

A Study on Spatial Statistical Perspective for Analyzing Spatial Phenomena in the Framework of GIS: an Empirical Example using Spatial Scan Statistic for Detecting Spatial Clusters of Breast Cancer Incidents

공간현상 분석을 위한 GIS 기반의 공간통계적 접근방법에 관한 고찰 : 공간 군집지역 탐색을 위한 공간검색통계량의 실증적 사례분석

이 경 주* 권 일**
Gyoung Ju Lee Ihl Kweon

요 약 지리적 공간상에서 발생하는 대부분의 현상은 서로 인접한 곳에서 유사한 값을 가지는 특성이 있다. 이는 공간자기상관성과 관련이 있으며 공간분석의 존재 이유를 나타내는 개념이다. 또한 지리적 공간상에서 위치에 따라 값의 분포가 다양한 패턴을 보이게 된다. 이러한 패턴은 공간적 변이를 내포하고 있다. 즉, 특정 위치에서 항상 같은 값을 관찰할 수 있다고 단정하기는 불가능하기 때문에 이러한 변이는 본질적으로 확률론적 특성을 지닌다. 이러한 공간자료의 특성들을 무시하고 일반적 통계분석 등을 수행할 경우 공간자기상관성으로 인하여 통계분석에서 가정하는 자료 값들 간 독립성이 위배되고 분석결과는 왜곡될 가능성이 크다. 그러므로 공간자료 분석을 위해서는 공간자기상관성과 확률론적 변이를 적절하게 반영할 수 있는 수단이 필요하다. GIS는 공간적 위치정보를 처리하는데 적합하고 공간통계학은 공간적 변이를 다루는데 유용하다. 따라서 GIS를 기반으로 공간통계학을 통합하는 분석방식은 공간자료의 특성들을 고려하여 유의미한 분석을 하기에 적합한 장점이 있다. 본 연구의 목적은 공간자료 분석에 있어서 공간통계학과 GIS를 결합하는 접근방식의 유용성을 논의하고 실증적 사례분석을 통하여 구체적인 활용성을 살펴보는 것이다. 이를 위하여 공간통계학을 주요 방법론으로 활용하는 공간역학(spatial epidemiology) 분야를 예시적으로 살펴보았다. 구체적으로는 공간검색통계량을 이용하여 미국 Erie 및 Niagara 카운티(New York 주) 내의 유방암 발생의 공간적 군집패턴 분석 논의하였다.

키워드 : GIS, 공간통계학, 공간적 군집패턴, 공간검색통계량

Abstract When analyzing geographical phenomena, two properties need to be considered. One is the spatial dependence structure and the other is a variation or an uncertainty inhibited in a geographic space. Two problems are encountered due to the properties. Firstly, spatial dependence structure, which is conceptualized as spatial autocorrelation, generates heterogeneous geographic landscape in a spatial process. Secondly, generic statistics, although suitable for dealing with stochastic uncertainty, tacitly ignores location information implicit in spatial data. GIS is a versatile tool for manipulating locational information, while spatial statistics are suitable for investigating spatial uncertainty. Therefore, integrating spatial statistics to GIS is considered as a plausible strategy for appropriately understanding geographic phenomena of interest. Geographic

* Full-time Lecturer, Dept. of Urban Engineering, Chungju National University, lgjracer@ut.ac.kr

** Professor, Dept. of Urban Engineering, Chungju National University, ikwon@ut.ac.kr (Corresponding Author)

hot-spot analysis is a key tool for identifying abnormal locations in many domains (e.g., criminology, epidemiology, etc.) and is one of the most prominent applications by utilizing the integration strategy. The article aims at reviewing spatial statistical perspective for analyzing spatial processes in the framework of GIS by carrying out empirical analysis. Illustrated is the analysis procedure of using spatial scan statistic for detecting clusters in the framework of GIS. The empirical analysis targets for identifying spatial clusters of breast cancer incidents in Erie and Niagara counties, New York.

