

DATA ANALYTICS IN SPATIAL EPIDEMIOLOGY: A SURVEY

Sharmila Banu Kather^{a*}, BK Tripathy^b

^aSchool of Computer Science & Engineering, VIT University, Vellore Campus, Vellore 632014 India

^bSchool of Computer Science & Engineering, VIT University, Vellore Campus, Vellore 632014 India

Article history

Received

16 March 2016

Received in revised form

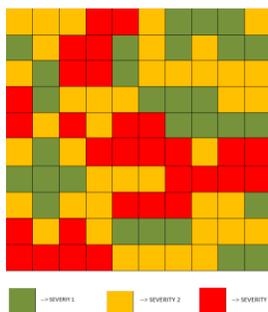
5 July 2016

Accepted

15 September 2016

*Corresponding author
sharmilabanu.k@vit.ac.in

Graphical abstract



Abstract

Spatial data analysis is being used efficiently and the governments have realized that georeferenced data yields more insight with time and locations. Epidemiology is about the study of origin and distribution of diseases and dates back to the 1600s with the instance of cholera in London. Data Science has been evolving and when analyzed with Soft Computing techniques like Rough Set Theory (RST), Fuzzy Sets, Granulation Computing which encompasses the data in its original nature, results can be obtained with accrued accuracy. This survey paper highlights Spatial Data Mining methods used in the field of Epidemiology, identifies crucial challenges and discusses of the use of Soft Computing methods.

Keywords: Spatial epidemiology, data mining, spatial auto co-relation, rough set theory, fuzzy sets, ecological fallacy, demographic shift, incomplete data

© 2016 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

Documenting information about events and people has always been a civilized habit of man's higher order skills. The information or data collected is used to look for interesting and life-saving patterns. Data mining has an elaborate history and is expanding along two dimensions. The first one is the direction in which new algorithms are obtained and the other one is to enhance the existing algorithms or making them more adaptive. When data is collected and studied within a specific spatial frame of reference, it is referred to as Spatial Data Mining. Spatial frames of reference (FoR) are ways of organizing mentally and communicating verbally certain aspects of our spatial knowledge. They represent coordinate systems used to compute and specify the location of objects with respect to other objects. For example, the mutual positioning of some objects in a figure can be described in several ways in English

depending on which object's location is in the focus of our mental attention and our communicative intention, i.e., which object is the located object, or locatum, and which other object in this visual scene is selected as the reference object, or relatum [54]. Such verbal descriptions typically entail a choice of a spatial frame of reference.

Let us consider the statement, "The pond is behind the house". It is a description of the relationship between the pond as the locatum and the house as the reference object in the intrinsic frame of reference. But if we state, "pond is the left of the house" then it is a description of the relationship in the relative frame of reference.

Spatial data mining is the process of uncovering potentially useful and unknown patterns from spatial data [1], [2]. It is the unique nature of spatial data which involves spatial auto-correlation and complexity that calls for redefined exploration of spatial classification, clustering, prediction and co-

location mining [1], [2]. The amount of spatial data generated is explosive as it is generated from a wide range of sources like remote sensing, environmental planning, satellite feeds, GIS etc., [3]. This field of study has impacted different areas like retail, banking, crime analysis, disease mapping, defense etc.

Spatial data mining has been gaining momentum with the huge amount of geo-spatial data being generated every day. Patterns generated from these data are crucial to a lot of government and research organizations. Installation and use of an elaborate Geographic Information System (GIS) infrastructure is considered as a dire need by a lot of governments. Location based services, location prediction, co-location pattern mining, spatial clustering, classification, outlier detection are some of the spatial data mining methods used extensively in a lot of applications. Below, we discuss some of the important methods briefly.

Prediction of critical events that involves lives, revenue and other resources helps with mitigating the loss with such events. This has impact in all major dimensions from global warming, natural disasters, epidemics, vegetative covers, stocks and share markets, consumer trends etc., to bird nesting incidents. Predicting an event is done using the data collected about the occurrence of the event and the location. Further, a grid is enforced on the area of study. Then for each cell in the grid, specific domain attribute values are measured. The attribute value in these cells where the event of interest has occurred is studied further for prediction. Spatial relevance and analysis can contribute effectively to such predictions as these parameters change across countries and continents. Uncovering spatial autocorrelation that is unique of spatial data has been found to be useful and helpful in uncovering spatial structure of errors [1]. Location based services are also focused in different fields to provide exclusive consumer services.

