Journal of Data Science **2**(2004), 177-193

Geostatistical Analysis of Chinese Cancer Mortality: Variogram, Kriging and Beyond

Dejian Lai

University of Texas and Jiangxi University of Finance and Economics

Abstract: In this study, we used geostatistical tools to spatially analyze the Chinese cancer mortality rates. We first quantified the spatial variations of the observations using the variogram and fitted spherical and exponential parametric models to the sample variograms. Then, utilizing the fitted variogram function, we performed ordinary Kriging on the Chinese cancer mortality rates based on both models and produced a set of contour maps.

Key words: Chinese cancer mortality, contour mapping, geostatistical analysis, Kriging, spatial correlation, variogram.

1. Introduction

The mainland of China is located between longitude east 70° and 135° and latitude north 15° and 60°. China, the largest developing country with over one fifth of the world population, does not have a comprehensive official statistics for the mortality of its population, especially for its 80% rural population. Although some of the mortality information has been collected for the immediate year prior to the population census in 1982, 1990 and 2000 (SSB 1985, 1993, 2002), the information on the causes of death was not collected.

Demographic data of China are collected by various governmental agencies. The State Statistical Bureau of China (SSB, it is now called National Bureau of Statistics in English. However, the Chinese name remains the same) conducts population censuses and surveys. Since the establishment of the People's Republic of China in 1949, there have been five population censuses. The first and the second population censuses were held in 1953 and 1964, respectively. Because of constraints of the technology at that time, the data collected from these two censuses were very limited and there was a lack of cross-tabulated tables. Beginning with the third population census in 1982, computer technology has been used in data processing and extensive cross-tabulation of the data become possible. The fourth and the fifth population censuses were held in 1990 and 2000, respectively. In addition to the population censuses. There are two large scale surveys conducted in 1987 and 1995 on one percent of the population. There are also annual one per thousand population surveys. The population surveys administrated by the State Statistical Bureau paid more attentions on the economic activity of the population than on its social development. In China, as far as the mortality information of the population is concerned, since majority of the deaths occurred outside hospitals, especially the old people living in rural areas, it is not an easy task to collect data on causes of death for the entire population. Due to culture and tradition as well as cost, there are very fewer coroners in China except for forensic purposes. Hence, China does not have comprehensive cause specified mortality data. For these reasons, statistical analysis based on survey data are essential and useful for further inferences.

The Ministry of Public Health (MOPH) of China collects most health related information such as statistics on hospitals, allocation of resources with a special attention on infectious diseases. There has been a strict residential registration system in China. Chinese population are generally categorized into two strata: rural (agricultural) and urban (non-agricultural) residents. Demographic information on the urban residents are much better than that of the rural residents. In this article, we focus our attention on the rural population of China. It is worth noting that permanent migration between rural and urban classification were strictly controlled. Most of the professional (non-agricultural) positions are reserved for urban residents. Passing the college entrance examinations seemed the only way for people of agricultural (rural) status can be permitted to change their status from rural to urban. Recently, this type of control is getting relaxed in some cities. The task for registration of residential status belongs to the Ministry of Public Security. Theoretically, the Ministry of Public Security has the basic records on deaths, birth and migration. However, at least for the rural population, these administrative data were far from accurate. China has implemented strong family planning programs, especially for its urban residents, in the last two decades. The Commission of Family Planing monitors the fertility closely. Perhaps, China has the most rigid fertility control system in the world. But in the vast remote rural areas, the system is much less rigid. The nuptial information is collected by the Ministry of Civic Affairs and education statistics are collected by the Ministry of Education. All these demographic data are reported to the State Statistics Bureau and some of the summary information is not publicly available.

In 1976, there was a large scale sampling survey conducted by the Ministry of Public Health on mortality in 65 rural counties in China. The survey collected detailed information on the causes of death for over 20 million rural residents. Rural residents are mostly living in counties although some of the counties may be under the administration of cities. There are more than 2,000 counties in China. The survey was based on multi-stage sampling design, which resulted 65 counties with population size greater 100,000 residents for every counties selected. The selected counties were located in 24 provinces. Considering the distribution of the counties, the sample of the counties were not strictly based on random selection. However, within a counties, a three-stage random sampling was implemented. For each selected county, two *xiangs* (communes) were selected. Then one *villiage* (brigade) was selected from each *xiang*. The *villiage* as an administrative unit is a cluster of natural villages. In each *villiage*, two production teams were selected. A production team may contain several natural villages when the natural villages are small. For a large village, there may be more than one production teams. Since 1980s, the production team as an administration unit disappeared, which was replaced by *group of peasants* with less administrative role. The comprehensive results from the survey were published in 1990 together with the results of a nutritional survey conducted in the same counties (Chen *et al.* 1990). Several articles on the survey results have been published prior to or after the release of the results (Armstrong 1980, Li *et al.* 1981, Hsing *et al.* 1991). However, there seems no analytical spatial statistical analysis on the results except a few descriptive spatial studies on the provincial level (Lam 1986, Lai 1997).

Spatial statistical (geostatistical) techniques have been widely used in many disciplines (Cressie 1993, Cliff and Ord 1981). Recognizing that the data from many surveys on public health issues are spatially correlated, we need to extend the traditional statistical methods that are based on the independent assumptions of the observations to the geostatistical techniques that can incorporate the spatially dependent structure of the data for a better statistical inference. Reviews of the spatial statistical analysis on disease can be found in Gesler (1986) and Marshall (1991). More recent references on cancer mapping are given by Best and Wakefield (1999) and Pickle *et al.* (1999).

Graphic study of health data was not new to public health workers. The famous study on cholera by John Snow in 1848 is a classic example (Snow 1855). Extensive descriptive statistical tests have been used in detecting disease patterns (Besag and Newell 1991, Cliff and Ord 1981, Cuzick and Edwards 1990, Walters 1992a, 1992b, 1994). However, from an analytical point of view, few applications of the well-established geostatistical methods on health data were found. One example is Oliver *at al* (1992) on a rare disease.

