HÖGSKOLAN DALARNA

STATISTICS

One-year Master thesis

The importance of spatial aggregation and the road network on deciding optimal location of hospitals in Dalarna

Author:

Mengjie Han

Bo Zhu

Supervisor:

Kenneth Carling

Johan Håkansson

Examiner:

Lars Rönnegård

Högskolan Dalarna

09/06/ 2009

Contents

Abstract

In Dalarna province, two hospitals, Falu hospital (in Falun) and Mora hospital, can provide the 24 hours emergency care for citizens. Locations of them are fairly crucial such that each person living in Dalarna can reach the nearest hospital in the shortest time. Modifiable areal unit problem (MAUP), population information and road network are three main factors in evaluating the positions. ArcGIS provides geographical data regarding the units, population and road system. The empirical results help us require the knowledge if the construction of the third emergency hospital is feasible and where the best locations are.

Key words:

MAUP (Modifiable areal unit problem); ArcGIS; Optimal location; Population density;

Spatial aggregation; Road network; R

1 Introduction

1.1 General problem

Sweden, like any other country, has several levels of administrative, geographical areas. Commonly used levels are counties, municipalities and parishes. Another, frequently used, area division is SAMS of which there are about 9,000 in Sweden. For various analyses of the societal transports, SAMS is frequently applied.

Now, given the division, if you were the person who makes the decision, faced with the problem of locating the hospital, what would you choose? Usually the idea is based on the central point to choose the location, which is reasonable. In calculating such measures it is necessary to define the central point of each SAMS. The definition is a straightforward geometric calculation by taking the shape of the area into consideration. However, if you need to choose two hospitals in the same area, how to choose their location? If you need to choose three or even more hospitals, what will you do? Other than location, are there any other factors to consider? To find out the answers, we have embarked on a series of studies.

1.2 Purpose

As the applied problem, we are interested in the optimal location of the hospitals in Dalarna. This application gives rise to methodological problems, and the case serves to examine the magnitude of problem of the methodology. We want to examine whether the current locations minimize the sum of the distance from citizens' locations to the hospitals. When we calculate the points for some special purpose, the geometric method is not appropriate. One alternative solution is to re-define the shape, to use different scale aggregate data and to weight the distance using population density. Then if the current location of the hospitals is not good enough, we will find the optimal location by calculating with different number of replications are tested, such as using Population-weighted, Road network and the spatial aggregation. For further research, we should find which scale of aggregate data is enough to meet the requirement of location decision and we want to compare them with each other in this case.

1.3 Method

In this paper, we use low-level grid data instead of administrative division. In calculating the possible positions of hospitals, mass random points are generated. In order to specify the empirical distributions of the sum of distance from every point to the nearest hospital, different sample size are experimented, and the Kolmogorov–Smirnov test is employed to specify it. To measure the distance between the population and the hospitals, and to get the optimal location, we use a computer simulation to find the approximate road system, and then use population data from year 2002 at atomic squares.

1.4 Outline

In this article we first give a background of the hospitals' situation in Dalarna, and mention the related earlier definition to this kind of problem (MAUP). We then describe the data and the road system related with our study. In the following section, we present our findings in different ways. The paper ends with a discussion of the findings and possible ways to explain them.

2 Background

2.1 Background

In all the hospitals in Dalarna[1] province, there are two hospitals that can provide the 24 hours emergency care for citizens. One is Falu hospital (in Falun) and the other is Mora hospital.

Falun has 55 000 inhabitants, the emergency hospital was established in Falun near the Falu Coppermine. This hospital is the centre of specialist health care in Dalarna and has a large number of specialties.

Mora municipality has 20 000 inhabitants, Mora hospital is an emergency hospital for the population within the Siljan area, the north and west of Dalarna and it is also a part of the local health care system of the region mentioned above together with the primary health care units in the area. The hospital has highly specialized resources due to the fact that it has a large geographical serving area and it´s patients often have a long way to travel. The total number of inhabitants in its serving area is roughly 75.000.

If someone lives in Dalarna and has some emergency needs, she will call the nearest hospital for help, either the Falu hospital or the Mora hospital. Since the hospitals are founded by the province's tax payments, it basically requires that everyone living in this province should have access to a hospital. In doing that, the number of section and the costs are important factors to consider. In geographical field, this problem is associated with the modifiable areal unit problem (MAUP).

[1] The information about the hospital comes form: http://www.ltdalarna.se/

2.2 Review of MAUP

The study of physical geography naturally lends itself toward large spatial scale analyses. Over the past 15 years, the development of sophisticated Geography Information System (GIS) software has led to a quantitative study within physical geography conducted at large spatial scales such as the landscape and regional scales (Rosswell, 1991; Cain, 1997; Davis, 1998; McDermid, 2005). Of particular importance to the study of large scale phenomena is the modifiable areal unit problem (MAUP). In geography, we use modifiable areal units in quantitative analysis (Openshaw and Taylor, 1979). Openshaw (1984) pointed that "the areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregate." (Openshaw, 1984)

MAUP consists of two components: one is the scale problem, which is the variation in numeric results that occurs due to the numbers of zones used in an analysis; the other is the aggregation problem or zonation effect, which refers to which zoning scheme used at a level of spatial aggregation (or the variation in numeric results arising from the grouping of small areas into larger units) (Openshaw and Tylor, 1979). The first concern focuses on the issue of scale and variation. When areal units are aggregated into fewer and larger units for statistical analysis, values associated with the variation of the data decrease which will affect any associated statistical analysis. The second concern focuses on aggregation and the variation in results from statistical analysis as a result of alternative combinations of areal units at similar scales (Openshaw, 1984).

MAUP is a source of statistical bias that can radically affect the results of statistical hypothesis tests. Studies of the MAUP date back to the 1930s with the emphasis greatest in the late 1960s and 1970s. The results from studies of the MAUP have been highly variable and somewhat incomplete, thus making it difficult to make broad inferences about how the MAUP influences the performance of univariate, bivariate and multivariate statistics. However, some general patterns have arisen (Dark 2007). In univariate statistics, when the MAUP is present the mean does not change and the variance declines with increasing aggregation (Gehlke and Biehl, 1934; Openshaw, 1984). In the natural sciences, research has focused on the issue of scale and not aggregation. One of the major contributions in the field of natural sciences was to acknowledge the existence of natural scales at which ecological processes and physical characteristics occur within the landscape. Wiens (1989) and Levin both argue that a variety of statistical and mathematical tools can be used for scaling. However, they conclude that these techniques are appropriate only when applied for short-term or small-scale predictions. While studies relating to the issue of scale, there has been little concern of aggregation. As GIS data is being widely used, concerns about aggregation should become increasingly important.

The concept and the effect of MAUP have been documented in quantitative geography (Forthetingham and Wong, 1991). If areal units are imposed onto a discrete geographical distribution for the purpose of aggregation, the areal values will be on the locations of the boundaries. For example, this situation happens when administrative boundaries are frequently taken into account to form units of distribution of human population. Even the simplest real-world aggregation problem presents different possible methods.

