated at the P_α points u_s discretizing a given entity v_α yields the ATA risk estimates for this entity:

$$\hat{r}(v_\alpha) = \frac{1}{n(v_\alpha)} \sum_{s=1}^{P_\alpha} n(u_s) \hat{r}(u_s) \quad (4)$$

where $u_s \in v_\alpha$ with $s=1, \dots, P_\alpha$, and $n(u_s)$ is the population count assigned to the interpolation grid node u_s . Constraint (4) is satisfied if the same K areal data are used for the ATA kriging of $\hat{r}(v_\alpha)$ and the ATP kriging of the P_α risk values.

Deconvolution of the semivariogram of the risk: An important step in the application of the kriging techniques is the inference of

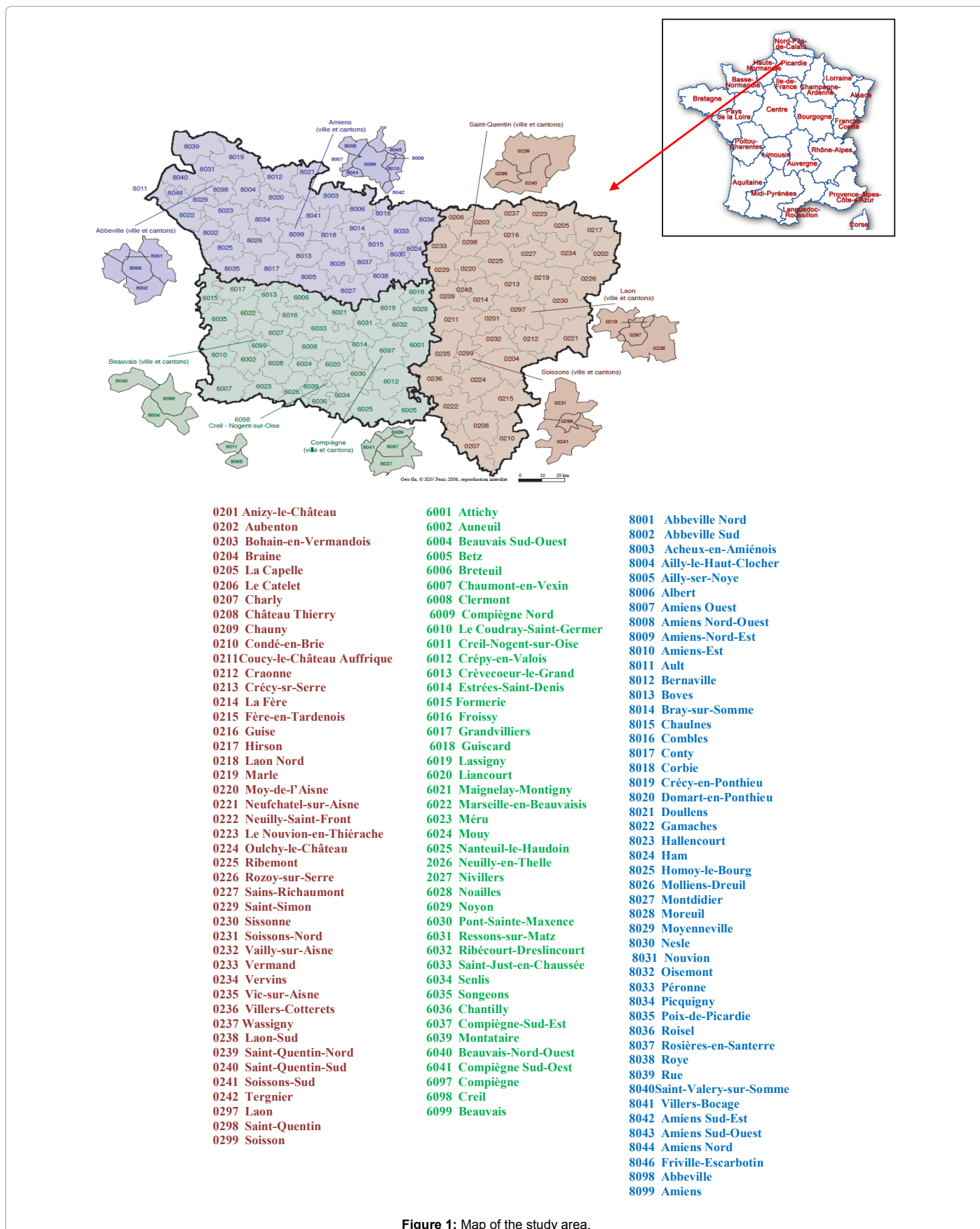


Figure 1: Map of the study area.