In our study, we applied geostatistical methods on the Chinese cancer mortality rates of the 63 rural counties (For reasons given in the next section, we omitted the data of 2 counties) and produced contour maps of the cancer mortality rates of the rural areas in mainland China (longitude east of 100°). Our study also established analytic geostatistical models for the Chinese cancer mortality rates of the rural China. The models can be used to estimate the cancer mortality

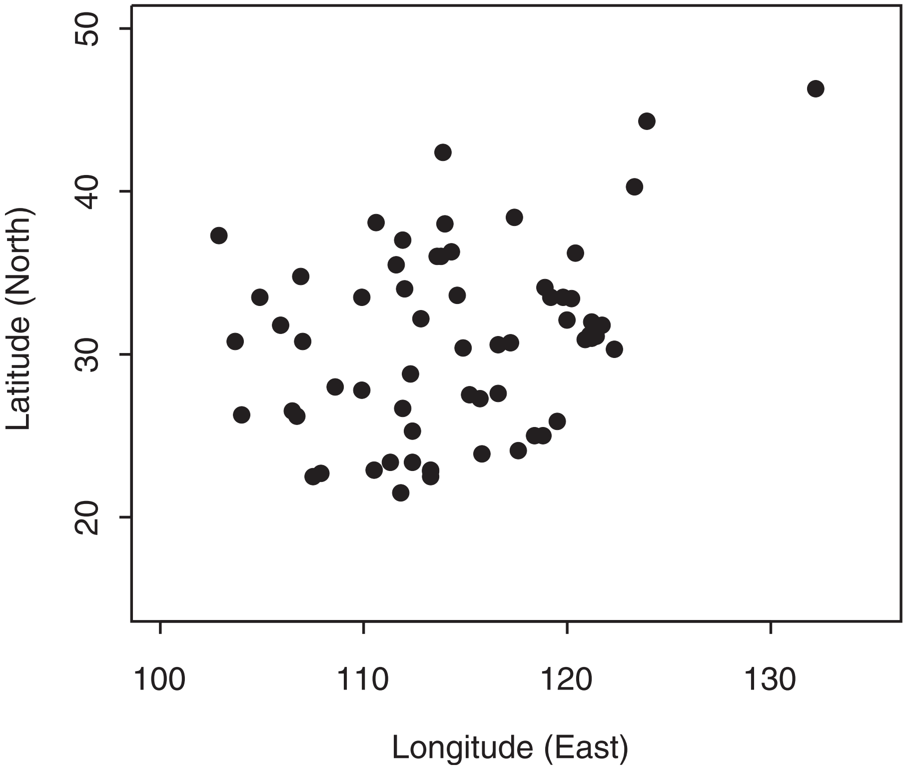


Figure 1: Location of the surveyed rural counties of China east of longitude east 100°.

rates by using the geographic locations (longitude, latitude) for other rural counties under concern that were not in the sample. The models can also be used to smooth the observed cancer mortality rates for the counties in the sample. There has been a lack of geostatistical study on Chinese mortality in the literature. Our study provided an example of using statistical models although the data were collected more than twenty years ago.

2. Materials and Methods

Most of the Chinese people (more than 90%) live in the east of longitude east 100° (SSB 2002) and there were only two samples in our data set in the west of longitude east 100°. Therefore we focused our attention on the areas of longitude east 100° without including the two far west observations of Tuoli (83.7°, 46.0°) and Dunhuang (94.8°, 40.1°).

In the mortality survey, in addition to cancer mortality, mortality due to noncancer were also collected for 49 of the 65 counties. The cancer mortality rates were classified by 16 cancer sites: nasopharynx, esophagus, stomach, liver, rec-

Table 1: The longitude (east), latitude (north) and the total, male, female cancer mortality rates in 1975 between ages 0 and 64 of the 65 rural counties of China.*

County	Longitude	Latitude	Total	Male	Female
Shanghai	31.1	121.4	89.93	120.08	59.77
Qingpu	31.2	121.1	93.94	116.50	71.37
Songjiang	31.0	121.2	76.01	96.09	55.93
Cixian	36.3	114.3	159.29	196.12	122.46
Jingxing	38.0	114.0	103.55	126.33	80.78
Huanghua	38.4	117.4	66.74	71.79	61.70
Huguan	36.0	113.6	145.72	187.15	104.29
Jiangxian	35.5	111.6	168.73	221.17	116.28
Jiexiu	37.0	111.9	59.59	59.49	59.69
Shangshui	33.6	114.6	41.97	47.82	36.12
Linxian	36.0	113.8	186.25	221.33	151.18
Songxian	34.0	112.0	135.59	172.54	98.63
Xiuyan	40.3	123.3	85.30	107.40	63.21
Changling	44.3	123.9	67.14	79.61	54.67
Baoqing	46.3	132.2	53.76	72.82	34.69
Laoshan	36.2	120.4	94.29	115.02	73.57
Shuyang	34.1	118.9	89.47	115.57	63.38
Huaian	33.5	119.2	174.05	211.08	137.02
Yangzhong	33.4	120.2	212.19	253.01	171.36
Jianhu	33.5	119.8	148.93	190.08	107.77
Qidong	31.8	121.7	105.66	147.61	63.71
Haimen	32.0	121.2	106.59	150.29	62.89
Taixing	32.1	120.0	140.71	176.32	105.10
Zongyang	30.7	117.2	121.43	146.29	96.57
Qianshan	30.6	116.6	91.10	111.77	70.43
Daishan	30.3	122.3	145.55	206.76	84.35
Jiashan	30.9	120.9	116.60	146.91	86.28
Zhangpu	24.1	117.6	114.54	166.25	62.82
Nanan	25.0	118.4	125.20	175.53	74.86
Changle	25.9	119.5	170.98	150.22	91.75
Huian	25.0	118.8	138.49	183.38	93.60
Lean	27.3	115.7	44.78	50.54	39.01

* Cancer mortality rates are given in per 100,000.