Outlier detection and analysis have been crucial in spatial data sets involving ecological study, health care and urban planning. Outliers which are deviations from consistent measurements have contributed to previously unknown discovery of knowledge. Spatial outliers are based on spatial attributes. Scatter plots are used to detect spatial outliers and are quantitative in nature [4]. Spatial co-location patterns are studied to learn about the multi-dimensional presence or absence of an event in different locations within a spatial frame of reference. This interesting study has helped in identifying disease patterns (West Nile Disease), evolution of animals and plants, crime analysis etc. The outcomes of this study are rules which directly lead to knowledge discovery.

Spatial clustering algorithms group spatial entities that are similar. Hierarchical, partition-based, density-based and grid-based algorithms are used based on the technique using which clustering is carried out. They are used to identify 'hotspots' which are useful

in crime pattern and epidemiological analysis. Spatial clustering has been used for epidemiological data by considering the position of events apart from the domain attribute space. The authors have employed local as well as global measure for spatial auto co-relation and have advocated the inclusion of temporal analysis for further processing.

Change detection algorithms are used for potential Earth science problems like land cover changes and study of habitats. The datasets used to study such problems involve high-dimensions and demand the need for time-series component [5]. Spatio-temporal data mining algorithms have been employed to analyze change detection in earth science. But such detection algorithms have not been categorically analyzed for epidemic models. Spatial data mining algorithms are employed for analysis in urban planning and determining sprawl of urban, rural and peripheral areas. Rough set theory (RST) based indiscernibility and approximation technique has been used to distinguish peri-urban fringes from urban and rural areas [6]. The authors have compared the results of RST based analysis to Map algebra and have highlighted the benefit of approximate reasoning in spatial analysis. The potential of powerful soft computing technique like RST is yet to be explored in epidemiological modeling when large spatial zones are involved.

Disease mapping is considered as an important phase towards understanding the spread of a disease, its impact, spatial relevance and planning effective intervention or precautionary strategies. Geographic Information Systems are more conducive for spatial epidemiology considering the dimension and magnitude of data recorded. Spatial clustering [7] and indices for auto-correlation, geostatistics have been used in spatial epidemiology. Uncertainty based models like rough set theory introduced by Pawlak [8], its variants and fuzzy set theory introduced by Zadeh [9] have been used in the study of epidemiology. Further, research works employing hybrid approaches for spatially oriented epidemiological study have been reported [10], [11], [12]. The hybrid approaches deal with incomplete and ambiguous data offering the benefit of approximate reasoning. Combination of granular computing and soft computing approaches is yet to be explored and it can reap dual benefits of approximation reasoning and lossless dimensionality reduction.

In this section we introduced the context based on spatial data mining, analysis of spatial data and its application in epidemiology. Section 2 discusses the impact of spatial epidemiological studies using Geographic Information Systems (GIS) and its promising applications in healthcare. Section 3 deals with both statistical and artificial intelligence based methods of data analysis in spatial epidemiology. Section 4 highlights the key epidemiological problems and section 5 presents challenges involved in such studies. In section 6 concluding remarks on this study are presented.

2.0 SPATIAL EPIDEMIOLOGY

Spatial epidemiology involves representing and studying geographically mapped health data. Environmental, demographic, behavioral and infectious risk factors are accounted for in the analysis. The epidemiological study at-large is case-based, cohort-based or ecology based. The origin of relating epidemiology and geographic locations dates back to 1800s with Sir John Snow discovering the root cause for cholera which he pinned it down on a water tap that he tracked using a map and distribution of disease in London. Figure 1 shows the distribution of cholera cases around a water pump in the broad way street of London. This finding was phenomenal and is a break-through in epidemiology with maps as early as the mid-19th century.

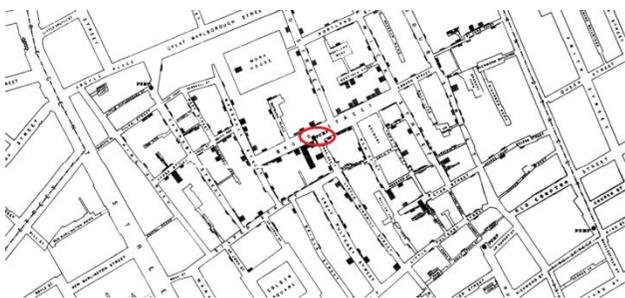


Figure 1 A part of map by John Snow showing the Cholera distribution in London, 1854

The figure is from <https://commons.wikimedia.org/wiki/File:Snow-cholera-map-1.jpg>

Disease mapping, clustering or correlation studies are predominant objectives in a spatial epidemiological study. Exclusive health surveys were conducted as early as 1936 to identify health atlases in the United Kingdom [13]. Later most of the developed countries documented their own national and regional health atlases [14]. These atlases provided a quick visualization of health census in the country and were used for descriptive purposes. While constructing disease maps, size of the geographic unit, method to analyze and the respective attributes are very crucial. This demands domain experts, administrators, engineers, statisticians to work cohesively and collectively.