MAUP is associated with ecological fallacy. Ecological fallacy occurs when it is inferred that results based on aggregate zonal date can be applied to the individuals within the zone itself (Shawna and Danielle 2007). Any statistics or models, which are based on aggregated spatial datasets, may be valid at current aggregated resolution, but any attempts to infer lower or higher resolution may be invalid. For example, the income of a certain area represents the average income. As regards to the same income level, we consider two extreme cases. One is that all the values are much higher values and much lower values. They cancel out each other to produce the average value. The other is that individual values are all distributed near the average values. It has the same result. Obviously, aggregate zone data is misleading.

However, sometimes, the individual data of an area is not easy to collect, or can be obtained only at very high cost. Therefore, it is necessary to use aggregate data in statistical model and inference.

3
Description of data

3.1 Data description

3.1.1 Population

The population data used in this paper is point data, which is published on Statistiska Centralbyrån [2] (Statistics Sweden). The subject is related with the population, which is sorted by municipality, age and gender. These 15729 records are taken from year 2002 and each one represents an atomic location (squares) of inhabitants of Dalarna.

The Statistics Sweden provides the original data of the population (see Appendix I: Table A1), which provides the population information of inhabitants lived in Dalarna. The population data records consist of the records for every 250m×250m square area with at least 1 person living in, on the year 2002. The population data records is saved in database file which are extracted from the ArcGIS map using ArcGIS software, in which each row is one record of square area in Dalarna. Every record contains 8 items which are square area number (RUT_ID), total number of the population (ANTAL_INV), number of the population in different ages (A0_15…A65_W), the x-axis Coordinate of the square area (POINT_X) and the y-axis Coordinate of the square area (POINT_Y). Looking through the original data, the population age information is not relevant for the problem of this essay, and therefore, we remove these columns.

[2] http://www.ssd.scb.se/databaser/makro/Visavar.asp?yp=tansss&xu=C9233001&huvudtabell=FolkmangdNov&deltabell=K1&deltabellnamn=Population+1+November+by+municipality%2C+age+and+sex%2E+Year&omradekod=BE&omradetext=Population&preskat=O&innehall=FolkmangdNov&starttid=2002&stopptid=2008&Prodid=BE0101&fromSok=&Fromwhere=S&lang=2&langdb=2

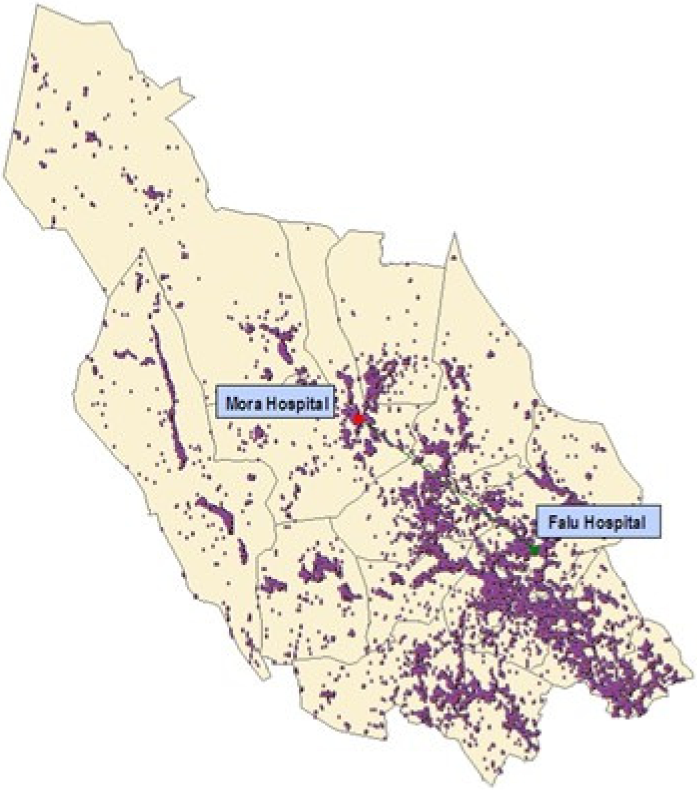
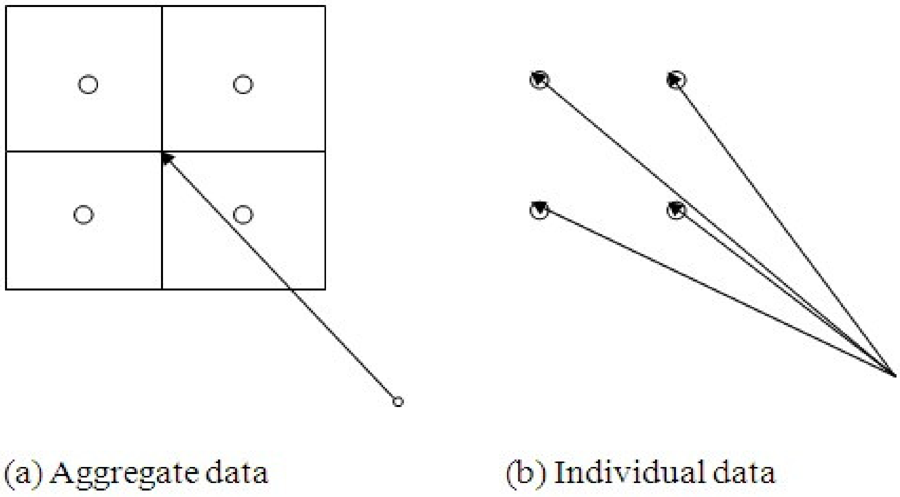


Figure 1 Dalarna Map

The population data records can be seen on the Map. Figure 1 shows the locations of all the 15729 atomic squares in Dalarna. Since we are concerned with the individual distance to the nearest hospital, we note that the “crow flight distance” (C-F distance) is the shortest line distance between the centers of the atomic squares and the hospital. Every point represents a square with the length of 250 meter. The coordinate of every point is expressed as x-y values. The point locates in the centre of the square. Locations of hospitals are also labeled on the map, which represent the actual positions.



(a) Aggregate data

(b) Individual data

Figure 2

Aggregate Data and Individual Data

Figure 2 shows by an example how the atomic squares are spatial aggregated into a higher level of aggregation. This kind of data is deemed as individual (the right part of the Figure2) data because it comes from atomic unit known as the square area. In addition, the aggregate data (the left part of the Figure2) denotes the sum of several squares. In that sense, the attribute value such as area and population are redefined with the central point changed.