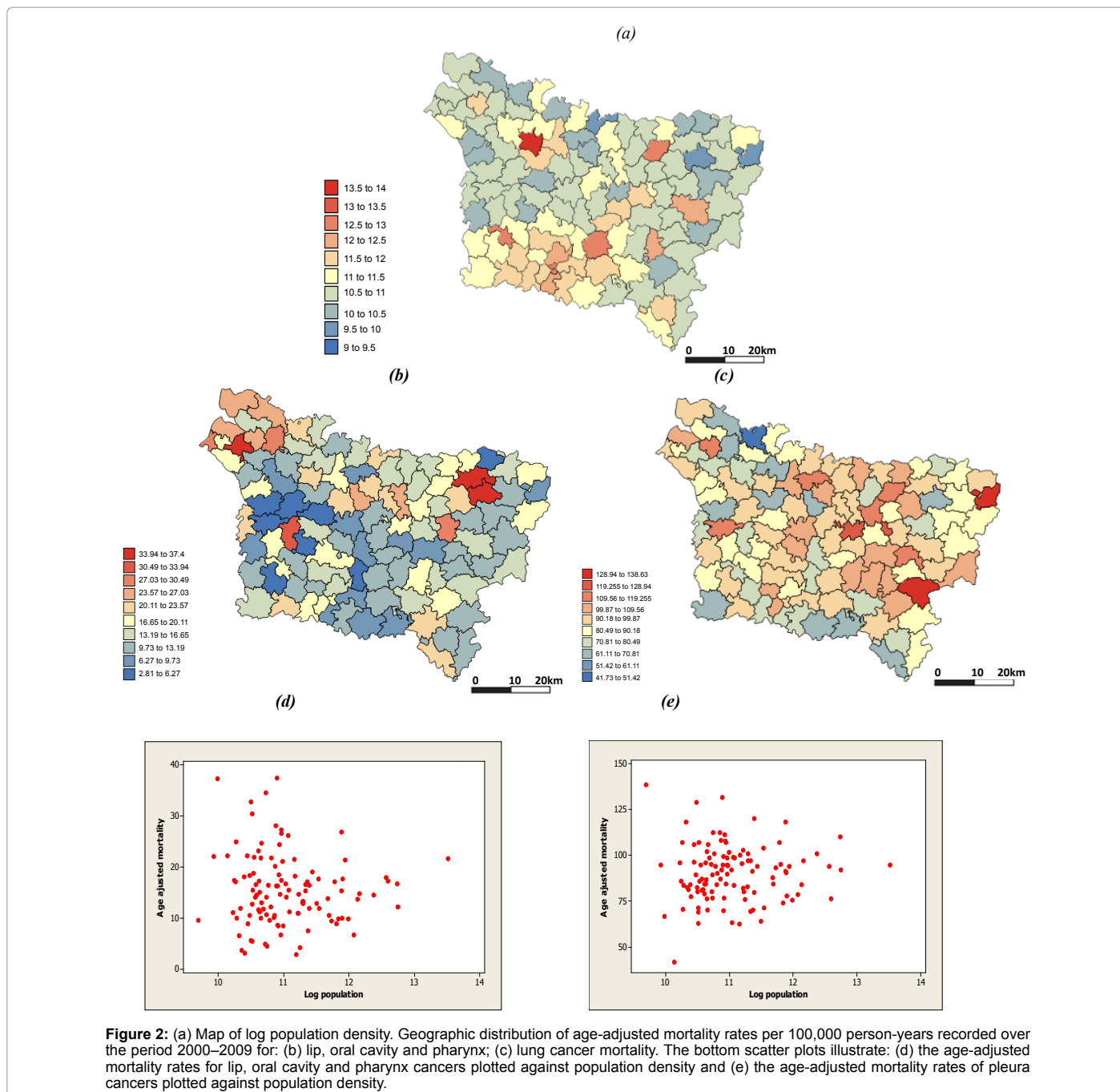


Figure 2: (a) Map of log population density. Geographic distribution of age-adjusted mortality rates per 100,000 person-years recorded over the period 2000–2009 for: (b) lip, oral cavity and pharynx; (c) lung cancer mortality. The bottom scatter plots illustrate: (d) the age-adjusted mortality rates for lip, oral cavity and pharynx cancers plotted against population density and (e) the age-adjusted mortality rates of pleura cancers plotted against population density.