Keywords : GIS, Spatial Statistics, Geographic Hot-Spot, Spatial Scan Statistic

1. Introduction

Most phenomena observed in geographic vessels such as disease/crime incidents in urban areas have intrinsically spatial nature. The spatial nature makes it significant to deal with underlying processes in a special framework like spatial analysis due to two distinctive properties inherent spatial data[19]. Firstly, observations distributed over geographical space are interrelated to other nearby observations. This property is based on the so-called first law of geography that the things closer together tend to be more similar than those farther apart[24]. This property leads to the spatial dependence structure among observations, establishing a fundamental conceptual ground for requiring special framework of spatial data analysis[6, 10].¹⁾ Eventually, there is no room for considering the term 'spatial' as special in analyzing spatial data if the relationships among geographic observations are homogeneous regardless of spatial proximity quantified mostly as distances among observations. Secondly, the spatially heterogeneous relationships among geographic incidents are stochastic in the occurrence. This implies that statistical variability or uncertainty inherent in spatial data needs to be methodologically accounted for in spatial analysis[17, 21].

The two properties in spatial phenomena necessitates the special framework for analyzing spatial data, attracting significant research interests in

many relevant fields[5]. Most highlighted methodological focus among others is to incorporate various statistical measures to GIS. Statistics have the well-established theoretical and empirical advantages in dealing with stochastic processes characterized by variability of uncertainty, while GIS is systematically equipped with various means for analyzing, visualizing spatial data and broadening the process of hypothesis testing[1, 9].

Statistical techniques, however, have a fundamental problem of assuming independence among observed data. If the assumption is violated, increased substantially is the possibility that generic statistical methods may lead researchers to wrong conclusions[20, 21]. This is due mostly to the fact that observations at nearby locations are not independent as discussed above[9]. Spatial statistics explicitly considers the geographic locations information in the definition of relevant measure. GIS provides versatile, analytical tools for efficiently handling locational information. Considering respective advantages together, spatially stochastic data can be properly analyzed in the framework of combining statistics with GIS[1, 20].

In other perspective, research highlights in the spatial statistics literature are put on identifying local anomalies in addition to gauging the overall or global trend of interested geographic phenomena[23, 22, 15]. Most statistical results are presented in the form of descriptive measures such as average, standard deviation, range, etc., as a representative

1) The spatial dependence structure often refers to as 'spatial autocorrelation', the term defining the degree of similarity among geographically adjacent observations. The similarity is often due to a spatial spillover effect and/or distance decay effect(Getis, 1999).

quantity for the entire population or within the whole study area. The global measures provide only a limited information and sophisticated insights for grasping geographic processes are barely obtainable. Since spatial processes are heterogeneous over different geographic locations, local details of spatial variations provide a broad, sophisticated spectrum of practical information for comprehending the spatial phenomena of concern[4]. Furthermore, even when global statistics fails to reject the null hypothesis of no spatial pattern of clustering, local statistics are useful to uncover local anomalies of increased incidence[23]. Local statistical measures are rigorously developed for analyzing spatial data in various purposes and GIS is apt to portraying significant features of concerned local areas by mapping the associated local statistics. For example, if finding out the anomalous local areas showing elevated rates of disease or crime incidents that are statistically significant is of concern, relevant local statistics can be effectively visualized on a map to aid to identify where to prioritize for allocating scarce public resources such as police patrols, further epidemiological investigations.

A notable categorization of utilizing both global and local spatial statistics is the work of Besag and Newell(1991) where three types of spatial statistical tests were suggested. They are general tests, focused tests, and tests for the detection of clustering. General tests focus on investigating the global spatial pattern characteristics such as overall tendency towards clustering or random distribution. It is a single summary measure of denoting overall geographic patterns. This category includes Moran's I (1948), Getis-Ord's G (1992), Geary's C (1954) statistics, K -function. Focused tests intend to examine whether spatial clusters are formed around one location or more pre-specified foci. Score test[14], Tango's C_F statistic, Stone's test (1988), etc. fall into this category. Tests for the detection of clustering point to the scan type statistics devised for exploring sub-areas that correspond to statistically significant local spatial clusters. Examples of this type

are Openshaw et al.(1987)'s Geographical Analysis Machine(GAM), Kuldorff(1997)'s Spatial Scan Statistic, Besag and Newell(1991)'s test, Fotheringham and Zhan(1996)'s method, Turnbull et al.(1990)'s CEPP(Cluster Evaluation Permutation Procedure).