Table 2: The longitude (east), latitude (north) and the total, male, female cancer mortality rates in 1975 between ages 0 and 64 of the 65 rural counties of China (continued).*

County	Longitude	Latitude	Total	Male	Female
Nanchang	27.6	116.6	92.03	137.22	46.84
Xiajiang	27.5	115.2	51.91	57.51	46.31
Linwu	25.3	112.4	56.36	78.40	34.32
Mayang	27.8	109.9	37.05	42.15	31.95
Qiyang	26.7	111.9	52.89	59.33	46.44
Yuanjiang	28.8	112.3	88.69	91.66	85.72
Zaoyang	32.2	112.8	48.28	59.41	37.15
Echeng	30.4	114.9	89.89	113.67	66.12
Cangwu	23.4	111.3	43.46	57.49	29.42
Chongzuo	22.5	107.5	54.81	80.18	29.44
Fusui	22.7	107.9	85.80	132.77	38.83
Rongxian	22.9	110.5	25.31	28.59	22.04
Qingzhen	26.5	106.5	31.57	33.58	29.56
Yinjiang	28.0	108.6	29.94	27.04	32.83
Huishui	26.2	106.7	12.38	12.31	12.44
Xuanwei	26.3	104.0	60.37	64.56	56.18
Wenjiang	30.8	103.7	41.48	46.97	35.99
Cangxi	31.8	105.9	138.03	154.6	123.43
Quxian	30.8	107.0	83.93	101.7	66.15
Shanyang	33.5	109.9	101.53	116.3	86.71
Jiaxian	38.1	110.6	86.12	98.4	73.84
Longxian	34.8	106.9	60.51	66.0	54.95
Sihui	23.4	112.4	51.87	66.9	36.75
Panyu	22.9	113.3	59.39	73.6	45.13
Zhongshan	22.5	113.3	53.29	65.7	40.88
Wuchuan	21.5	111.8	30.78	41.5	20.00
Shunde	22.8	113.3	76.63	104.6	48.61
Wuhua	23.9	115.8	39.11	43.5	34.67
Tianzhu	37.3	102.9	107.90	134.1	81.70
Dunhuang	40.1	94.8	85.25	109.7	60.81
Wudu	33.5	104.9	71.25	58.4	84.07
Tuoli	46.0	83.7	113.21	136.1	90.23
Xianghuang Qi	42.4	113.9	93.80	100.2	87.31

* Cancer mortality rates are given in per 100,000.

tum, colon, colorectal, lung, breast, cervix, leukemia, bladder, penis, lymphoma, choriocarcinoma and brain.

For our analysis in this article, as an example, we used the total cancer mortality rates of age groups 0-64 and the geographic locations (longitude, latitude) of the 63 Chinese rural counties from the published source (Chen *et al.* 1990). For males, the mean cancer mortality rate was 0.11212 percent with standard deviation 0.05881, minimum 0.01231 and maximum 0.25301. For females, the mean cancer mortality rate was 0.06726 percent with standard deviation 0.03267, minimum 0.01244 and maximum 0.17136. The detailed description of the data can be found in Chen *et al.* (1990). For an easy reference, we extracted the mortality rates together with the longitude and latitude of the 65 counties for the total, males and females (Tables 1 and 2). The longitude and the latitude of the counties under study are plotted in Figure 1.

Let $Z(X_i)$ denote the cancer mortality rate of place (county) X_i, where $X_i = (x_{i1}, x_{i2})$, x_{i1} is the longitude and x_{i2} is the latitude. For a stationary process $Z(X_i)$, we define the variogram, which is also called semi-variogram, by

$$\gamma(h) = (1/2)E(Z(X_i) - Z(X_i + h))^2, \tag{2.1}$$

where $h = (h_1, h_2)$ is the difference of the longitude and the latitude between the two places $(X_i,\ X_i + h)$.

To estimate $\gamma(h)$, we used Matheron's method (Cressie 1993) via S-plus (Insightful 2001):

$$2\hat{\gamma}(h) = (1/|N(h)|) \sum_{N(h)} (Z(X_i) - Z(X_j))^2, \tag{2.2}$$

where the sum is over $N(h) = \{(i, j), \|X_i - X_j\| = \|h\|\}$, $|N(h)|$ is the number of distinct elements of $N(h)$ and $\|h\|$ is the Euclidean distance of h from the origin (0,0). In most applications, $\|X_i - X_j\|$ may not be exactly the same within a given neighborhood (group), which is similar to the case of constructing histogram. In our calculation, we used the width of the group being the maximum of the distances divided by the number of groups. Using S-plus function variogram, the number of groups can be specified. In our application, we selected the group number being 20, which seemed a reasonable number. The tolerance of the width was chosen to be one-half of the width. In estimating the variogram, we can specify the direction. For example, to take into account the east-west variation, one may compute the variogram according to equation (2.2), but restrict the pairs in the east-west direction. In actual application, observations were usually not exactly on the east-west direction. In this case, one needs to specify the angle tolerance. The S-plus function variogram can take this option. In our study, since we do not have a specified direction in mind, we used omnidirectional variogram.