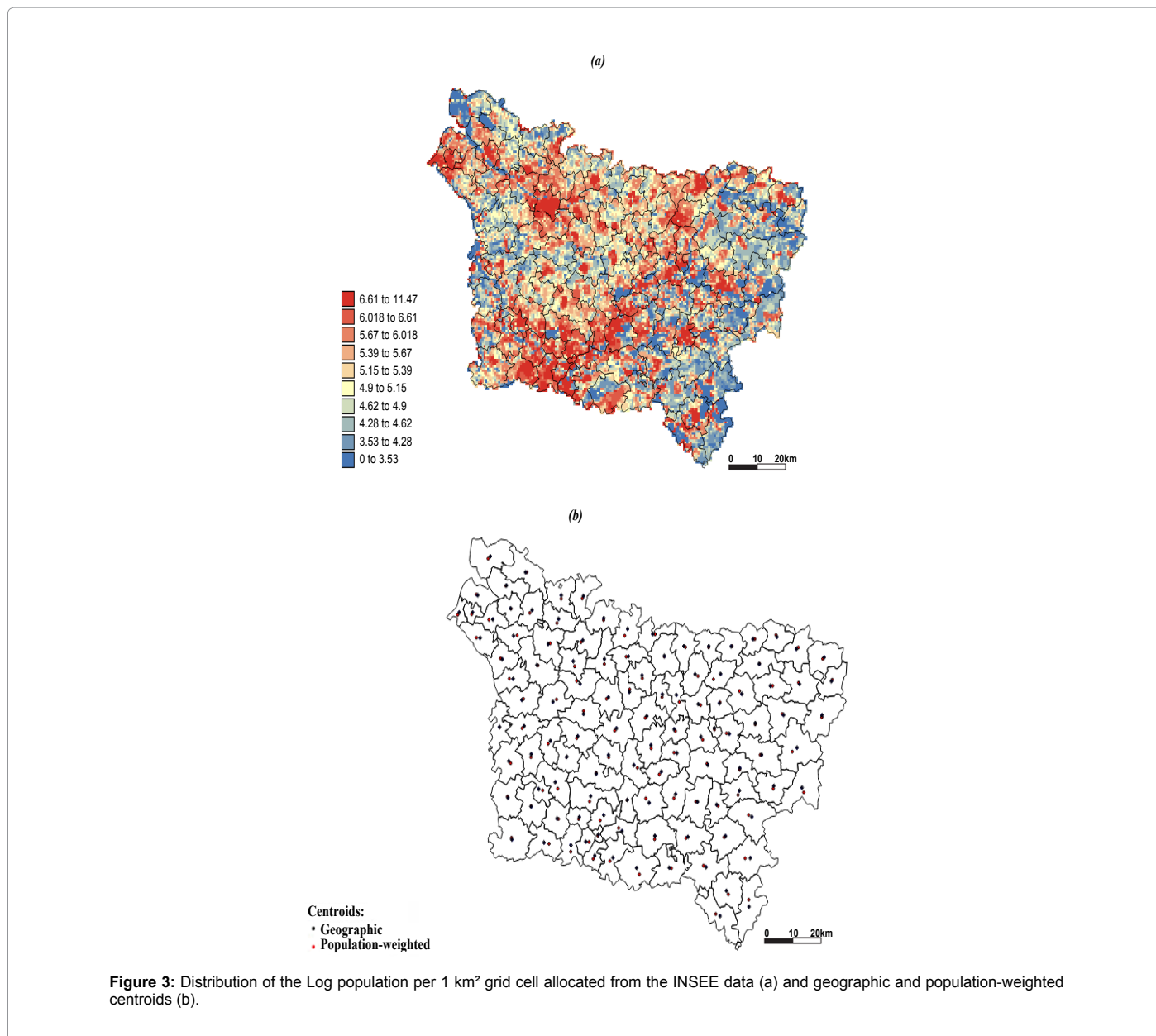
Cancer mortality	Numbers of cases	Age-adjusted rates/per 100 000 person-years
Lip, oral cavity and pharynx cancer mortality		
Cumulative	1327	16.26
Minimum	1	2.81
Maximum	128	37.4
lung cancer mortality		
Cumulative	7338	97.89
Minimum	11	41.37
Maximum	534	138.63

Table 1: Cumulative, maximum and minimum number of mortality and age-adjusted rates/per 100 000 person-years by county, 2000-2009.

the point-support variogram $\gamma_R(h)$ or, equivalently, the point-support covariance $C_R(h)$ defined as $C_R(0) - \gamma_R(h)$. This function cannot be estimated directly from the experimental variogram because the latter is computed from areal rate data. The regularized semivariogram of the risk can be estimated as:

$$\gamma_R(h) = \frac{1}{2 \sum_{\alpha, \beta}^{N(h)} \frac{n(v_\alpha)n(v_\beta)}{n(v_\alpha) + n(v_\beta)}} \sum_{\alpha, \beta}^{N(h)} \left\{ \frac{n(v_\alpha)n(v_\beta)}{n(v_\alpha) + n(v_\beta)} [z(v_\alpha) - z(v_\beta)]^2 - m^* \right\} \quad (5)$$

where, $N(h)$ is the number of pairs of areas (v_α, v_β) , the population-weighted centroids of which are separated by the vector h . The usual



squared differences $[z(v_\alpha) - z(v_\beta)]^2$ are weighted by a function of their respective population sizes, $n(v_\alpha)n(v_\beta) / [n(v_\alpha) + n(v_\beta)]$, which are inversely proportional to their standard deviations.