With the recent advances in data science and advanced soft computing paradigms, constructing decision systems for epidemiological data can only be more sophisticated. Rough Sets [10] and rough fuzzy sets [11] have been applied for epidemiological studies. Inclusion and critical analysis of spatial attributes in the decision systems makes the study more relevant. Spatial auto co-relation which is unique to spatial data and accounts for clustering of formation of hotspots gives an insight into the results arrived in such studies. Tobler's first law of geography proposed in 1970 is the intuition behind the results

[15]. The law states that '*all things are related; but near things are more strongly related than distant ones*'. The underlying concept is spatial auto co-relation and Neighborhood Rough Set has been used to address this natural relation [16].

The advancement in the field of Geographic Information Systems (GIS) has given added breakthrough to deal with health data. It is viewed as an *evidence-based* practice tool and weighs more when used for community health. It is documented in [12], [51], [52], [53] that when properly used, GIS can help in epidemiological perspective by providing informed decision making, perform risk analysis and plan interventions and mitigation strategies. Although setting up GIS infrastructure can be very expensive, it is imperative to consider spatial attribute as an important feature [12]. Besides, if such infrastructures can be used for multiple causes then it becomes beneficial.

Increased use of GIS for health care has been reported for the last two decades in the developed countries. In the developing countries, it is in the primitive stages. However, in either case, it is normally used as isolated systems by various agencies and hence tapping the collective benefit cost-effectively is a distant dream. This calls for integration of public health departments from both the federal levels and national level. Although inspecting the relationship between disease incidence and geographic location can be done using GIS, it is suggested that it can only be exploratory and not confirmatory [17]. Merging health datasets from smaller regions and investigating health outcomes and patterns with respect to the demographics will yield promising results. GIS employs the use of *choropleth maps*, *smoothed maps*, *spatial filters* and geostatistical measures like *kriging* and *trend surface analysis*.

GIS has been used to store and explore electronic medical records of asthma patients in a specific location [18]. Any unusual increase in hospitalization of asthma patients might result in scrutiny of the environmental factors and industrial set-ups in incident locations. GIS is also in African countries for emergency response as well as effective and transparent decision-making [19]. Using demographic data, targeted health education drives were also planned. The work also exhibits the fact that GIS has also been used for delivering health care services using appropriate interfaces. Association between asthma incidences and vicinity of industries along with wind patterns in Northern England has been studied and documented [20]. Report on the discovery of information relating to the past based on the current and prevalent drug use based on UK data [21] is documented. GIS applications are used in hospital resource management and to avoid blocks to services [22]. In the next section we discuss on different methods used in spatial data mining.

3.0 METHODS

Spatial data mining includes clustering, classification, prediction and other determinant techniques. Some of those that are used frequently in spatial epidemiology are discussed further.

3.1 Point Pattern Analysis

Basic components that are mapped about a disease to be represented in GIS or any decision system are points, lines and regions. They correspond to disease location, transmission means or vicinity to affecting sources and the area under study. Disease locations or points of incidents are of particular interest as researchers look for georeferenced information. A pattern based on the distance metric among such disease points is derived in [23]. The distance metric employed is used to look for random pattern as well as collective pattern in the space under investigation. Ripley's K function and Kernel density estimation are employed to study the distribution of events. The former technique identifies dependence between the events while the latter identifies hotspots.

3.2 Spatial Auto Co-Relation Indexes

Moran's index is a global index used to measure how an attribute under analysis is correlated with a geographic location. It takes a value between -1 and +1. If the value is positive; the region is in close association with nearby regions. Negative value indicates disassociation between the regions and zero value hints at an outlier instance. Moran's index gives a degree of association between a set of observed values and weighted average of neighboring values. It is expressed as a matrix. Moran's I is defined as

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

where N is the number of spatial units indexed by i and j and X is the variable of interest \bar{X} is the mean of X and w_{ij} is an element of a matrix of spatial weights [23].

Local indicators are used when local associations are of primary interest and contribute in hotspots. To measure this index, the study region is divided into smaller ones and the attribute in one area is analyzed in relation with the same attribute in proximal area. This leads to the highlighting of clusters.