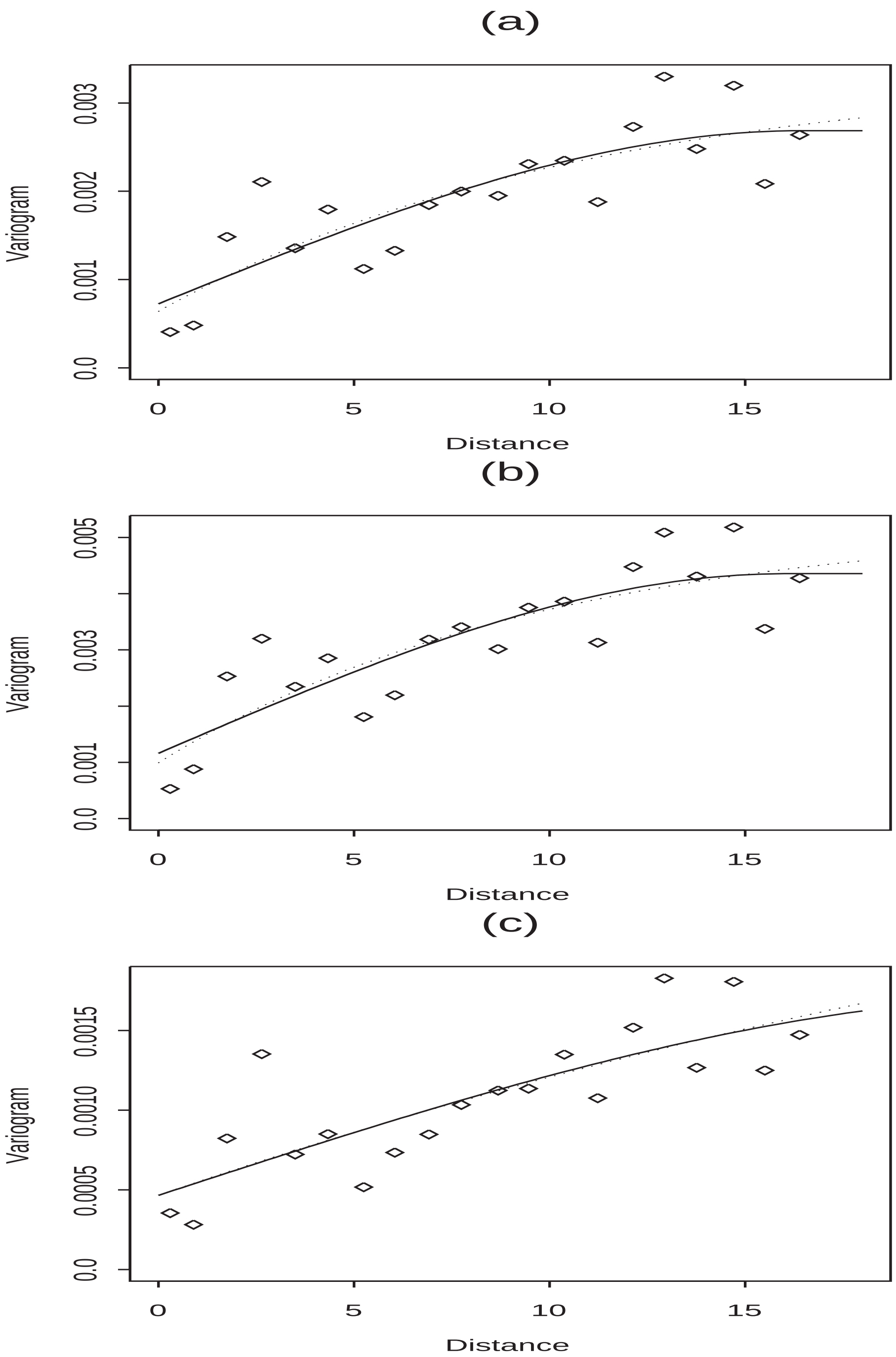


Figure 2: Empirical variogram and the fitted curves (solid: spherical model; dashed: exponential model): (a) total population; (b) male population; (c) female population.

That is, the angle tolerance is 90°. The method of estimating the variogram used in our study was a classic approach based on method of moments. One may select the robust version (Cressie 1993) which is also available in S-plus function *variogram*.

In order to carry out Kriging for optimal linear spatial prediction, one needs to fit a parametric model over the estimated empirical variogram. Several possible choices of the models such as the linear model, exponential model, rational quadratic model, wave model and spherical model are available (Cressie 1993). The plot of the empirical variogram $\hat{\gamma}(h)$ shown in Figures 2(a) to 2(c) suggests a spherical model (solid line in Figure 2). We also tried the exponential model (dashed line in Figure 2). From the empirical (sample) variogram $\hat{\gamma}(h)$ for various h (We had 20 of them in our study), we first fitted the spherical parametric model over the empirical $\hat{\gamma}(h)$. The spherical model is defined (Cressie 1993) as

$$\gamma(h,\theta) = \begin{cases} 0 & h = 0, \\ c_0 + c_s\{(3/2)(\|h\|/a_s - (1/2)(\|h\|/a_s)^3\}, & 0 < \|h\| \leq a_s, \\ c_0 + c_s, & \|h\| \geq a_s, \end{cases} \tag{2.3}$$

where $\theta = (c_0, c_s, a_s)'$ and $c_0 \geq 0$, $c_s \geq 0$, $a_s \geq 0$. In the spherical model described in equation (2.3), c_0 denotes the nugget effect that measures the microscale variation which may be resulted from discontinuous process of the mortality rates among the counties; $c_0 + c_s$ is the sill that quantifies the variability of the mortality rates in long distance, if c_s =0, $\hat{\gamma}(h)$ is a constant for all $\|h\| > 0$ that indicates no spatial correlation of the mortality rates among the counties; a_s is the range that can be interpreted as the mortality rates are uncorrected beyond distance a_s.

In addition to the spherical model, for comparison purpose, we then fitted also the exponential model on the empirical variograms. The exponential model is defined (Cressie 1993) as

$$\gamma(h,\theta) = \begin{cases} 0 & h = 0, \\ c_0 + c_s\{1 - \exp(-\|h\|/a_s)\}, & \|h\| \neq 0, \end{cases} \tag{2.4}$$

where the parameters in the exponential model have the same interpretation as that in the spherical model. However, the correlation decays to zero in the exponential model. In the spherical model, there is no correlation for the observations when the distance is beyond a_s.

Cressie (1993) provided detail explanations of the meaning of the parameters in the variogram model. Using non-linear least squares regression, we estimated the parameters of the above variogram models of the cancer mortality rates for the total population, the males and the females of the Chinese rural counties.