Results and Discussion

Figure 2 shows the spatial distribution of mortality due to the cancer of lip, oral cavity and pharynx as well as lung cancer, age-adjusted per 100 000 person-years. It should be noted that the population is not evenly distributed throughout the study area (Figure 2a), and the rate calculated for a less populated county tends to be less reliable. This implies that the interpretation of the map must be carried out with caution. The scatter plot at the bottom of Figure 2 illustrates this effect, commonly known as the "small number problem," that translates into the larger spread of mortality rates for smaller populations.

The highest age-adjusted mortality rates per 100 000 person-years recorded from 2000–2009 for lip, oral cavity and pharynx cancers were more concentrated in the north of the region, but they are generally spread throughout the area, whereas the highest rates of lung cancers were located in the eastern part of the area.

The spatial distribution of population used to avoid the constraints of county geographical boundaries in the estimation is mapped in Figure 3a. This map shows a large variability of population concentration within each county. This variability was taken into account; the geographic centroids are replaced by population-weighted centroids (Figure 3b).

Figure 4 shows the omnidirectional semivariogram of the lip, oral cavity and pharynx, and lung cancer mortality risk computed from county-level rates using an estimator (5). The semivariogram model (see theoretical regularized model in the Figure) is used to estimate the

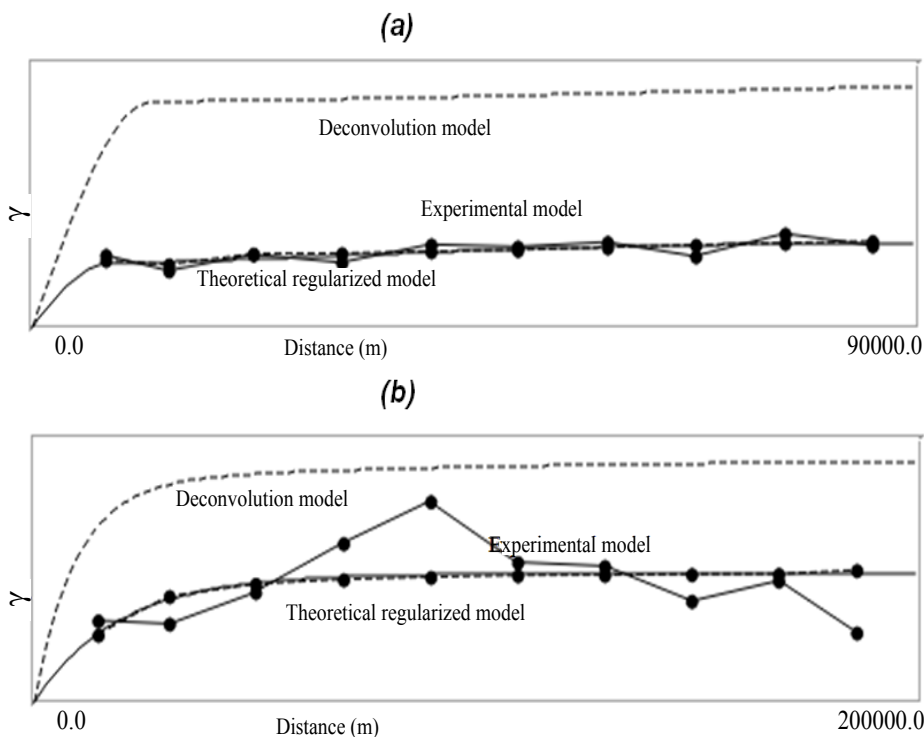


Figure 4: Semivariogram and the associated deconvolution of the lip, oral cavity and pharynx mortality risk (a) as well as the lung cancer mortality risk (b) computed from county-level rates using an estimator (5).