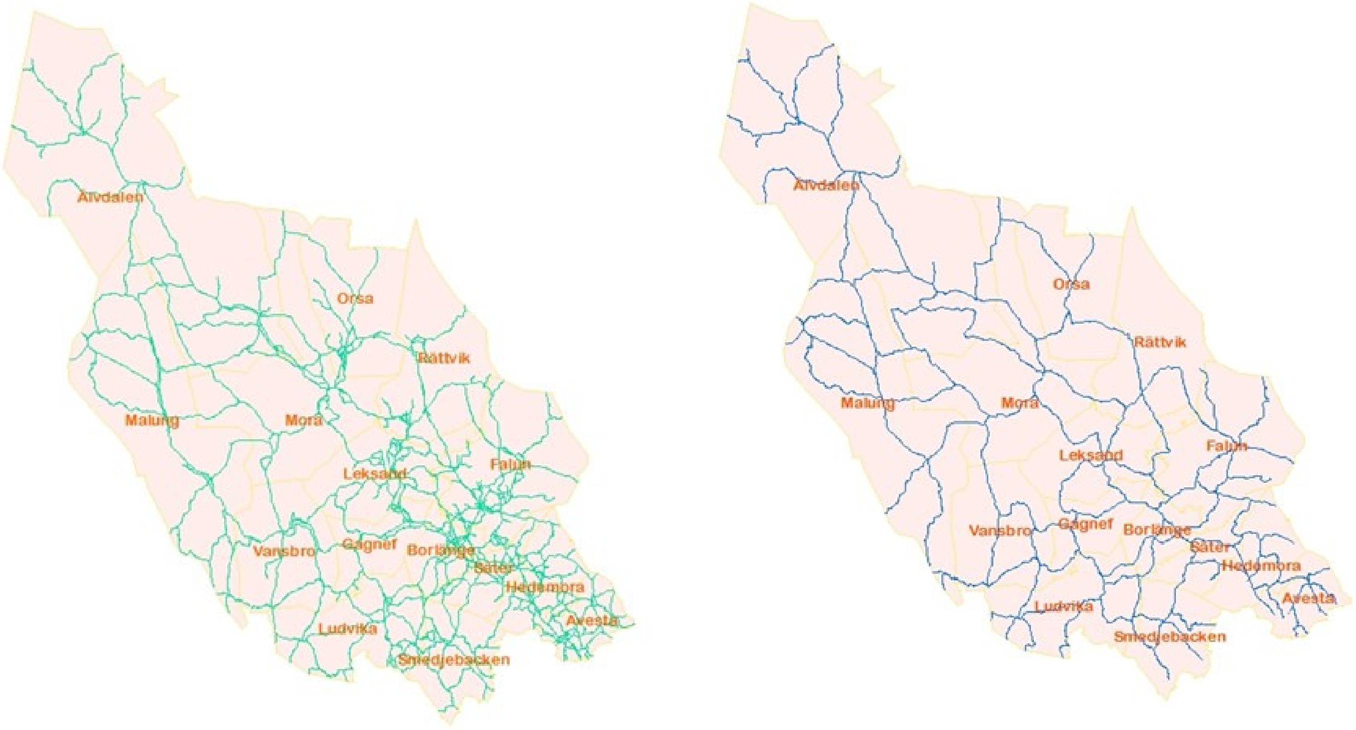
3.1.2 Road system distance

A more related discussion of the C-F distance is the distance between the atomic square and the hospital along the public roads, this distance is defined as road system distance (R-S distance). Swedish national road are roads with road numbers from 1 through 99 in Sweden. The national roads are usually of high quality and sometimes pass through several counties. Roads with lower numbers are in southern Sweden, and roads with higher numbers are in northern Sweden. There are many cases where two or more routes in this system share the same physical road for a considerable distance, giving the country several kilometers of double-numbered road.

The network of national roads covers all of Sweden, and has a total length of 8,769 km (not including E-roads. The figure is shorter than before, since road 45 is now E45.). The national roads are public roads owned by the Government of Sweden and administered by the Swedish Road Administration, which is a government agency in Sweden.[3] The road system used in this paper is Swedish national roads, which are also distributed in Dalarna (Figure 3).

The left part of the Figure 3 gives us an overview of all roads in Dalarna that is a quite complicated road system. To make the calculation possible, we create an approximate road system (right part of the Figure 3). Some roads in short length are removed leaving the main roads on the map. The road system can now be calculated based on the approximate road system.

[3] http://en.wikipedia.org/wiki/Swedish_national_road



Road

Approximate road

Figure 3 Road System

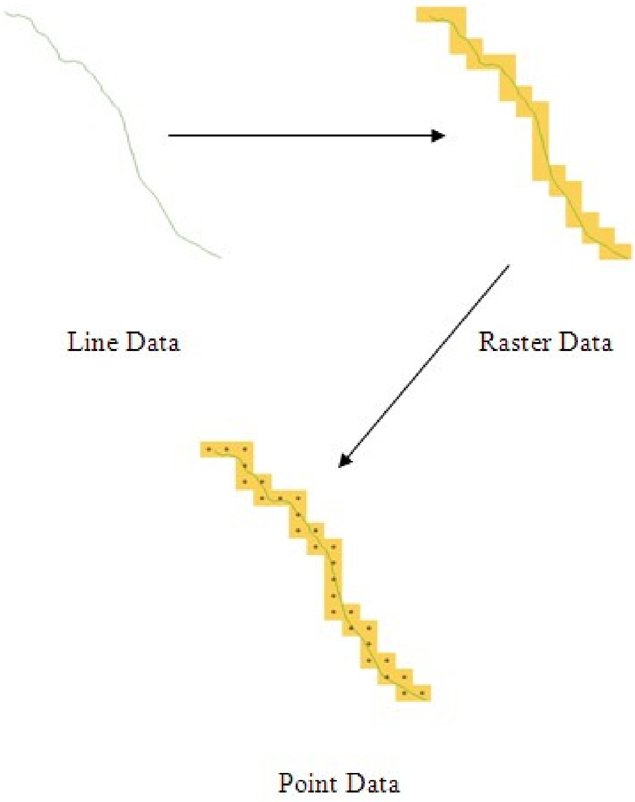


Figure 4 Data Transformation

The data formation in the map is feature data. Road is line data (Appendix I: Table A2) that contains endless points. In order to extract the coordinate information from the line data, we must transfer it into raster data (Figure 4). Every grid can be transferred back into point data (Appendix I: Table A3). In that sense, the coordinate can be extracted. We only consider the starting points and ending points and assume the roads between them are straight lines, which is the reason why we refer to the road system as approximate.

3.2 Pretreatment of coordinates

Integer and decimal fraction of the original data are in large numbers. We first neglect the decimal fraction since it is a very small part of the coordinate value. And then the integer fraction need paralleled move towards the origin point that is the minimal value of two element of the coordinate vector. Finally, the values are shrunk by 250 times in order to facilitate calculation such that the measurement unit is 250 meter.

4 Results

4.1 Optimal location of the hospitals

In the map of Dalarna province (Figure 1), we can see lots of squares (points in this picture). The bigger points are locations of the real hospitals which can provide the 24 hours emergency care for citizens known as Falu hospital and the Mora hospital. The coordinates for the hospitals are Falu: (723, 308) and Mora: (483, 492).

One principle we are following is to minimize the sum of total distance for the population. In that sense, three factors are experimented with. These factors are population weighted or unweighted, spatial aggregation at different levels, C-F distance or R-S distance. To examine and discuss the results of the factors, we compare the result on locating two hospitals. Furthermore one hospital and three hospitals are also considered to draw conclusions.

4.2 Minimum distance function

To explain procedures for calculating optimal locations, we need some notations and formulas first. (x_i, y_i) is the coordinate of atomic square i. N_i is the number of inhabitants in the atomic square i. (X_h, Y_h) is the coordinate of the h th hospital. i =1, 2, …, I (I = 15729 in the application) and h = 1, 2, …, H.