With the established model of the variogram and the observed values from the sampling survey, we performed the ordinary Kriging of the Chinese cancer mortality rates. The ordinary Kriging is an optimal spatial prediction of the Z process. Mathematically, for a selected place (county) with location X_0 , we search for the linear predictor $\sum_{i=1}^{n} \lambda_i Z(X_i)$, where $Z(X_1), \ldots , Z(X_n)$ are known observations from the survey, that minimizes $E(Z(X_0) - \sum_{i=1}^{n} \lambda_i Z(X_i))^2$ subject to $\sum_{i=1}^{n} \lambda_i = 1$ (Cressie 1993). That is, by the Lagrange multiplier method, we minimize

$$E(Z(X_0) - \sum_{i=1}^{n} \lambda_i Z(X_i))^2 - 2m(\sum_{i=1}^{n} \lambda_i - 1),$$

with respect to λ_1, ..., λ_n and m. The solution of $\lambda_0=(\lambda_1, ..., \lambda_n, m)'$ is given by

$$\lambda_0 = \Gamma_0^{-1} \gamma_o, \tag{2.5}$$

where $\gamma_0 = (\gamma(X_0 - X_1), ..., \gamma(X_0 - X_n), 1)'$ and

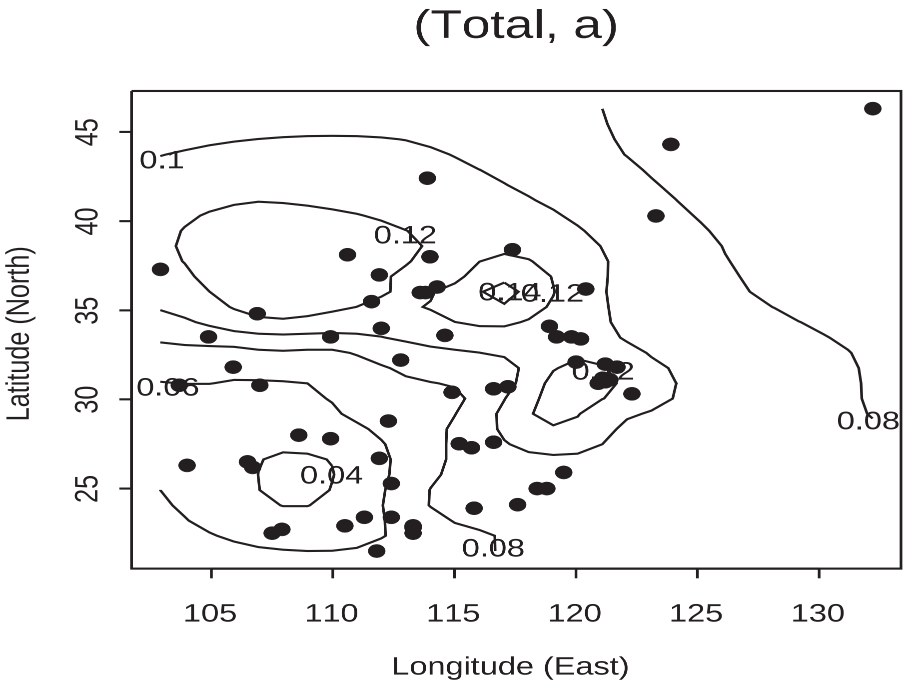
$$\Gamma_0 = \begin{cases} \gamma(X_i - X_j), & i = 1, ..., n, j = 1, ..., n, \\ 1 & i = n+1, j = 1, ..., n, \\ 0 & i = n+1, j = n+1. \end{cases}$$

Ordinary Kriging provides us an optimal way of estimating the value of Z (cancer mortality rate) at any given location (county). We used these estimates to form contour maps of the cancer mortality rates of the region under study.

3. Results

The empirical variograms and their fitted spherical and exponential curves are shown in Figures 2(a) to 2(c) for the cancer mortality rates of the total population, the males and the females, respectively. These plots suggest that spatial correlation of the cancer mortality rates exist. If there were no spatial correlation among the cancer mortality rates, the variograms would be flat across the horizontal line (distance).

The estimates ($\hat{\theta}$) of the parameters (θ) in the spherical model of the cancer mortality rates of the Chinese rural counties were (0.0007, 0.0020, 16.4102), (0.0012, 0.0032, 16.0325) and (0.0005, 0.0008, 23.4848) for the total population, the males and the females, respectively. By the exponential model, they were (0.0006, 0.0027, 10.9062), (0.0010, 0.0043, 9.9175) and (0.0005, 0.0031, 37.0290), respectively. Although there were large differences between the ranges from these two model, from Figure 2, there seemed little difference between these two models within the distance interval for the empirical variograms estimated from our data set.



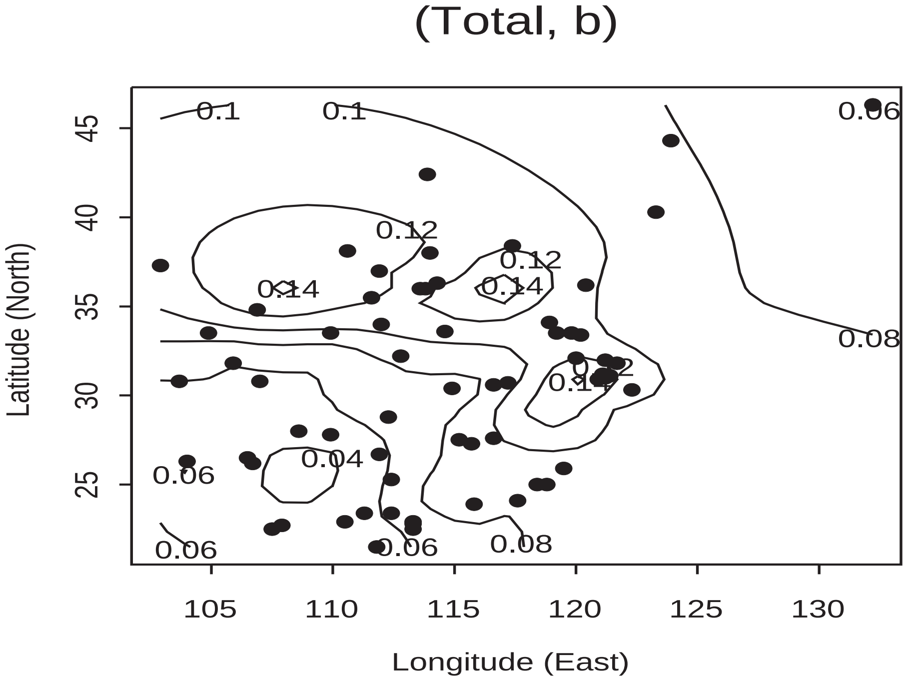


Figure 3: Contour map of the cancer mortality rates of the total population (0-64): (a) spherical model; (b) exponential model.