3.3 Kriging

Kriging is a prediction technique where an event is predicted at a geographic location based on the weighted sum of neighborhood data. This prediction is used in generating maps for values that are unknown. The weights (associated with data) are determined from a mathematical model, Variogram cloud. To arrive at impartial results the weights are set to one. It is one of powerful interpolation techniques and is based spatial auto-correlation which upholds the *Tobler's law of geography* [15]. Adaptive Kriging technique has been used in an integrated data structure used for the habitat modelling of Highland Haggis in Scotland as early as 1995 [44]. This technique is used along with Gaussian Simulation to relate distribution of specific locations in space. [45]. Kriging is used in modelling the irregular distribution of iron grade in Iran [55]. The technique has been used extensively in health care research with respect to distribution and spatial analysis. Kriging has been used to model cancer mortality rate [46] and levels of benzene in urban environment [47]. It is used for mapping disease risks like cholera and dysentery in susceptible regions of Bangladesh [48], to design vector – host models for encephalitis [49], to study spatio-temporal air pollution traits and asthma patient visits in Taiwan [50] are some to quote amidst myriad applications of Kriging in epidemiology.

3.4 Artificial Intelligence Techniques

Artificial intelligence is used in spatial epidemiology through data mining techniques like classification [24] and regression trees [25], multivariate adaption splines and tree based classifiers. Inferring results has been simple and has denoted unambiguous statements. Classification and regression trees are used for classification of drugs and it is used in preventive measures of diabetic complications [26]. Multivariate adaption splines have been useful in avoiding false positives (a test result which wrongly indicates that a particular condition or attribute is present) in high-dimensional data [27]. Although these techniques have been used in epidemiological analysis, it is essential that the results are validated with domain experts to refrain from spurious classification and prediction.

3.5 Rough Set Theory Based Classification

Researchers have used RST extensively for prediction modeling and dimensionality reduction. Uses of RST in medicine for predicting success of surgeries, myocardial infarctions, post-operative long & short term survival of patients, predicting diseases and length of treatment period is documented [28]. Rough Set Theory is employed to identify distribution of neural tube defects in new born children [10]. Dimensionality reduction using *reduct* of rough sets and identifying spatial factors involved in this defect

were the highlights of the work. The work proved to be one of the eye openers to potential research in ecology, epidemiology and medicine.

4.0 KEY EPIDEMIOLOGICAL PROBLEMS

The transmission and distribution of Severe Acute Respiratory Syndrome (SARS) was studied and analyzed by [29], risk exposure pattern is documented in [30], distribution of measles in Turkey has been analyzed [31], Rough Set theory has been used to predict *pancreatitis* [32], ambulation after spinal cord injuries is predicted in [33], a rough set based predictor was built for myocardial infarction [32], spatial decision rules in neural-tube birth defect was uncovered using RST in [10]. Spatial analysis employing statistical models and spatial regression methods were used to study population dynamics [7] and weighted centroid method was used to predict outbreak of *Escherichia Coli* in [35].

To better express the multifaceted nature of the real world and address the limitation of knowledge and uncertainty of factual data, *fuzziness* can be used to represent some attributes of data. It has been used to represent the classification of land-cover types in [36] and effect of environmental factors on birth defects has been substantiated based on discernibility [10], ability to handle inconsistent data, applicability to any number of outcomes, dimensionality reduction, suitability for spatial data are some of the features that make Rough Sets very conducive to epidemiological study.

A geographic phenomenon may tend to be closely related and distant related entities based on the distance. This is spatial auto correlation and upheld by Tobler's *law of geography* [15]. In RST, an object tends to have *roughness* where the object is a subset of universe with some property [8]. Lower and Upper approximations are used to define an object. The roughness of an object can be précised upon collecting more attributes about the object. It is affirmed that *roughness* is not a *fuzzy* concept by nature and so fuzzy sets cannot be used to represent roughness [11]. *Rough Fuzzy Sets* which is an extension of rough sets can be used to construct the decision system for spatial analytics.

Rough Fuzzy Decision system has been used for spatial data analysis which is based on rough fuzzy sets [9]. Rough Fuzzy sets use similarity relation instead of equivalence relation and apply fuzzy granule for similarity relation [37]. Determining the spatial characteristics of the disease distribution will help in identifying the worst affected population and respective demographics. It can further help to build and test theories, plan and evaluate epidemiological surveys, forecast trends and test control measures. The proposed approach is first of its kind and would serve as an important tool for public health researchers and practitioners.