The population weighted summed C-F distance, S, in the case of H hospitals can be expressed as

$$S=\sum_{i=1}^{I}N_i\cdot\min[\sqrt{(x_i-X_1)^2+(y_i-Y_1)^2},\ ...,\ \sqrt{(x_i-X_H)^2+(y_i-Y_H)^2}]\qquad(1).$$

In the case of only hospital, equation (1) is simplified to a function of X_1 and Y_1, namely,

$$S(X_1,\ Y_1)=\sum_{i=1}^{I}N_i\cdot\sqrt{(x_i-X_1)^2+(y_i-Y_1)^2}\qquad(2).$$

By taking partial derivatives with respect to the two variables, the solution expressions can be derived as

$$\begin{cases} X_1=\dfrac{\sum_{i=1}^{I}N_ix_i\Big/\sqrt{(X_1-x_i)^2+(Y_2-y_i)^2}}{\sum_{i=1}^{I}N_i\Big/\sqrt{(X_1-x_i)^2+(Y_2-y_i)^2}} \\ Y_1=\dfrac{\sum_{i=1}^{I}N_iy_i\Big/\sqrt{(X_1-x_i)^2+(Y_2-y_i)^2}}{\sum_{i=1}^{I}N_i\Big/\sqrt{(X_1-x_i)^2+(Y_2-y_i)^2}} \end{cases}\qquad(3).$$

The above expressions are not the final results. Although the mathematical solutions exist theoretically, it will be very complicated. Furthermore, when the number of hospitals is equal or greater than 2, no mathematical solutions actually exist to minimize S, because the positions of the nearest hospitals for each atomic square are unknown parameters. Hence, it seems difficult to find an analytical solution to this problem at finding an optimal location. Instead we consider a simulation approach.

4.3 Statistical simulation of location

4.3.1 Location of hospitals

Instead of using mathematical methodology, we generate uniform random numbers on the map for H hospitals. These points represent the possible location of hospitals. Since what we are interested in is to compare computing results and the current hospitals, two pairs of random numbers are generated. We repeat the procedure R times. The optimal location is then said to be the smallest of all $S^{(R)}(X_1,\ Y_1,\ X_2,\ Y_2)$, being the min($S(X_1,\ Y_1,\ X_2,\ Y_2)$), based on R times replications. To find the optimum under the restriction that the citizens must follow the road system to the hospital, we proceed in an identical fashion, only replacing the C-F distance with the R-S distance.

4.3.2 Uniformly points for population unweighted

The uniformly distributed points x_i *and* y_i are located within the Dalarna border known as the unweighted method. Figure 5 shows the procedure. The first step is to generate dots of the whole range of x and y, because the area is an irregular polygon. And the second step is to eliminate dots that lie outside the region.

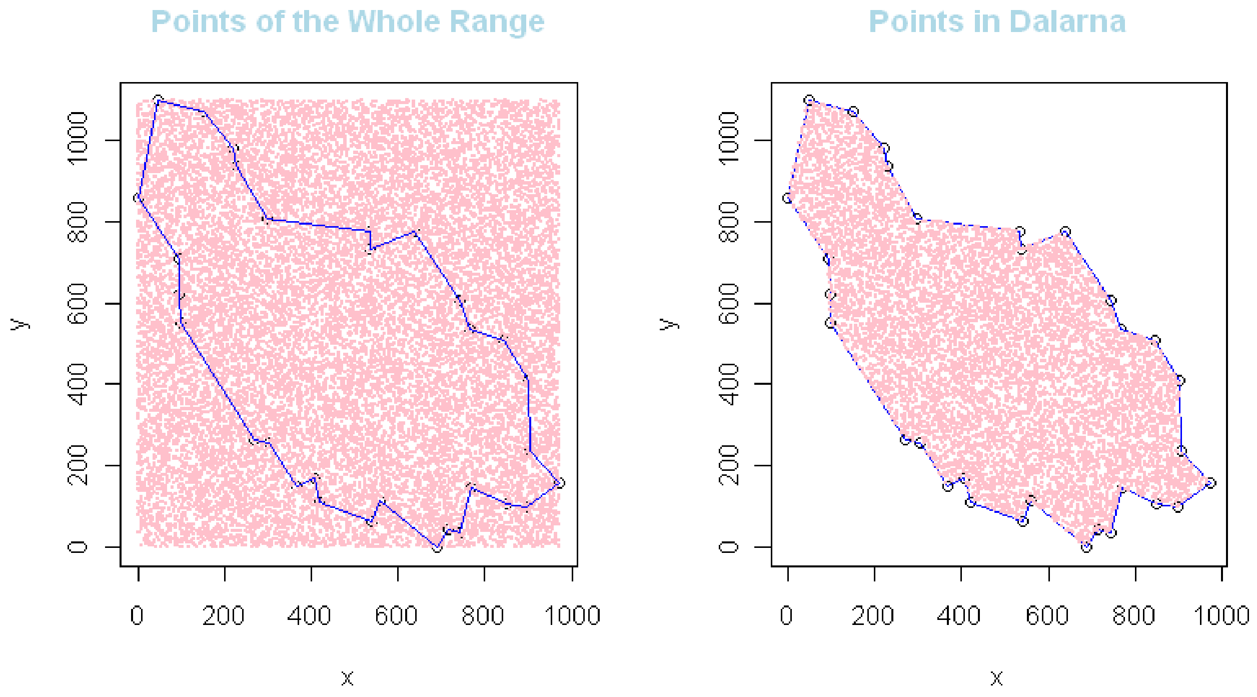


Figure 5 Random Points

Based on the uniform points, the empirical distance min ($S(X_1, Y_1, X_2, Y_2)$) can be calculated. Here we compare four different numbers of R: 500, 1000, 2000 and 5000.

4.3.3 Kolmogorov–Smirnov test and distribution specification

Above we have outlined the procedure for finding the optimum making use of simulations of locations R times in remains to discuss the volume of R. Figure 6 gives the empirical distribution histograms of all of S in equation (1) for different Rs. Four theoretical known distribution curves are added. The dashed line is normal distribution with mean of the mean of $S(X_1, Y_1, X_2, Y_2)$ while the dotted curve is also normal with mean of median of $S(X_1, Y_1, X_2, Y_2)$. The dot dash curve is gamma and the solid is lognormal. The graphs suggest that these four theoretical distributions provide different locations and scales.

In order to specify the distribution of S, Kolmogorov–Smirnov (K–S test) test is available. It uses goodness of fit statistics in minimum distance estimation. It also uses the supremum of the absolute difference between the empirical and the estimated distribution functions.