Among other geographic domains, intensive attention has been paid to spatial epidemiology for finding out the local areas with unusually high rate of disease incidents and infer possible environmental factor that may have potential impacts. The cholera outbreak in London's Soho area in 1845 is a famous, historical example of this research interest. Dr. John Snow identified the clusters of many deaths in just visual map and blocked the Broad Street pump where the cluster is around. He could end the spread of the disease that claimed thousands of lives. Detecting clusters of cholera saved a lot of lives, even though detecting cluster in Snow's case was through the naked eye.

Recent advances in spatial statistics and GIS have sophisticated the methodological procedure for detecting and visualizing potential local areas where outbreaks or pandemic of various diseases (e.g., breast cancer, leukemia, influenza, etc.) are of substantial geographic concern. Arguably, among the three types of utilizing spatial statistics, tests for the detection of clustering are the most appropriate correspondents to the methodological advance associated with Snow's classical spatial analysis. This is partly due to the fact that attentions are often paid to effectively visualize analytic outcomes so that intuitive interpretations are easily accompanied by significant outcomes equipped with solid analytic procedures. Eventually, visually intuitive local information that procedure for detecting local clusters intends to provide make it easy to pinpoint where to concentrate scarce public resources in the decision-making processes. Furthermore, the spatial statistical analysis framework of this category coupled with GIS can be methodologically extended to benefit other spatial domains with similar research interests such as criminology, road ecology, etc. by providing useful local insights that may have crit-

ical impacts on relevant decision-making processes.

Based on the argument that emphasizes the utility of the third category in addressing many geographical issues, this article aims at demonstrating the procedure of utilizing spatial statistics of the third category(detection of spatial clusters) in the framework of GIS to identify local anomalies of geographic phenomena. The spatial scan statistic, among others, was chosen to illustrate the procedure for locating local hot-spots of disease incidents. Application domain for the illustration is adopted as spatial epidemiology. Specifically, the breast cancer incidents occurred in 1993-1997, in Erie and Niagara Counties, New York was analyzed. A software was developed as a simple GIS platform where location information is internally manipulated in calculating the spatial scan statistic.²⁾ The strategy of utilizing spatial statistics on a GIS platform have potential benefits by extending fruitfully to various geographic research areas targeting for exploring and analyzing spatial data to identify local non-random features and provide associated spatial decision-making information. The paper is organized as follows. After the introduction(section 1), section 2 introduces spatial scan statistic. Section 3 applies the statistic to empirical data which contains breast cancer incidents in the study area and presents the interpretation of the analysis results. In section 4, conclusion is made and further research issues on integrating spatial statistics to GIS are discussed.

2. Methodology

To define spatial scan statistic, required are location information on cases and controls. In the con-

text of spatial epidemiology, cases refer to as individuals with disease while controls indicates individuals without disease in a study area. To obtain the statistic, the numbers of cases and controls falling into subregions (e.g., census tracts, zip code boundaries, etc.) that constitute a study area are marked and the likelihood ratios of alternative and null spatial patterns are calculated.

The null spatial pattern indicates that there is no local tendency of clustering in the study region and therefore, the probability of an individual being a case takes a constant value. The alternative spatial pattern, on the other hand, implies that the likelihood of more cases being clustered around a center conditioned on a certain radius is greater than the constant obtained in the null pattern. Then, the ratio of the two likelihoods are calculated. Eq.(1) denotes the likelihood ratio difference based on Bernoulli model.³⁾ Usually, the centroid of each subregion is used as the center of the circular window and the distances are computed using the coordinates of centroids in calculating the statistic. The parameters comprising the likelihood ratio of Bernoulli model are presented in Figure 1.

$$LR = \frac{L_A}{L_0} = \frac{(p_A)^c (1-p_A)^{n_i-c} (q_A)^{C-c} (1-q_A)^{(N-n_i)-(C-c)}}{(p_0)^C (1-p_0)^{N-C}} \quad (1)$$

If more cases are located inside the circled window than the counterparts outside and as the likelihood ratio increases, the probability that incidents inside may not happen purely by chance also increases. Then, the statistic is simply defined as the maximum value of the likelihood ratio as denoted in Eq(2).

$$Spatial\ Scan\ Statistic = Max(LR) \quad (2)$$

2) The integration of spatial statistics (spatial scan statistic in this paper) and GIS is based on tight-coupling strategy(Malczewski, 1999).