Lip, oral cavity and pharynx cancer mortality

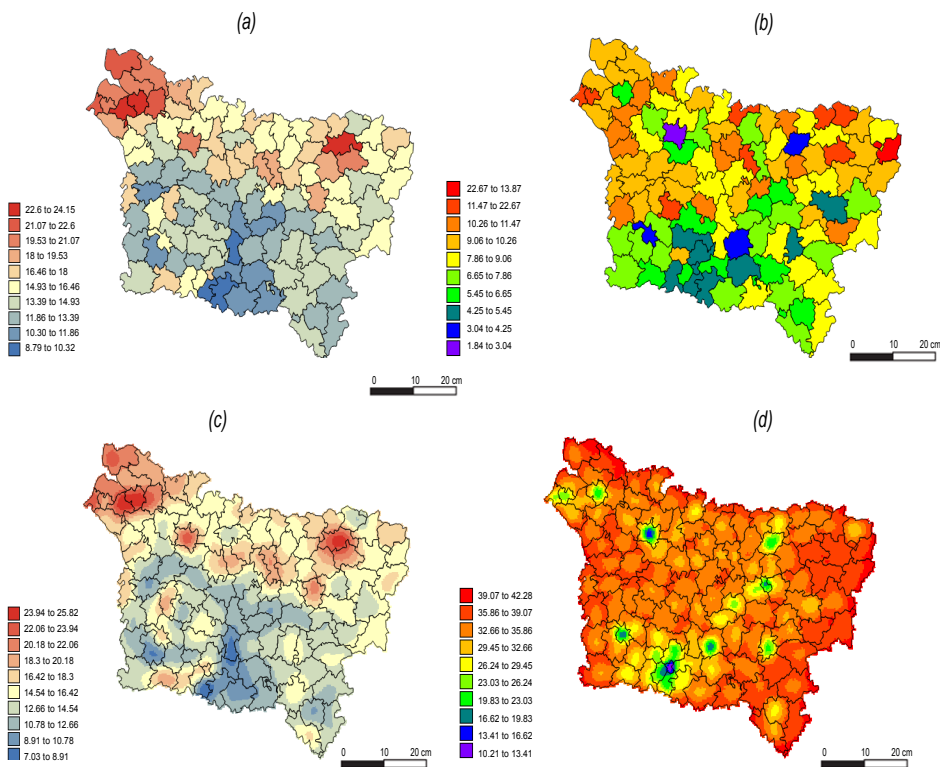


Figure 5: Maps of the lip, oral cavity and pharynx cancer risk estimated by using ATA Poisson kriging (a), and ATP Poisson kriging (c) with the corresponding prediction variance for ATA (b) and ATP (d).