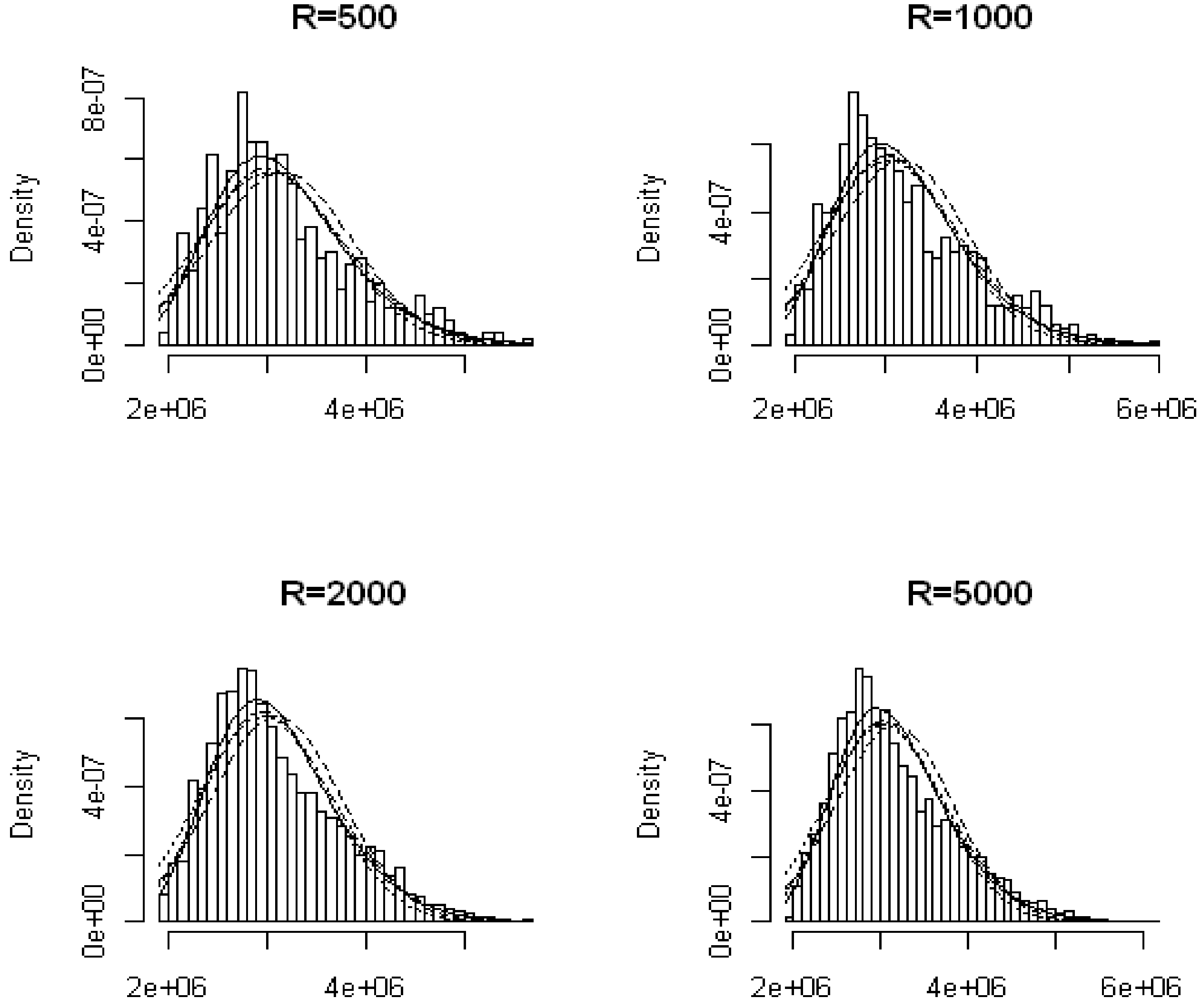


Figure 6 Theoretical fitting lines under different sample size

The Kolmogorov–Smirnov test (Kotz, 2006) is a form of minimum distance estimation used as a nonparametric test to compare a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample K–S test). The K–S statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. The null distribution of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution (in the two-sample case) or that the sample is drawn from the reference distribution (in the one-sample case). In each case, the distributions considered under the null hypothesis are continuous distributions. The K–S test can be modified to serve goodness of fit test.

The K–S is given by Kolmogorov distribution. The cumulative distribution function is

$$\Pr(K \le x) = \frac{\sqrt{2\pi}}{x}\sum_{i=1}^{\infty}\exp(\frac{-(2i-1)^2\pi^2}{8x^2}) \tag{4.}$$

The K–S statistic is $D_n = \sup_x | F_n(x) - F(x) |$, where $F_n(x)$ is the empirical distribution function and $F(x)$ is the reference distribution function. Therefore the null hypothesis is rejected at level α if $\sqrt{n}D_n > K_\alpha$ where K_α is found from $\Pr(K < K_\alpha) = 1 - \alpha$.

In this case the lognormal distribution produces the largest p-value (equal to 0.257) based on R=5000 that means we cannot reject $S(X_1, Y_1, X_2, Y_2)$ comes from lognormal distribution.

What we are interested in is the minimum value of $S(X_1, Y_1, X_2, Y_2)$. The lognormal distribution suggests that probability that we get smaller value than min($S(X_1, Y_1, X_2, Y_2)$) is 0.011, if we let R=5000 which is a quite small probability such that we can believe that the minimum sum is difficult to decrease.

4.4 Result

The experimented results are listed for the case of two hospitals. Different results under different cases are listed in the Table 1. If we consider the current position of hospitals as the compared coordinates known as (723,308) and (483,492), the results show that the smallest square and taking population into account produce minimum error. C-F distance is used to finish these results. In consideration of the real problem, road-based analysis is reasonable to be added.

Table 1 Optimal Location of Different Scales and Different Methods

Square	Unweighted		Population weighted	
	South	North	South	North
250*250 m^2	(657, 333)	(316, 669)	(692, 261)	(473, 536)
1000*1000 m^2	(665, 390)	(232, 591)	(733, 235)	(508, 447)
5000*5000 m^2	(603, 400)	(267, 622)	(700, 271)	(494, 495)
10000*10000 m^2	(680, 383)	(305, 583)	(664, 248)	(428, 425)
50000*50000 m^2	(738, 319)	(337, 561)	(696, 253)	(448, 468)

If the road network is taken into account, a more realistic result is provided. But some assumptions must be put up with. First, the main roads are continuous. That means if two points locate on the road, only one way is supposed to be found to connect these points. Second, based on the coordinates, one road is likely to be divided into several segments. Every segment is supposed to be straight line. Third, if points do not exactly locate on the road, points can reach the nearest road. The distance is component of the total distance. Once points are located on the road, distance can be measured through the main road network. As depicted in Figure 3, the work can be done in the following method. Every main road contains two nodes, starting node and ending node. They represent the start points or destinations for persons and hospitals respectively. Once the positions of squares and hospitals are specified, their nearest node points are specified. The distance can be calculated. Results for one hospital and three hospitals are calculated in the same way. Since the 250 square meters scale is more believable, we only list this scale (Others are listed in the appendix II Table A4-A9) in terms of graphs (Figure 7).