3) Kuldorff(1997) suggested two types of the statistic. Each type assumes different underlying probability model. One is Bernoulli model and the other is Poisson model. If data are in the form of binary counts such as 'with(success)' and 'without(not success)', Bernoulli model is more suitable. Poisson model, on the other hand, deals with population at-risk exposed to the potential risk factors. Therefore, if the data reflects some continuous risk factors of disease incidence such as the years lived in a certain area, etc. Poisson model fits better. In this study, Bernoulli model was chosen for the ease of explanation in the illustration.

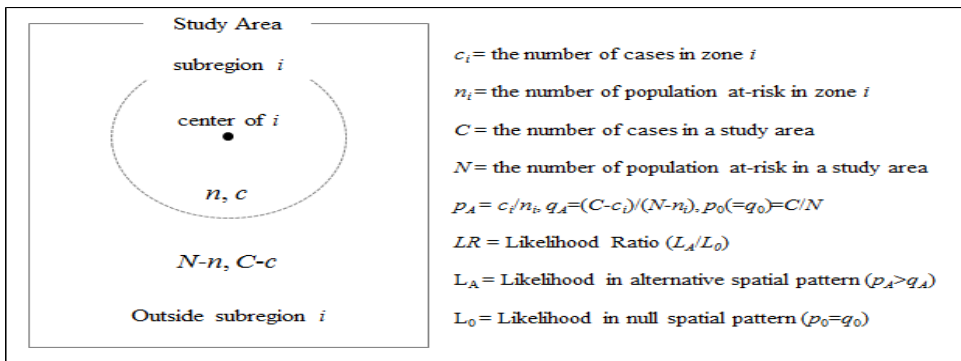


Fig. 1. The parameters of likelihood ratio of Bernoulli model

Note that the location of center and radius of the window are two inherent parameters that determines LR , even if not explicitly specified in the definition. For all centers and a range of radii of the circular windows, a set of LR s is calculated and maximum is selected as a test statistic. The outer limit of the search radius is set so that the number of controls (individuals without disease) in subregion i (c_i) does not exceed 50% of the regional total of controls (C).

To test the statistical significance of the statistic, Monte Carlo simulation is required.⁴⁾ In the simulation, incidence cases are randomly located in a multinomial framework based on the size of population in each subregion and the statistic for the randomly redistributed cases is calculated. This is repeated many times (e.g., 999 times) and the observed maximum value of likelihood ratio is compared with a set of simulated maximums. If the observed LR exceeds a statistical threshold, the null hypothesis of no spatial clustering is rejected for the location of the maximum on a given radius. For example, if the number of simulation is 999 and the threshold is set to, for example, 95% ($\alpha=0,05$), the observed value should to be greater than the upper 950th simulated value to reject the numm hypo-

thesis. In this case, it is barely stated that the observed maximum likelihood ratio is the product of random spatial process and thus, it can be assured that the location marked with the observed maximum LR is statistically significant conditioned on a preset critical value. Same procedure can be applied to comparing the second, the third, etc. maximum likelihood ratios with their simulated counterparts.⁵⁾

3. Empirical analysis and interpretation

Empirical analysis was carried out for Erie and Niagara counties, New York. The study area is composed of total 80 zip code boundaries. In each zip code unit, the variable denoting the case is the number of breast cancer incidents by zip code unit during the period of 1993 to 1997. The number of female population denotes the control variable since breast cancer occurs in the female. The dataset was downloaded from the webpage of the New York State Department of Health.⁶⁾

Figure 2 shows the study area partitioned in zip code boundaries. The three circles represent the location and the size of the first, the second, and the third clusters detected. In the left side map of Figure

4) It is fortunate that the spatial scan statistic is not exposed to the multiple testing problems since the statistical significance of the statistic is tested only for the maximum one and therefore the complicated issues of applying Bonferroni adjustment for multiple hypothesis testings are conveniently avoidable.
 5) Sometimes, it is addressed that comparing the second, the third maximums to the first one to test the statistical significances is too conservative.
 6) The website is <http://www.health.state.ny.us/nysdoh/cancer/csii/nyscsii.htm>.

2, the clusters are identified by zip code and the most likely cluster is found in the upper middle part of Erie county. Location of the first cluster is marked in Amherst area which is densely populated as one of old suburbs in Erie county. Although the whole study area is white-female dominant except the City of Buffalo colored in light tone, female population ratio is relatively small in the area containing the first cluster. It is intriguing that all the clusters are located in close proximity in suburban area, neighboring the City of Buffalo.