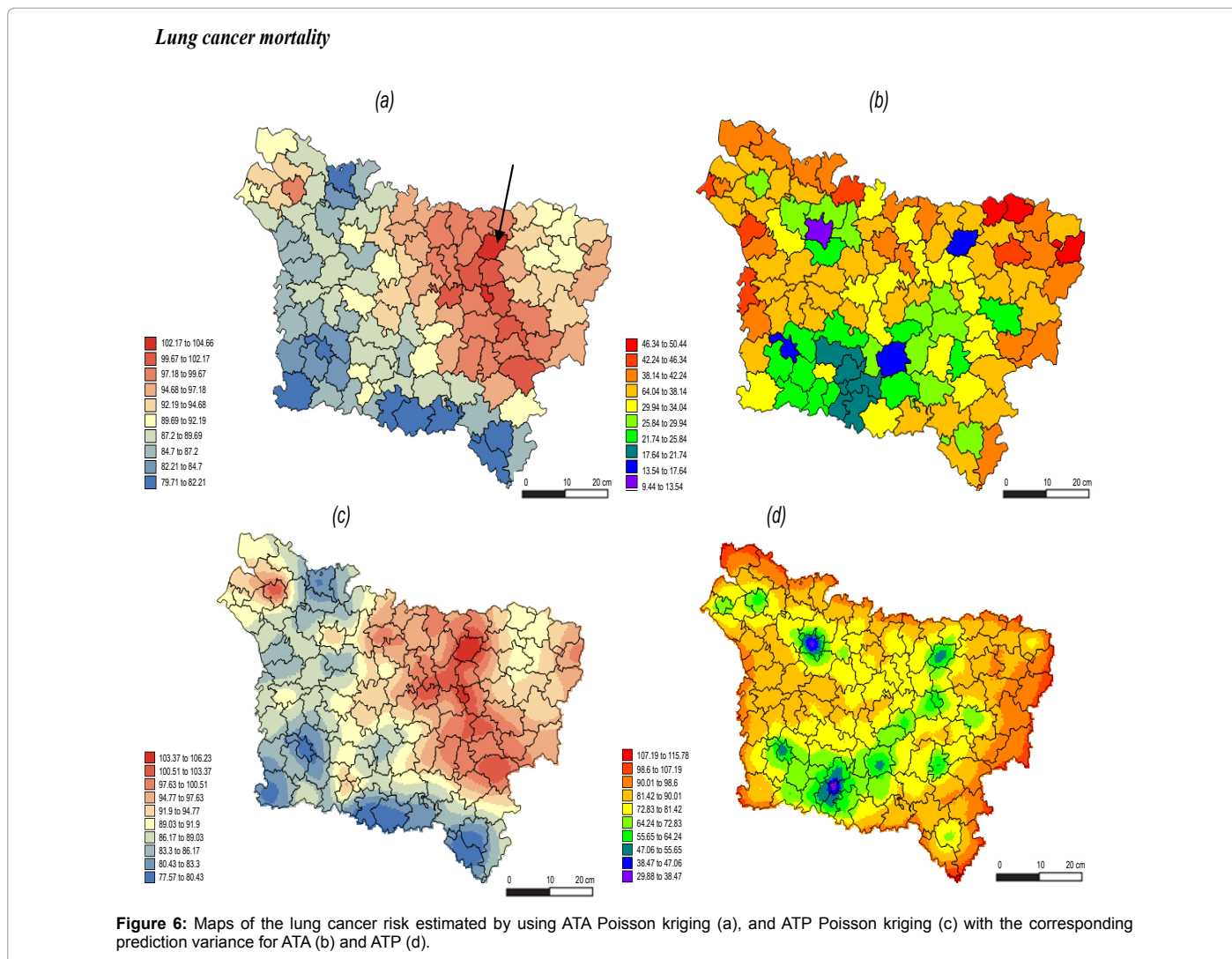


Figure 6: Maps of the lung cancer risk estimated by using ATA Poisson kriging (a), and ATP Poisson kriging (c) with the corresponding prediction variance for ATA (b) and ATP (d).

lip, oral cavity and pharynx, and lung cancer mortality risk and the associated prediction variance at the county-level (ATA kriging) or at the nodes of a 1 km spacing grid (ATP kriging).

The experimental variograms were fit using a spherical model with a range of 12.5 km for lip, oral cavity and pharynx cancer mortality and an exponential model with a range of 26.7 km for lung cancer mortality. Each model was deconvoluted using the iterative method [15]. The deconvoluted variogram model was then used to compute aggregated risk values at the county-level using ATA and ATP kriging, see Figure 3. The kriging estimate is based on the $K=32$ closest observations selected based on population-weighted distance between the counties. The noise due to the small population size was filtered; the original rate map is less smooth than all the other maps.

The lip, oral cavity and pharynx cancer mortality rate varies between 2.81 and 37.40 per 100 000 inhabitants. After the application of Poisson kriging, the minimum rate increased from 2.81 to 8.79 deaths/100 000 inhabitants, and the maximum rate decreased from 37.40 to 24.46 deaths per 100 000 inhabitants. Notably, the high rates recorded in sparsely populated counties, such as Sains-Richaumont county, (37.40 deaths/100 000 person-years), north of the Aisne department, are strongly smoothed (24.15 deaths/100 000 person-years). The highest

rates recorded in densely populated counties, such as Abbeville North county, (26.60 deaths/100 000 person-years), remain almost the same after smoothing (24.90 deaths/100 000 person-years). The map shows that the situation is favorable in the south of the region, and it is rather unfavorable in the northeast and northwest (Figure 5).

The lung cancer mortality rate varied from 41.70 to 138.63 per 100 000 inhabitants. The rate, after application of Poisson kriging ranged from 79.53 to 104.3 per 100 000 inhabitants. The highest rates recorded in densely populated counties remained the same after smoothing, such as Abbeville North county. Conversely, the highest rates recorded in the least populated counties were highly smoothed, for example, Aubenton county (Figure 6).

Compared to Figure 5, the map in Figure 6 shows a rather unfavorable situation in the northwestern region, specifically the in the Aisne department, after application of Poisson kriging in terms of lung cancer rates.

The ATP kriging risk maps are viewed as the products of the disaggregation of the ATA kriging risk maps because the ATP risk estimates are non-negative and their sum is equal to the original areal county ATA risk (Table 2). The ATP kriging map shows that high risks

	Mean	Min	Max
Amiens county			
Observed rates	21.66		
ATA Poisson kriging	20.85		
ATP Poisson kriging	20.39	19.89	20.82
Sains-Richaumont county			
Observed rates	37.40		
ATA Poisson kriging	24.15		
ATP Poisson kriging	24.28	22.52	25.23

Table 2: Summary of Kriging Estimates for Lip, oral cavity and pharynx cancer mortality by county.

are not confined to a single county but can potentially spread to areas around the county with extreme risk, (i.e., the high cancer mortality risk found in Guise county, spread to the nearby Ribemont county, (Figure 6c), which is why designing prevention strategies should not be performed at the level of a single county without taking into account the associated neighboring areas.