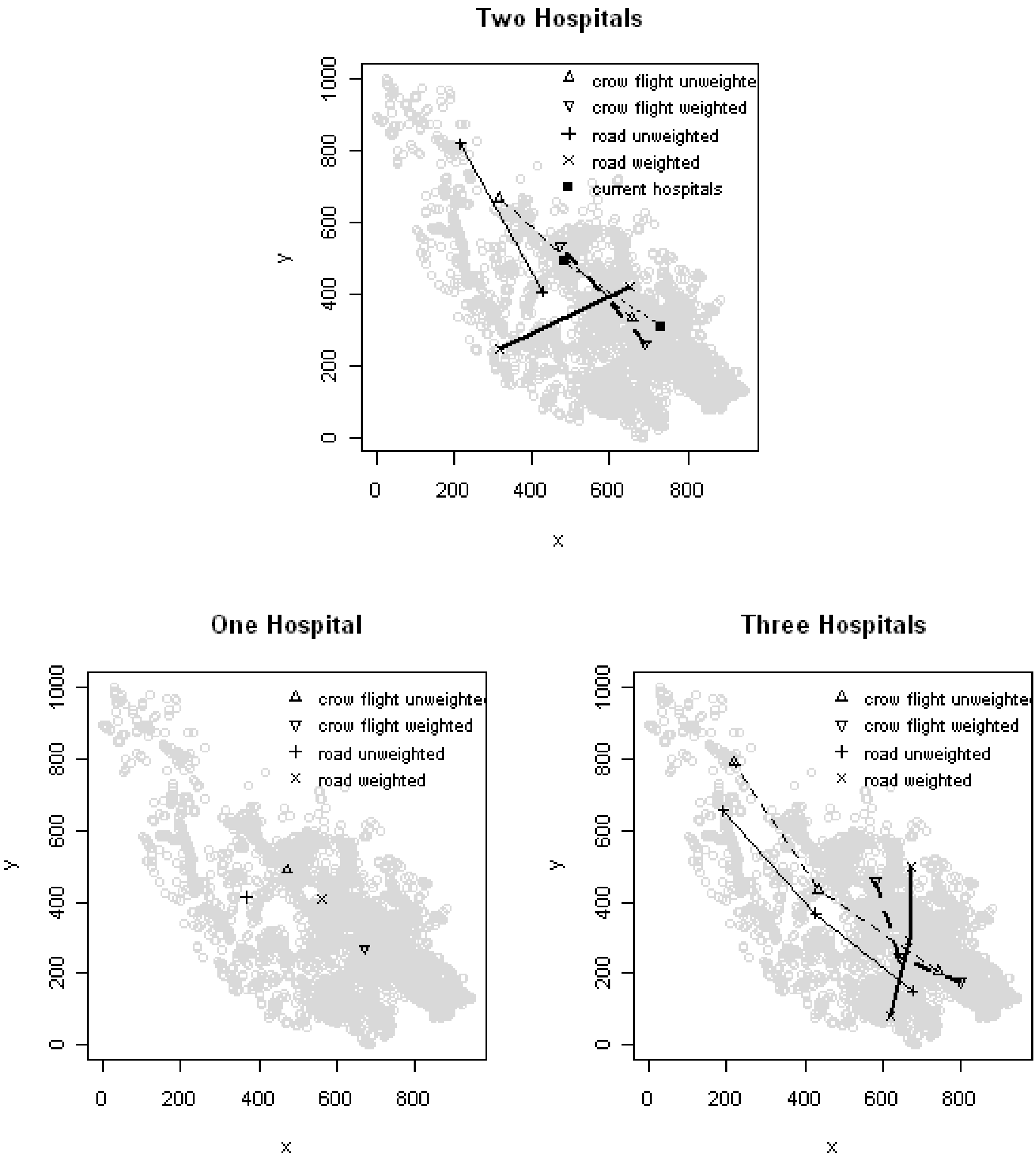


Figure 7 Optimal hospitals for 250 square meters

5
Discussion

Going back to our aims, what we are interested in is how the optimal location changes by considering different factors. The results provided by Figure 7 and Table A4-A9 help us to derive the following findings.

The most important finding is that the location of current two hospitals are most close to the result given by crow flight and weighted by population. That, to some extent, illustrates that the MAUP problem exists in this problem. And if the aggregate data is used, we are more likely to underestimate the average distance. Therefore, considering the current conclusion, 250 square meters is appropriate atomic unit. Moreover, population distribution really plays an important role in deciding the positions. In regard to the number of hospitals, the sum of distance decreases not so much if we only consider crow flight distance. But the road network provides a sharp decrease, which means if the reduced amount of the sum of distance is crucial, government may consider building the third emergency hospital.

From Figure 7, it is no surprise that the high population density pulls the weighted results to the south and east. The road network spreads out toward west-east direction compared to the crow flight distance in spite of the seemingly symmetric road system. Therefore, the construction of the third hospital must take the road system into account based on the weighted atomic units.

Reference

Cain, D. H., Riitters, K. and Orvis, K. 1997: A multi-scale analysis of landscape statistics. *Landscape Ecology* 12, 199-212.

Dark, S. J. and Bram, D. 2007: The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography* 31 (5) pp. 471-479.

Davis, F. W., Stoms, D. M. and Hollander, A. D. 1998: *The California Gap Analysis Project-Final Report*. Santa Barbara, CA: University of California Press.

Fotheringham, A.S. and Wong, D.W.S. 1991: The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23: 1025-44.

Gehlke, C. E. and Biehl, H. 1934: Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association Supplement* 29, 169-70.

Kotz, S. 2006: Encyclopedia of Statistical Science (Second Edition), Volume 6, 3875-3878. ISBN 0-471-743-76-3 (v.6)

Levin, S.A. 1993: Concepts of scale at the local level. In Ehleringer, J.R and Field, C.B., editors, *Scaling Physiological Processes: Leaf to Globe*, San Francisco, CA: Academic Press, 7-19.

McDermid, G. J., Franklin, S.E. and LeDrew, E. F. 2005: Remote sensing for large-area habitat mapping. *Progress in Physical Geography* 29, 449-74.

Openshaw, S. 1984: The Modifiable Areal Unit Problem. Norwich: Geo Books. ISBN 0-86094-134-5.

Openshaw, S. and Tylor, P.J. 1979: A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In Wriglry, N., editor, *Statistical applications in spatial science*, London: Pion, 127-44.

Ormsby, T., Napoleon, E. and Bruke, R 2001: Getting to know ArcGIS desktop: basic of ArcView, ArcEditor, and ArcInfo. ISBN 1-879102-89-7.

Paze, A. 2004 Anisotropic variance functions in geographically weighted regression models. *Geographical analysis* 36 (4).

Rosswell, T., Woodmansee, R.G. and Matson, P.A. 1991: What does remote sensing do for ecology? *Ecology* 72, 45-54.

Tagashira, N. and Okabe, A. 2002: The Modifiable Areal Problem in a Regression Model Whose Independent Variable Is a Distance from a Predetermined Point. *Geographcal analysis* 34 (1).

Wiens, J.A. 1989 : Spatial scaling in ecology. *Functional Ecology* 3, 385-86.

Appendix

Appendix I: Format of Original data

Table A1: the population data showing square area number, total number of the population, number of the population in different ages, and the X, Y-axis Coordinate of the square area

RUT_ID	ANTAL_INV	A0_15	A16_24	A25_64	A65_W	POINT_X	POINT_Y
13102506866250	3	0	0	0	3	1310375,0035	6866374,9998
13105006866250	8	3	3	3	3	1310625,0006	6866374,9998
13107506866250	3	0	0	0	3	1310874,9977	6866374,9998
13110006866000	6	3	0	3	3	1311124,9989	6866124,9984
13115006864250	3	0	0	3	0	1311625,0014	6864374,9987
13117506864250	3	0	0	3	0	1311875,0026	6864374,9987

Table A2: the road data the population data showing the information of start points and end points, number of roads and the addition information of all records.