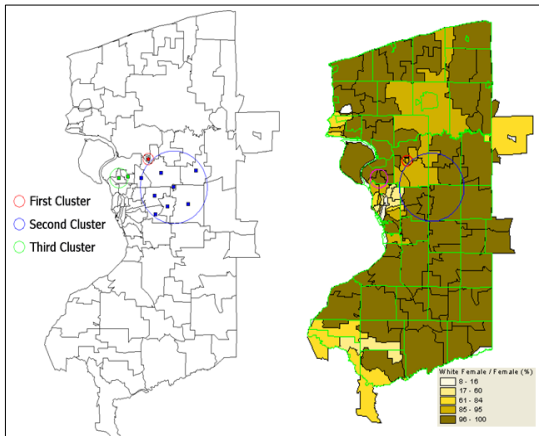


Fig. 2. Detected clusters and white female / overall female population (%)

To test the statistical significance of detected clusters, the null distribution assuming no clustering with spatial randomness was simulated through 999 Monte Carlo replications, using the custom program. As a result, *p*-values of most clusters are 0.001, which means that the rank of LLR (Log

Likelihood Ratio) of observed cases among 999 Monte Carlo replications is 1. Therefore, it is extremely unlikely that the clusters of breast cancer during the study period(1997~1997) have been formed just by chance. They are all statistically significant and it is statistically plausible to say that there are a spatial process forcing the patterns to be formed and therefore, further epidemiological investigation is worth making into to find out possible causes.

The information of clusters is summarized in Table 1. Here, the extent of the most likely cluster roughly corresponds to just one zip code boundary (14228) with very high relative risk. The relative risk (RR) is defined as the ratio of observed to expected cases. It is defined as follows in Eq.(3)

$$RR = \frac{\frac{c_i}{n_i}}{\frac{C}{N}} = \frac{c_i}{p_i \times \frac{C}{N}} = \frac{O_i}{E_i} \quad (3)$$

The notation of Eq.(3) is presented in Figure 1. The relative risk in subregion *i* indicates that how deviant observed case counts are from expected condition. Expected condition is simply the per capita counts in consideration of population size in each subregion. The relative risk in this cluster is about 7 times bigger than the county-wide one. Secondary clusters also shows higher rate of relative risk than the county-wide rate.

One thing that should be noted is that socio-demographic characteristics were not adjusted due to some limitations in this study. For example, as shown in the right side map of Figure 2, white female is dominant in study area. In most zip code

Table 1. Summary information of analysis results

	Zip codes in the detected cluster	Population at-risk	Observed count	Expected count	Relative Risk
1st	14228	1051	54	7.27	7.43
2nd	14026, 14221, 14043, 14225, 14086, 14031, 14227, 14226	109470	935	756.72	1.24
3rd	14217 14223	27664	256	191.23	1.34

Table 2. Log Likelihood Ratio of detected clusters and associated p-values

	Log Likelihood Ratio (LLR)	P-value
1st	62.9	0.001
2nd	24.22	0.001
3rd	10.49	0.001

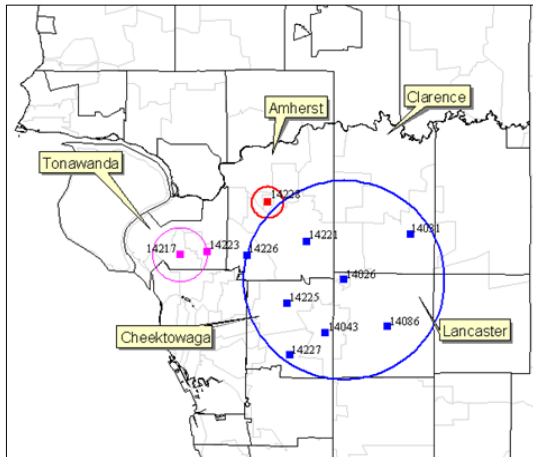


Fig.3. Detailed map of detected clusters

boundary, the rate of white female to overall female is over 95%. Thus, it is probable that the adjustment by race in female population would not make a significant difference in detecting clusters. Data availability was another limitation in adjusting process.