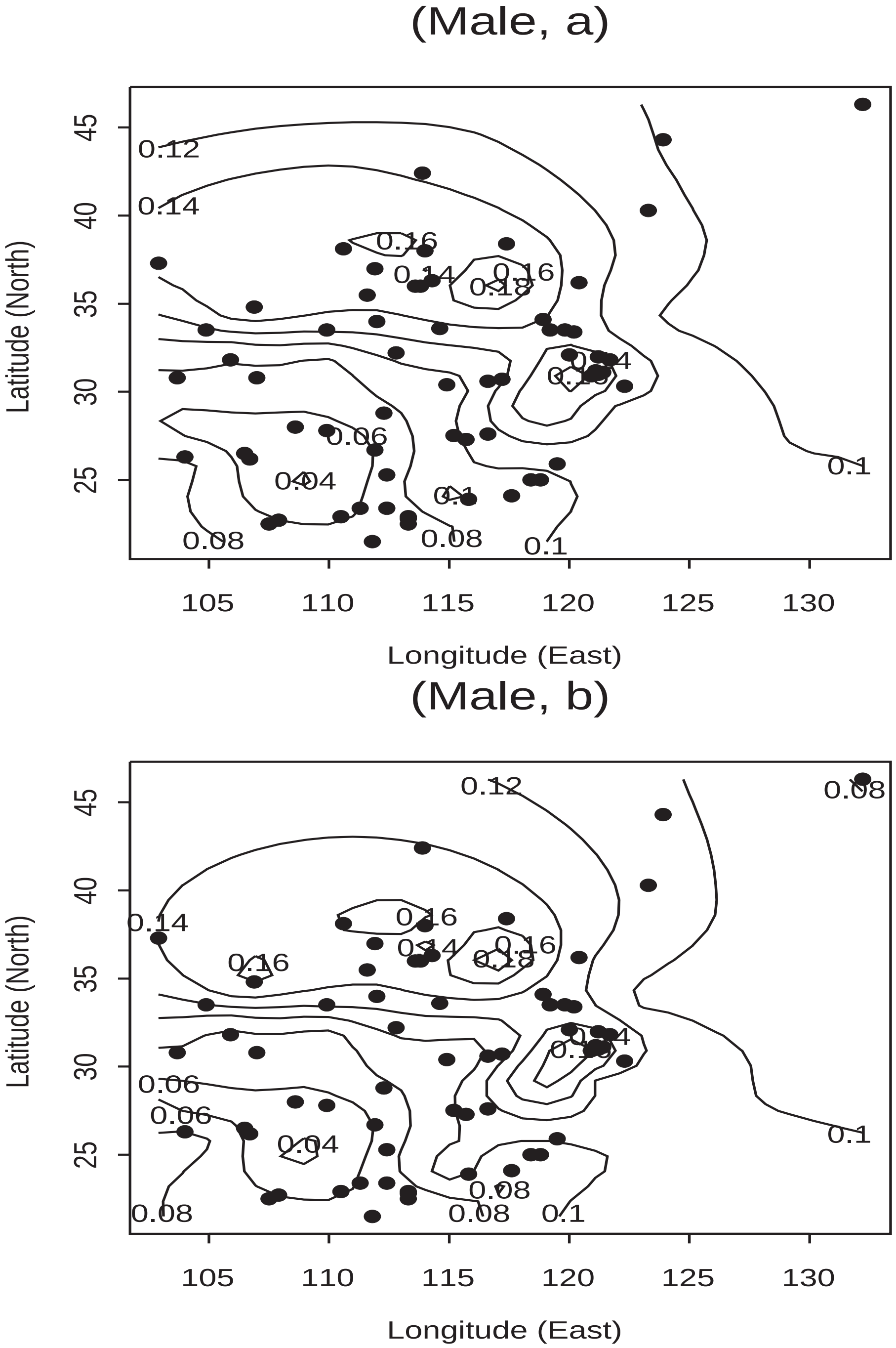
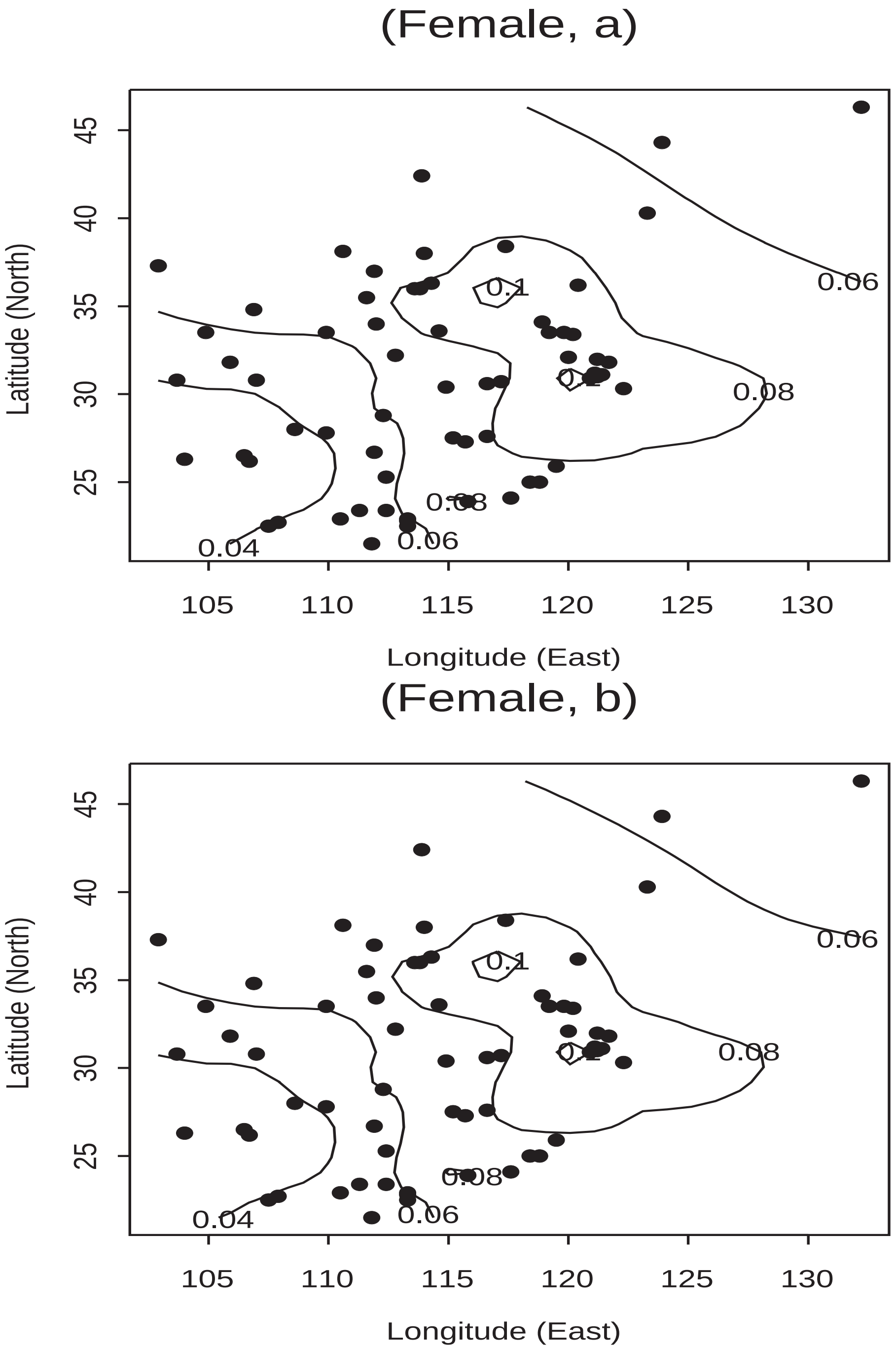