For a county with a large population, the ATA kriging variance map primarily reflects the highest degree of confidence in the estimated mortality risk. However, the distribution of the population can be highly heterogeneous in large counties with contrasted urban and rural areas. This information is taken into account by the kriging process. The ATP kriging variance maps highlight the location of urban centers, such as Amiens county, which are densely populated with low uncertainty in the risk assessment. Incorporating information from the high-resolution population map strengthens the impact of low or high rates in the vicinity of urban areas and helps in reducing the prediction variance around these areas. The variance of the risk estimates decreases as the area of geographical units increases: the grid-level to the county-level. The risk variance estimated for lung cancer at the county-level varies from 9.44 to 50.44 (Figure 6b), and the variance estimated at grid-level varies from 29.88 to 115.78 (Figure 6d). This uncertainty attached to the risk estimate can be incorporated in the analysis of relations along with socioeconomic and environmental factors, such as exposure indicators described in Caudeville et al. [5,6], modeled at a resolution of a 1 km² grid by weighting each estimation according to the inverse of its kriging variance. Thus, rates with a large variance will have a low weight in the analysis [21].

Several authors have already addressed the spatial relationships between health data and environmental data. One of the issues faced by spatial epidemiologists and for exposure assessment is the combination of data measured for very different spatial scales and with different levels of reliability. In reality, the analysis of cancer mortality maps is often hindered by the presence of noise caused by unreliable extreme rates computed from sparsely populated geographic units. A number of approaches have been developed to improve the reliability of risk estimates [22,23]. The most commonly used are Bayesian methods [24], which are commonly referred to as the BYM model. Bayesian methods prohibit any change of scales, an operation that is easily conducted within the framework of kriging. Goovaerts and Gebreab [25] conducted a simulation-based evaluation of the performance of geostatistical and full Bayesian disease-mapping models, and they found that the geostatistical approach yielded smaller prediction errors and more precise and accurate probability intervals and that it allowed for better discrimination between counties with high and low mortality risks.

The analysis of age-adjusted lip, oral cavity, pharynx, and lung cancer mortality rates illustrated the benefits of Poisson kriging: the

incorporation of the high-resolution population map for filtering the noise caused by small, sparsely populated areas and the estimation of the risk and associated uncertainty at fine spatial scales. The approach should facilitate the analysis of relationships between health data and putative covariates (i.e. environmental, socio-economic, or demographic factors) that are typically measured over different spatial scales [26]. These covariates could also be used directly as secondary information in area-to-point kriging, leading to more detailed risk maps at finer scale [27]. An important consideration in the interpretation of this study is that ATP kriging cannot actually create higher resolution data from areas (ATP kriging cannot realistically be a replacement for data collected at different scales). Whilst such kriging methods can provide another useful visualisation and analysis technique, they are not a substitute for higher resolution data. The original data is subject to the MAUP "modifiable area unit problem" [28,29], and therefore, the results of any analysis using this data will also have this limitation.

Conclusion

Characterizing spatial disparities in cancer mortality is a requirement for the reduction of diseases that are leading causes of death. The analysis of cancer mortality maps is often hindered by the presence of noise in mortality data, which is caused by low population densities with drastic variations in cancer rates. The methodology that we applied was based on geostatistics. It allows for both filtering noise caused by the "small number problem" and estimating the mortality risk at a fine resolution, while also taking into account the size and shape of county as well as the distribution of the population in each county. This methodology is more reliable for characterizing spatial disparities in cancer mortality, allowing for an estimation of the risk and the associated uncertainty on different scales. This form of Poisson kriging will facilitate the analysis of the relationships of cancer mortality rates with environmental and socio-economic data measured on very different supports.

Acknowledgment

The authors wish to acknowledge the financial support by the French Environment and Energy Management Agency ADEME and the French Picardy Region provided within the framework of the CIRCE project.