In Figure 3, the municipal boundaries that contain the clusters with statistical significance are Amherst, Tonawanda, Clarence, Cheektowaga and Lancaster area, which are suburbs neighboring the City of Buffalo. It looks that second cluster is very close to the first one and it covers a wide extents of adjacent geographies. The possible inference is that zip code in the first cluster may have strong influence to nearby zip codes contained in the second cluster which are also statistically significance

(p-value less than 0.01 in Table 2)

One final note is that analysis was carried out by making a stand-alone GIS program that implements spatial scan statistic.⁷⁾ Spatial scan statistic like other spatial statistics requires location information to be intensively manipulated in the analysis process and as stressed previously, GIS is most suitable framework equipped with a variety of location analysis tools such as inter-distance matrix generation function so that spatial pattern variations due to different geographic locations are seamlessly dealt with. This is the most highlighted feature of integrating spatial statistical analysis procedure to GIS framework. Statistical analysis results that accounts for stochastic properties of interested spatial process needs to be presented in various means for the insightful interpretations. Typically, map information associated with tables are efficient visual tools that are easily available in GIS. Notably, simply mapping a variable (e.g., the number of disease counts, etc.) does not shed much lights on comprehending interesting process that creates the spatial patterns like clustering since it is barely possible to consider plausibly the process of investigating variations or uncertainties by just looking at the variable itself on a map. This is the key conceptual touch point where the utility or usefulness of GIS is most highlighted in the effort of figuring out essential properties of concerned spatial processes. The tight-coupling strategy of integrating spatial sta-

7) The spatial scan statistic is computable using SaTScan™ software. The software was developed by Martin Kulldorff with Information Management Systems, Inc (2000). It is downloadable on the web. Although this software is known to have a good performance in analysis, results are provided exclusively in a text-based format, requiring additional awkward works to visualize outcomes in a commercial GIS program. For this reason, in this research, a standalone GIS software equipped with mapping functionality was developed to demonstrate the tight-coupling strategy. The software was written in Microsoft Visual Basic 6.0 IDE. As GIS mapping component in the IDE, ESRI MapObjects2.0 ActiveX control was used.

tistics and GIS demonstratively highlights the significance of analyzing spatial data in GIS-based spatial analysis framework. Figure 4 presents the overall procedure for utilizing the demonstration program.

4. Conclusion and discussions

In this paper, integrating spatial statistics to GIS was highlighted as a spatial analysis strategy. Demonstrated was the procedure of utilizing spatial statistics that identify geographic clusters (local hot-spots of concerned phenomena) in the domain of spatial epidemiology. Specifically, the spatial scan statistic was applied for detecting breast cancer clusters in Erie and Niagara counties, New York state. It was addressed that the integration strategy has some utilities. Firstly, spatial structure where some incidents tend to be concentrated in some local areas makes it plausible to utilize GIS framework internally well-manipulating locational information in the analysis procedure. Secondly, stochasticity of spatial process requires analysis procedure to be equipped with appropriate tools which are suitable for dealing with statistical uncertainty or variations.

Spatial statistical approach fits this condition.

The integration strategy takes such a good advantage of the two utilities of spatial statistics and GIS that it may produce insightful analysis outcomes in both terms of assuring analytical objectivity and interpretational efficiency by visualizing(mapping). This is especially true of when pinpointing local hot-spots of incidents of concerned phenomena like highly elevated number of breast cancer cases or geographical clusters. The spatial scan statistic is one of spatial statistics applicable to this matter, much benefited from the strategy. This is mainly due to the fact that the location and size of clusters detected from analytical procedure are most properly presented in the form of map. In other words, the study is signified by the fact that the rare disease (breast cancer) clusters that sometimes are hard to find out by naked eyes but detectable through applying spatial statistical framework. In addition, p -values of detected clusters were also estimated to determine if the clusters are found just by chance or if there is something interesting that forces the rare disease to be clustered for some reasons.