Figure 4: Contour map of the cancer mortality rates of the male population (0-64): (a) spherical model; (b) exponential model.


(Female, a)

(Female, b)

Figure 5: Contour map of the cancer mortality rates of the female population (0-64): (a) spherical model; (b) exponential model.

Using the fitted variograms, we established an optimal linear prediction (Kriging) of the morality rates for all possible places (rural counties in China) within the study area. As an example, suppose that we are interested in estimating the cancer mortality rates for Ningdu county, a rural county in Jiangxi province which was not in the data set. The location of the county is 115.48° of longitude east and 26.22° of latitude north (Defense Mapping Agency 1979). The estimated cancer mortality rates by the spherical model of age group 0-64 for the total population, the males and the females were 0.06943%, 0.0876% and 0.0502%, respectively. Using the exponential model, we got 0.0665%, 0.0832% and 0.0502%, which were slightly less than that from spherical model.

Based on the results of the Kriging, we formed contour maps for the study areas. The results using both variogram models are shown in Figures 3, 4 and 5 for the total population, the males and the females, respectively. From these contour maps, we can see that the cancer mortality rates were higher in the central China (around Henan and Shanxi provinces) and lower in the southwestern China (around Guizhou and Yunnan provinces). The cancer mortality rates in the eastern China (around Shanghai, Jiangsu and Zhejiang provinces) were also higher than most of other places in China. Both spherical model and exponential model produced similar results.

4. Discussion

People live in the environment of competing risks (Lai and Hardy 1999). The eastern China is the most developed area in China. In this area, the life expectancy at birth in 1989 for the total population of Jiangsu province, Shanghai city and Zhejiang province was 71.01, 73.27 and 70.71 respectively, which are higher than those of Guizhou province (65.87) and Yunnan province (64.04) in the southwestern China (SSB 1995). This fact indicates that people in the more developed areas usually did not die from "curable" diseases and live longer. Hence, people in eastern China have higher chance subjecting to degenerative diseases and other "non-curable" diseases such as cancer (Lai 1997). The life style of the population plays an important role also in developing cancer. In the central China (for example, Henan province and Shanxi province), the climate is usually drier and the winter is longer than that in the southern China. People in the central China have a long tradition of consuming preserved food that led to a higher cancer mortality rates in the central China.

The published results in Chen *et al.* (1990) included the cancer mortality rates by site. The methods described in previous sections can be applied to generate the contour maps of the cancer mortality rates by site. The purpose of this article is to show that geostatistical methods are useful in making statistical inferences for spatially correlated health survey. We therefore omitted the completed contour maps for cancer mortality rates by site.

Recently, the Ministry of Public Health of China and Chinese Academy of Preventive Medicine (MOPH and CAPM 1992) have established a relative comprehensive disease surveillance system over 145 points with 10 million population (less than 1% of the total population of China). If the geographic information and the statistics are available from these disease surveillance points (DSPs), the methods studied in this article can be applied to produce a more complete annual report (MOPH and CAPM 1992).

Extensive Pearson correlation coefficients were calculated between cancer mortality rates and other variables without taking into account the spatial structure of the observations (Chen *et al.* 1990). It is known that spatial correlation would lead to inaccurate conclusion based on Pearson correlation coefficient (Lai *et al.* 2003).