FNODE	TNODE	LPOLY	RPOLY	LENGTH	AV01_W	KKOD	KATEGORI
60	57	8	8	824,976	1	5231	Allmõn võg >7m, võgnummer >500
83	86	8	8	4378,38	2	5331	Allmõn võg 5-7m, võgnummer >500
85	86	8	8	10,125	3	5321	Allmõn võg 5-7m, võgnummer 100-499
73	88	8	8	5715,740	4	5421	Allmõn võg <5m, võgnummer 100-499
88	97	8	8	5155,570	5	5321	Allmõn võg 5-7m, võgnummer 100-499
104	97	8	8	3316,070	6	5421	Allmõn võg <5m, võgnummer 100-499

Table A3 the raster data (point) showing point number, number of roads, and the X,Y-axis Coordinate of the square area

POINTID	GRID_CODE	POINT_X	POINT_Y
1	93,000	1349726,9680	6894988,9395
2	93,000	1346846,3185	6894028,7230
3	93,000	1347806,5350	6894028,7230
4	93,000	1348766,7515	6894028,7230
5	93,000	1349726,9680	6894028,7230
6	60,000	1317079,6070	6893068,5065

Appendix II: Collection of results tables

Table A4: Optimal location (by coordinate) of two hospitals

		C-F distance		R-S distance	
	Spatial aggregation square (meter)	hospital 1	hospital 2	hospital 1	hospital 2
Unweighted	250	(657, 333)	(316, 669)	(427, 407)	(217, 821)
	1000	(665, 390)	(232, 591)	(481, 408)	(152, 800)
	5000	(603, 400)	(267, 622)	(594, 484)	(183, 779)
	10000	(680, 383)	(305, 583)	(508, 301)	(113, 826)
	50000	(738, 319)	(337, 561)	(641, 473)	(175, 759)
Weighted	250	(692, 261)	(473, 536)	(318, 248)	(651, 422)
	1000	(733, 235)	(508, 447)	(708, 149)	(399, 411)
	5000	(700, 271)	(494, 495)	(656, 181)	(449, 395)
	10000	(664, 248)	(428, 425)	(746, 166)	(444, 379)
	50000	(696, 253)	(448, 468)	(782, 210)	(436, 396)

Table A5: Coordinate results for three hospitals

		C-F distance			R-S distance		
	Spatial aggregation square (meter)	hospital 1	hospital 2	hospital 3	hospital1	hospital 2	hospital 3
Unweighted	250	(743, 206)	(435, 433)	(218, 789)	(676, 151)	(427, 367)	(190, 655)
	1000	(668, 323)	(483, 416)	(110, 781)	(440, 410)	(599, 538)	(177, 751)
	5000	(781, 406)	(423, 482)	(203, 769)	(394, 19)	(658, 362)	(145, 890)
	10000	(668, 367)	(494, 387)	(277, 685)	(657, 410)	(384, 493)	(151, 671)
	50000	(810, 243)	(419, 482)	(160, 781)	(604, 410)	(421, 417)	(171, 773)
Weighted	250	(799, 177)	(643, 246)	(581, 456)	(621, 82)	(666, 290)	(673, 498)
	1000	(838, 136)	(673, 258)	(484, 564)	(569, 158)	(670, 222)	(506, 360)
	5000	(847, 193)	(660, 277)	(414, 488)	(619, 156)	(655, 167)	(373, 484)
	10000	(723, 264)	(642, 268)	(455, 370)	(697, 244)	(511, 286)	(493, 340)
	50000	(698, 269)	(836, 275)	(448, 412)	(660, 272)	(573, 314)	(456, 437)

Table A6: Sums of distance and calculating time for two hospitals

		C-F distance		**R-S distance**	
	Spatial aggregation square (meter)	sum of distance	calculating time (min)	sum of distance	calculating time (min)
Unweighted	250	1963107	23	4289342	17
	1000	1450736	3	2407840	4
	5000	1473600	less than 1	2183200	1
	10000	1283200	less than 1	2081600	1
	50000	1361896	less than 1	2130349	less than 1
Weighted	250	32589808	41	107828989	31
	1000	29300784	7	57260608	3
	5000	24971200	less than 1	48298800	1
	10000	24416000	less than 1	46316800	less than 1
	50000	23940608	less than 1	49811008	less than 1

Table A7: Sums of distance and calculating time for three hospitals

		C-F distance		**R-S distance**	
	Spatial aggregation square (meter)	sum of distance	calculating time (min)	sum of distance	calculating time (min)
Unweighted	250	1647088	46	3007346	27
	1000	1100752	7	1842064	5
	5000	1157600	less than 1	1733600	1
	10000	984000	less than 1	1635200	1
	50000	925962	less than 1	1711916	less than 1
Weighted	250	30132756	77	42996260	52
	1000	25668432	12	37425264	7
	5000	20142400	less than 1	24431600	1
	10000	19609600	less than 1	20096000	less than 1
	50000	18548028	less than 1	18689616	less than 1

Table A8: Coordinate results for two hospitals

		C-F distance	R-S distance
	Spatial aggregation square (meter)	hospital 1	hospital 1
Unweighted	250	(473, 488)	(368, 414)
	1000	(484, 462)	(421, 399)
	5000	(471, 462)	(492, 390)
	10000	(527, 429)	(410, 411)
	50000	(589, 419)	(584, 469)
Weighted	250	(671, 269)	(562, 408)
	1000	(673, 272)	(457, 375)
	5000	(675, 267)	(771, 209)
	10000	(665, 261)	(722, 207)
	50000	(688, 258)	(545, 443)

Table A9: Sums of distance and calculating time for three hospitals

		C-F distance		R-S distance	
	Spatial aggregation square (meter)	sum of distance	calculating time (min)	sum of distance	calculating time (min)
Unweighted	250	2903114	12	6379128	9
	1000	2351248	3	3629040	3
	5000	2360000	less than 1	3435600	1
	10000	2283200	less than 1	2918400	less than 1
	50000	2458095	less than 1	3189955	less than 1
Weighted	250	44767773	22	167723961	17
	1000	40656432	6	102175664	5
	5000	36136000	less than 1	102798400	1
	10000	36203200	less than 1	99427200	less than 1
	50000	36010548	less than 1	98344228	less than 1