Some limitations employing the integration strat-

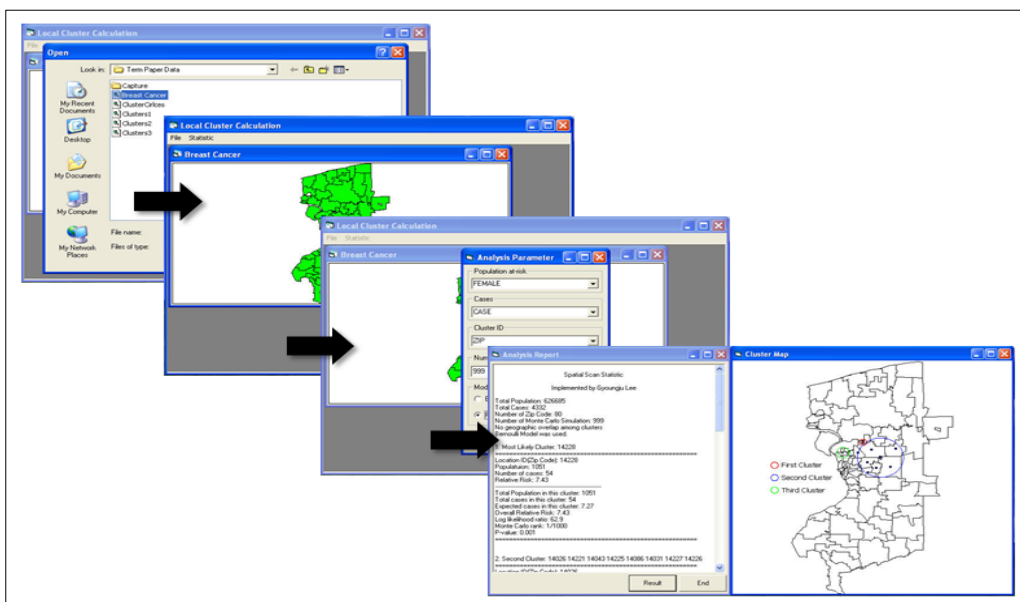


Fig. 4. Demonstration program and running sequence

egy for especially analyzing epidemiological patterns are associated with data problems. Some socio-demographic characteristics of population at-risk could have critical impacts on the analysis outcomes. For example, populations of certain ages groups may have more vulnerable to disease and the variations in population property needs to be appropriately accounted for in a dataset. It is not possible to state that the risk of being a case is constant over any age groups. Other known socio-demographic risk factors associated with breast cancer incidence are listed as race, age, family history of breast cancer, reproductive/hormonal factors, the age of menopause occurring, and the age of a woman at her first full-term pregnancy. The potential and still in-debate risk factors include the exposure to environmental contaminants. Not that the study limitedly aims at demonstrating the spatial statistical analysis procedure integrated with GIS. Although the risk factors were not considered and adjusted in this research, dealing with those risk factors with more detailed data is necessary for more sophisticated analysis and realistic outcome in further research.

Finally, the demonstration of taking advantage of integration strategy was for spatial epidemiology, but the strategy can be fruitfully extended to other relevant domains of spatial analysis. Crime hot-spots detection, for example, has been increasingly routine in a police department, aiding to inform police officers to make rational decisions associated with patrol resource redistributions or restructuring of police precincts. It is not quite recent that GIS has been adopted as a ordinary tool for the tasks. Further research efforts for extending the strategy to a variety of application domains are worth making for enhancing informed decision-making processes.

References

- [1] Bailey, T.C. and A.C. Gatrell. 1995. *Interactive Spatial Data Analysis*, Prentice Hall, Malaysia, PA, USA. pp.24-57.
- [2] Besag, J. and J. Newell. 1991. "The detection of clusters in rare diseases", *Journal of the Royal Statistical Society, Series A* 154(1):143-155.
- [3] Fotheringham, A.S. and F.B. Zhan. 1996. "A comparison of three exploratory methods for cluster detection in spatial point patterns", *Geographical Analysis* 28(3):200-218.
- [4] Fotheringham, A.S. and C. Brundson. 1999. "Local forms of spatial analysis", *Geographical Analysis* 31(4):340-358.
- [5] Fotheringham, A.S., Brundson. C and M. Charlton. 2000. *Quantitative Geography: Perspectives on Spatial Data Analysis*, Sage Publications, London, UK. 15pp.
- [6] Fotheringham, A.S. and P.A. Rogerson. 2009. Introduction. In: A.S. Fotheringham and P.A. Rogerson.(ed.), *The Sage Handbook of Spatial Analysis*. Sage Publications, London, UK. 1-4pp.
- [7] Geary, R.C. 1954. "The continuity ratio and statistical mapping", *The Incorporated Statistician* 5(3):115-145.
- [8] Getis, A. and J. Ord. 1992. "The analysis of spatial association by use of distance statistics", *Geographical Analysis* 24(3): 189-206.
- [9] Getis, A. 1999. "Spatial statistics. In: P.A. Longley et al.(ed.)", *Geographic Information Systems: Principles and Technical Issues*. John Willey & Sons, New York, NY, USA. 239-251pp.
- [10] Haining, R. 2009. "The special nature of spatial data. In: A.S. Fotheringham and P.A. Rogerson.(ed.)", *The Sage Handbook of Spatial Analysis*. Sage Publications, London, UK. pp.5-23.
- [11] Kuldorff, M. 1997. A spatial scan statistic. *Communications in Statistics: Theory and Methods* 26(6):1481-1496.
- [12] Kahng, B. K., I. Kweon, and T. H. Kim, 1997, An analysis methodology of spatial locational character and change of urban micro land use, with GIS & statistical analysis, in the case of Kangnam, Seoul, *Journal of Geographic Information System Association of Korea*, 5(1),