In our study, we focused our attention only on the spatial structure of the cancer mortality rates without correlating with other possible covariates. A better way of statistically analyzing the data might be to correlate the cancer mortality rates with the covariates by taking into account the spatial correlations. However, this comprehensive approach may not be able to produce the contour maps since we cannot have the covariate information for non-surveyed places.

It is worth noting that our modeling level is at the county. For estimating the cancer mortality rates for other places, we shall only apply the models and the contour maps to the counties. Hence, the resulting values from the model are not for an arbitrarily chosen location. Further, since the survey was conducted for the rural counties of China and there was not enough information for the far west of China, therefore the models established here are only suitable for the rural counties of China in longitude east of 100°. For simplicity, we used the longitude and latitude scale directly as the distance for Kriging without converting it to the actual distance.

We used S-plus in our study for computing the variogram, fitting the variogram models, Kriging and doing the contour maps. Links between S-plus and other spatial analysis software were reviewed in Gatrell and Bailey (1996).

In the Kriging, we did not make distributional assumptions on the mortality rates other than the stationarity of the processes. Therefore in using the model for prediction we decided not to do detailed statistical inferences on the predicted values from the model.

References

Armstrong, B. (1980). The epidemiology of cancer in the People's Republic of China. *International Journal of Epidemiology* **9**, 305-315.

Besag, J. and Newell, J. (1991). The detection of clusters in rare disease. *Journal of* *the Royal Statistical Society, Series A* **154**, 143-155.

Best, N. and Wakefield, J. (1999). Accounting for inaccuracies in population counts and case registration in cancer mapping studies. *Journal of the Royal Statistical Society, Series A* **162**, 363-382.

Chen, J. S., Campbell, T. C., Li J. Y. and Peto R. (1990). *Diet, Life-style and Mortality in China*. Oxford University Press.

Cuzick, J. and Edwards, R. (1990). Spatial clustering for inhomogeneous population. *Journal of the Royal Statistical Society, Series B* **52**, 73-104.

Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes: Models and Application*. Pion.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley.

Defense Mapping Agency, United States (1979). *Gazetter of the People's Republic of China*. Washington DC.

Gatrell, A. C. and Bailey T. C. (1996). Interactive spatial data analysis in medical geography. *Soc. Sci. Med.* **42**, 843-855.

Gesler, W. (1986). The uses of spatial analysis in medical geography: A review. *Soc. Sci. Med.* **23**, 963-973.

Hsing, A. W., Guo, W. D., Chen, J. S., Li, J. Y., Stone, B. J., Bolt, W. J. and Fraumeni, Jr. J. F. (1991). Correlates of liver cancer mortality in China. *International Journal of Epidemiology* **20**, 54-59.

Insightful (2001). *S-plus 2000*.

Lai, D. (1997). Spatial statistical analysis of Chinese cancer mortality: A comparison study of the D statistic. *Scandinavian Journal of Social Medicine* **25**, 258-265.

Lai, D. and Hardy, R. J. (1999). Potential gains in life expectancy or years of potential life lost: Impact of competing risks of death. *International Journal of Epidemiology* **28**, 894-898.

Lai, D., Dharma, R., Wolinsky, J. S. and Narayana, P. A. (2003). A comparative study of correlation coefficients in spatially MRSI-observed neurochemicals from multiple sclerosis patients. *Journal of Applied Statistics* **30**, 1221-1229.

Lam, N. S. N. (1986). Geographical patterns of cancer mortality in China. *Social Science and Medicine* **23**, 241-247.

Li, J. Y., Liu, B. Q., Li, G. Y., Chen, Z. J., Sun, X. D. and Rong, S. D. (1981). Atlas of cancer mortality in the People's Republic of China: An aid for cancer control and research. *International Journal of Epidemiology* **10**, 127-133.

Marshall, R. J. (1991). A review of methods for the statistical analysis of spatial patterns of disease. *Journal of the Royal Statistical Society, Series A* **154**, 421-441.

Ministry of Public Health and Chinese Academy of Preventive Medicine (MOPH and CAPM) (1992). *Annual Report on Chinese Diseases Surveillance*. Hauxia Press.

Oliver, M. A., Muir, K. R., Webster, R., Parkes, S. E., Cameron, A. H., Stevens, M. C. G. and Mann, J. R. (1992). A geostatistical approach to the analysis of pattern in rare disease. *Journal of Public Health Medicine* **14**, 280-289.

Pickle, L. W., Mungiole, M., Jones G. K. and White, A. A. (1999). Exploring spatial patterns of mortality: The new atlas of United States mortality. *Statistics in Medicine* **18**, 3211-3220.

Snow, J. (1985). *On the Mode of Communication of Cholera*, 2nd ed. Churchill.

State Statistical Bureau of China (SSB) (1995). *Monograph on the 1990 Population Census of the People's Republic of China*. Statistical Publishing House, Beijing.

State Statistical Bureau of China (SSB) (1985). *Tabulation on the 1982 Population Census of The People's Republic of China*. Statistical Publishing House, Beijing.

State Statistical Bureau of China (SSB) (1993). *Tabulation on the 1990 Population Census of The People's Republic of China*. Statistical Publishing House, Beijing.

State Statistical Bureau of China (SSB) (2002). *Tabulation on the 2000 Population Census of The People's Republic of China*. Statistical Publishing House, Beijing.

Walter S. D. (1992a). The analysis of regional pattern in health data: I. Distributional considerations. *American Journal of Epidemiology* **136**, 730-741.

Walter S. D. (1992b). The analysis of regional pattern in health data: II. The power to detect environmental effects. *American Journal of Epidemiology* **136**, 742-959.

Walter S. D. (1994). A simple test for spatial pattern in regional health data. *Statistics in Medicine* **13**, 1037-1044.

Received Octomber 19, 2002; accepted April 1, 2003.

Dejian Lai
School of Public Health
University of Texas Health Science Center
Houston, TX 77030, USA
dlai@sph.uth.tmc.edu

also:
Faculty of Statistics
Jiangxi University of Finance and Economics
Nanchang, P. R. of China