References

1. Leclerc A, Chastang JF, Menvielle G, Luce D (2006) Socioeconomic inequalities in premature mortality in France: have they widened in recent decades? *SocSci Med* 62: 2035-2045.
2. Melchior M, Goldberg M, Krieger N, Kawachi I, Menvielle G, et al. (2005) Occupational class, occupational mobility and cancer incidence among middle-aged men and women: a prospective study of the French GAZEL cohort*. *Cancer Causes Control* 16: 515-524.
3. Challier B, Viel JF (2001) Relevance and validity of a new French composite index to measure poverty on a geographical level. *Rev EpidemiolSantePublique* 49: 41-50.
4. Rey G, Jouglu E, Fouillet A, Hémon D (2009) Ecological association between a deprivation index and mortality in France over the period 1997-2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health* 9: 33.
5. Caudeville J, Bonnard R, Boudet C, Denys S, Govaert G, et al. (2012) Development of a spatial stochastic multimedia model to assess population exposure at a regional scale. *Sci Total Environ* 432: 297-308.
6. Caudeville J, Boudet C, Denys S, Bonnard R, Govaert G, et al. (2011) Characterization of environmental inequalities in Picardy based on a multimedia coupled model and a geographic information system use. *Environmental risks and health* 10: 239-242.
7. Waller LA, Gotway CA (2004) *Applied spatial statistics for public health data*. John Wiley & Sons, New York.

8. Lai D (2004) Geostatistical analysis of Chinese cancer mortality: variogram, kriging and beyond. *Journal of Data Science* 2: 177-193.
9. Oliver MA, Webster R, Lajaunie C, Muir KR, Parkes SE, et al. (1998) Binomial cokriging for estimating and mapping the risk of childhood cancer. *IMA J Math Appl Med Biol* 15: 279-297.
10. Goovaerts P (2005) Detection of spatial clusters and outliers in cancer rates using geostatistical filters and spatial neutral models. In: Renard PH, Demougeot-Renard H, Froidevaux R(eds), *geoENV V-Geostatistics for Environmental Applications*. Springer; Berlin 149-160.
11. Goovaerts P (2005a) Simulation-based assessment of a geostatistical approach for estimation and mapping of the risk of cancer. In: Leuangthong O, Deutsch CV (eds), *Geostatistics Banff 2004*. Dordrecht, The Netherlands 14: 787-796.
12. Kaiser M, Cressie N (1997) Modeling Poisson variables with positive spatial dependence. *Statistical and Probability Letters* 35: 423-432.
13. Monestiez P, Dubroca L, Bonnin E, Durbec JP, Guinet C (2005) Comparison of model based geostatistical methods in ecology: application to fin whale spatial distribution in northwestern Mediterranean Sea. In: Leuangthong O, Deutsch CV (eds) *Geostatistics Banff 2004*. Dordrecht, The Netherlands, Kluwer Academic Publishers 14: 777-786.
14. Monestiez P, Dubroca L (2006) Geostatistical modelling of spatial distribution of *Balaenopterophysalus* in the northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts. *Ecological Modelling* 193: 615-628.
15. Goovaerts P (2006) Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. *Int J Health Geogr* 5: 52.
16. Goovaerts P (2005) Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. *Int J Health Geogr* 4: 31.
17. Ali M, Goovaerts P, Nazia N, Haq MZ, Yunus M, et al. (2006) Application of Poisson kriging to the mapping of cholera and dysentery incidence in an endemic area of Bangladesh. *Int J Health Geogr* 5: 45.
18. Kyriakidis P (2004) A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis* 36: 259-289.
19. <http://www.or2s.fr/Portals/0/Autres%20sanitaire/Rap%20CIRCEnew.pdf>
20. Goovaerts P (2008) Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units. *Math Geol* 40: 101-128.
21. Goovaerts P (2009) Medical Geography: a Promising Field of Application for Geostatistics. *Math Geol* 41: 243-264.
22. Lawson AB (2001) Disease map reconstruction. *Stat Med* 20: 2183-2204.
23. Kafadar K (1994) Choosing among two-dimensional smoothers in practice. *Computational Statistics and Data Analysis* 18:419-439.
24. Besag J, York J, Mollie A (1991) Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43:1-59.
25. Goovaerts P, Gebreab S (2008) How does Poisson kriging compare to the popular BYM model for mapping disease risks? *Int J Health Geogr* 7: 6.
26. SaibMS, Caudeville J, Carre F, Ganry O, TrugeonA, et al. (2014) A. Spatial Relationship Quantification between Environmental, Socioeconomic and Health Data at Different Geographic Levels. *Int J Environ Res Public Health* 11: 3765-3786.
27. Gotway CA, Young LJ (2004) Ageostatistical approach to linking geographically-aggregated data from different sources. In Technical report # 2004-012. Department of Statistics, University of Florida.
28. Opensha S, Taylor PJ (1979) A Million or So Correlation Coefficients: Three Experiments on the Modifiable areal Unit Problem. In *Statistical Methods in the Spatial Sciences*; Pion: London, UK.
29. Openshaw S (1984) *The Modifiable Areal Unit Problem Concepts and Techniques in Modern Geography*; Geo Books: Norwich, UK, Available online: <http://qmrq.org.uk/files/2008/11/38-maup-openshaw.pdf> (accessed on 31 March 2014).