27-41

- [13] Kweon, I., and J. W. Kim, 2002, Urban Land Use Planning with a PSS-based Land Use Change Projection Model, *Journal of Geographic Information System Association of Korea*, 10(4), 515-532
- [14] Lawson, A.B. 1993. "On the analysis of mortality events associated with a prespecified fixed point", *Journal of the Royal Statistical Society, Series A* 156(2):363-377.
- [15] Lee, G., Yamada, A. and P.A. Rogerson. 2010. "GeoSurveillance: GIS-based exploratory spatial analysis tools for monitoring spatial patterns and clusters. In: Fisher, M. and A. Getis. (ed.)", *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Springer, Heidelberg, Germany. pp.135-149.
- [16] Malczewski, J. 1999. *GIS and Multicriteria Decision Analysis*, John Wiley & Sons, New York, NY, USA. 306
- [17] Mallows, C. 1998. The zeroth problem, *American Statisticians* 52(1):1-9.
- [18] Moran, P.A.P. 1948. "The interpretation of statistical maps", *Journal of the Royal Statistical Society, Series B* 10:245-251.
- [19] Opemshaw, S., Charlton, M., Wymer, C. and A. Craft. 1987. "A mark 1 geographical analysis machine for the automated analysis for point data sets", *International Journal of Geographical Information Systems* 1(0):335-358.
- [20] O'Sullivan, D. and D.J. Unwin. 2003. *Geographic Information Analysis*, John Willey & Sons, Hoboken, NJ, USA. 21-55pp.
- [21] Rogerson, P.A. 2001. *Statistical Methods for Geography*, Sage Publications, London, UK. 12-15pp.
- [22] Rogerson, P.A. 2005. "A set of associated statistical tests for the detection of spatial clustering", *Ecological and Environmental Statistics* 12(3):275-288.
- [23] Rogerson, P.A. and I. Yamada. 2009. *Statistical Detection and Surveillance of Geographic Clusters*, CRC Press, Boca Raton, FL, USA. 85pp.
- [24] Stone, R. 1988. "Investigation of excess environmental risks around putative sources: statistical problems and a proposed test", *Statistics in Medicine* 7(6):649-660.
- [25] Tobler, W.R. 1970. "A computer movie simulating urban growth in the Detroit Region", *Economic Geography*. 46(2):234-240
- [26] Turnbull, B.W., Iwano, E.J., Burnett, W.S., Howe, H.L. and L.C. Clark. 1990. "Monitoring for clusters of disease: application to leukemia incidence in upstate New York", *American Journal of Epidemiology* 132(1):136-143.

 Manuscript received : 2011.09.23

Revised : 1차 2012.01.10 / 2차 2012.02.09

Accepted : 2012.02.24



Gyoung Ju Lee

2008. Ph.D. State University of New York at Buffalo

2008~2010. Associate Research Fellow at Korea Research Institute for Human Settlement(KRIHS)

2010~Present. Full-Time Lecturer, Department of Urban Engineering, Chungju National University
Research Expertise : Geographic Information Science, Spatial Statistical Analysis, Land Use Planning, National Territorial & Urban Simulation

Ihl Kweon

1996. Ph.D. Hanyang University

2006~Present. Director, Korea Spatial Information Society

2008~Present. Director, The Korean Regional Development Association

2011~Present. Director, Korea Planners Association
1996~Present. Professor, Department of Urban Engineering, Chungju National University
Research Expertise : Urban Planning, Planning Support Systems(PSS), Spatial Statistical Analysis