Appendix III: Main procedures in R

```
##########################################################################################################################
## Path Function of Any Two Dots                                                                                          ##
## nod.co is a matrix which provides the information about which two nodes are connected. The row numbers are the nodes numbers and the elements are ##
## corresponding connected noeds nodes to its row number.                                                                  ##
##########################################################################################################################
road.connect<-function(x,y){
   num1<-num2<-num3<-num4<-num5<-num6<-num7<-num8<-num9<-num10<-num11<-1
   while(x!=y){
   k1<-k2<-k3<-k4<-k5<-k6<-k7<-k8<-k9<-k10<-T
   c1<-nod.co[x,num1];while(c1>=1 & c1!=x){if (c1==y){nod<-c(x,y);y<-x}
   else{
     while(x!=y & k1==T){
     c2<-nod.co[c1,num2];while(c2>=1 & c2!=x){if (c2==y){nod<-c(x,c1,y);y<-x}
     else{
       while(x!=y & k2==T){
       c3<-nod.co[c2,num3];while(c3>=1 & c3!=c1){if (c3==y){nod<-c(x,c1,c2,y);y<-x}
       else{
         while(x!=y & k3==T){
         c4<-nod.co[c3,num4];while(c4>=1 & c4!=c2){if (c4==y){nod<-c(x,c1,c2,c3,y);y<-x}
         else{
           while(x!=y & k4==T){
           c5<-nod.co[c4,num5];while(c5>=1 & c5!=c3){if (c5==y){nod<-c(x,c1,c2,c3,c4,y);y<-x}
           else{
             while(x!=y & k5==T){
             c6<-nod.co[c5,num6];while(c6>=1 & c6!=c4){if (c6==y){nod<-c(x,c1,c2,c3,c4,c5,y);y<-x}
             else{
               while(x!=y & k6==T){
               c7<-nod.co[c6,num7];while(c7>=1 & c7!=c5){if (c7==y){nod<-c(x,c1,c2,c3,c4,c5,c6,y);y<-x}
               else{
                 while(x!=y & k7==T){
                 c8<-nod.co[c7,num8];while(c8>=1 & c8!=c6){if (c8==y){nod<-c(x,c1,c2,c3,c4,c5,c6,c7,y);y<-x}
                 else{
                   while(x!=y & k8==T){
                   c9<-nod.co[c8,num9];while(c9>=1 & c9!=c7){if (c9==y){nod<-c(x,c1,c2,c3,c4,c5,c6,c7,c8,y);y<-x}
                   else{
                     while(x!=y & k9==T){
                     c10<-nod.co[c9,num10];while(c10>=1 & c10!=c8){if (c10==y){nod<-c(x,c1,c2,c3,c4,c5,c6,c7,c8,c9,y);y<-x}
                     else{
                       while(x!=y & k10==T){
                       c11<-nod.co[c10,num11];while(c11>=1 & c11!=c9){if (c11==y){nod<-c(x,c1,c2,c3,c4,c5,c6,c7,c8,c9,c10,y);y<-x}
                       c11<-0}
                       num11<-num11+1;k1<-k2<-k3<-k4<-k5<-k6<-k7<-k8<-k9<-k10<-T
```

```
                      if (num11>7){k10<-F;num11<-1}}
                    }
                    c10<-0}
                    num10<-num10+1;k1<-k2<-k3<-k4<-k5<-k6<-k7<-k8<-k9<-k10<-T
                    if (num10>7){k9<-F;num10<-1}}
                  }
                  c9<-0}
                  num9<-num9+1;k1<-k2<-k3<-k4<-k5<-k6<-k7<-k8<-k9<-k10<-T
                  if (num9>7){k8<-F;num9<-1}}
                }
                c8<-0}
                num8<-num8+1;k1<-k2<-k3<-k4<-k5<-k6<-k7<-k8<-k9<-k10<-T
                if (num8>7){k7<-F;num8<-1}}
              }
              c7<-0}
              num7<-num7+1;k1<-k2<-k3<-k4<-k5<-k6<-k7<-k8<-k9<-k10<-T
              if (num7>7){k6<-F;num7<-1}}
            }
            c6<-0}
            num6<-num6+1;k1<-k2<-k3<-k4<-k5<-k6<-k7<-k8<-k9<-k10<-T
            if (num6>7){k5<-F;num6<-1}}
          }
          c5<-0}
          num5<-num5+1;k1<-k2<-k3<-k4<-k5<-k6<-k7<-k8<-k9<-k10<-T
          if (num5>7){k4<-F;num5<-1}}
        }
        c4<-0}
        num4<-num4+1;k1<-k2<-k3<-k4<-k5<-k6<-k7<-k8<-k9<-k10<-T
        if (num4>7){k3<-F;num4<-1}}
      }
      c3<-0}
      num3<-num3+1;k1<-k2<-k3<-k4<-k5<-k6<-k7<-k8<-k9<-k10<-T
      if (num3>7){k2<-F;num3<-1}}
    }
    c2<-0}
    num2<-num2+1;k1<-k2<-k3<-k4<-k5<-k6<-k7<-k8<-k9<-k10<-T
    if (num2>7){k1<-F;num2<-1}}
  }
  c1<-0}
  num1<-num1+1
  }
return(nod)
}
```

```
##################
## Distance Function##
###################
distance<-function(x1,y1,x2,y2){
dis<-sqrt((x1-x2)^2+(y1-y2)^2)
return(dis)
}
#############################################
## distance between any two nodes                        ##
## (X, Y) corresponds to the coordinate of each node     ##
##############################################
dis.road<-matrix(0,48,48)
for (i in 1:47){
   for (j in (i+1):48){
   a<-road.connect(i,j)
      road.sto<-numeric(0)
      for (k in 1:(length(a)-1)){
      road.sto[k]<-distance(X[a[k]],Y[a[k]],X[a[k+1]],Y[a[k+1]])
      }
      dis.road[i,j]<-dis.road[j,i]<-sum(road.sto)
   }
}
###################################################
##Only weighted RS-distance computing is listed for atomic unit.##
##x and y are the coordinate of the centre of the square.          ##
##X and Y are the coordinate of all of the nodes.                  ##
##pop is the number of the corresponding squares.                  ##
###################################################
no250<-nohos1250<-nohos2250<-NULL;dis250<-numeric(0)
for (i in 1:length(x)){
   abc<-numeric(0)
   for (j in 1:48){
   abc[j]<-distance(x[i],y[i],X[j],Y[j])
   }
   no.least<-order(abc)[1]
   no250<-c(no250,no.least)
   dis250[i]<-distance(x[i],y[i],X[no250[i]],Y[no250[i]])
}

x01<-runif(5000,100,max(x))
y01<-runif(5000,0,900)
x02<-runif(5000,100,max(x))
y02<-runif(5000,0,900)
for (i in 1:5000){
  abchos1<-abchos2<-numeric(0)
  for (j in 1:48){
  abchos1[j]<-distance(x01[i],y01[i],X[j],Y[j])
  abchos2[j]<-distance(x02[i],y02[i],X[j],Y[j])
  }
  no.least.hos1<-order(abchos1)[1]
  no.least.hos2<-order(abchos2)[1]
  nohos1250<-c(nohos1250,no.least.hos1);nohos2250<-c(nohos2250,no.least.hos2)
}
d1<-as.numeric(0)
d2<-as.numeric(0)
c<-0
for (k in 1:length(x01)){
   d1[k]<-0
   d2[k]<-0
   for (l in 1:length(x)){
      if (dis.road[no250[l],nohos1250[k]]<dis.road[no250[l],nohos1250[k]]){
      dis1<-pop[l]*dis.road[no250[l],nohos1250[k]]
      d1[k]<-d1[k]+dis1
      }
      else{
      dis2<-pop[l]*dis.road[no250[l],nohos2250[k]]
      d2[k]<-d2[k]+dis2
      }
   }
}
n<-order(d1+d2)[1];abcdhos1<-abcdhos2<-numeric(0)
for (j in 1:48){
  abcdhos1[j]<-distance(x01[n],y01[n],X[j],Y[j])
  abcdhos2[j]<-distance(x02[n],y02[n],X[j],Y[j])
  }
  ju1<-order(abcdhos1)[1]
  ju2<-order(abcdhos2)[1]

min(d1+d2)+distance(x01[n],y01[n],X[ju1],Y[ju1])+distance(x02[n],y02[n],X[ju2],Y[ju2])+sum(dis250)
x01[n]
y01[n]
x02[n]
y02[n]
```