Research on spatial spread of incidence/prevalence of Parkinson's is documented in [36], the spread of human Lyme disease in [39], human cell lymphoma virus type 1 (HTLV-1) in a European Centre for Disease Prevention and Control's technical report have been recorded [40]. These works help in understanding the affected population across a specific area or country and across the world. The geo-referencing of this data gives an insight into the geographies affected with the same disease instances and calls for seeking spatial associations. Including geographical information leads to evidence based spatio-temporal approach in analyzing public health [5]. This work recalls the need for exploring spatial patterns in disease outcomes. It calls for the spatial indices used to identify spatial distribution of disease clusters [41]. Moran's Index, Spatial scan statistic, *k*-nearest neighbour etc., are discussed. Spatial auto correlation is also measured using these statistics.

The effects of environmental pollutants and its association with asthma incidences in North East England have been studied [42]. Environmental threats to human health have been analyzed and classified them as risk, health and hazard indicators [43]. The effect of pollutants from industries on public health can be a decisive study contributing to the welfare of people at large and social cause. Cities that have capitalized from Industries are marked by humongous outlet of effluents and air pollutants. Eventually people living in the affected area are prone to a spectrum of infections and diseases. Data Mining based on soft computing techniques like Rough Set Theory can be used to explore spatial conclusive rules on predominant diseases caused by pollution and look for spatial associations related to the diseases.

5.0 CHALLENGES

Data mining and soft computing techniques depend just as much on data as *statistical* techniques. The infrastructure for the organized collection of data for further analysis is the quintessential phase. This phase needs to be addressed with care and collaborative efforts of administrators, health care professionals, statisticians, engineers etc., The subsequent consideration is due for the attributes considering the high-dimensional nature of this data. Researchers and administrators could resort to powerful dimensionality reduction techniques.

The *health outcome* oriented data will include sensitive information of individuals. Ensuring privacy to sensitive data is becoming an important criterion in the data access policies of governments and corporations. Providing security to data using non-cryptographic based techniques has been in use for a long time. They are used to provide security to Health data, finance data and the like. Data can be distorted using various approaches to hide sensitive

information and provide privacy. A whole line of methods from statistical disclosure control to distortion based techniques are in use. The scale of geographic area studied for disease instances bears significance on the quality of results obtained.

Despite the infrastructure, choice of method applied to analyse data, privacy preserving methods, it is crucial that the results be validated through healthcare professionals and by integrating the research findings in healthcare management infrastructure. Carrying out epidemiological studies to analyse population demographics and health outcomes may incur *ecological fallacy* where a population level inference may be assumed to be individual level inference. Time variance may reflect myriad changes in studies conducted. Hence modeling for the study should consider this *demographic shift*.

6.0 CONCLUSION

We rely on statistical tools to analyse epidemic or endemic distribution of diseases to come up with interventions, precautionary measures or appropriate medical campaigns. This approach needs to be collaborated with high-dimensional data mining and soft computing techniques to access systematic knowledge layers hidden in data. Besides that data-driven nature of soft computing techniques, missing and vague data are better handled rather than the poor choice of ignoring or replacing incomplete data. GIS based infrastructure and integrated collection of health demographics data should become the goal posts of developing countries.

References

- [1] Shekar, S and S. Chawla. 2003. *Spatial Databases: A Tour*. Prentice Hall. 267: 271.
- [2] Shashi, S., M. Evans, J. Kang. 2014. Technical Report on Spatial Data Mining. Department of Computer Science, University of Minnesota
- [3] Deren, L and S.Wang. 2005. Concepts, Principles And Applications Of Spatial Data Mining And Knowledge Discovery. *Proceedings of International Conference on Spatio-temporal Computing conducted by International Society for Photogrammetry and Remote sensing*.
- [4] Luc A. 1994. Exploratory Spatial Data Analysis and Geographic Information Systems. *New Tools for Spatial Analysis*. 45: 54.
- [5] Boriah, S., V. Mithal, A. Garg, V. Kumar, M Steinbach, C. Potter & S. A. Klooster. 2010. A Comparative Study Of Algorithms For Land Cover Change. *CIDU*. 175-188.
- [6] Murgante B., G. Las Casas, A. Sansone. 2007. A Spatial Rough Set For Locating The Periurban Fringe. *SAGEO*.
- [7] Chi, G., and J. Zhu. 2008. Spatial Regression Models For Demographic Models. *Population Res. Policy Review*. 27(1): 17-42.
- [8] Pawlak, Z. 1982. Rough Sets. *International Journal of Man-Machine Studies*. 21: 127-134.
- [9] Zadeh, L. A. 1965. Fuzzy Sets. *Information and Control*. 8: 338-353.
- [10] Bai, H., Y. Ge, J. Wang and Y.L. Liao. 2010. Using Rough Set Theory To Identify Villages Affected By Birth Defects: The Example Of Heshun, Shanxi, China. *International Journal of Geographical Information Science*. 24(4): 559-576.
- [11] Bai, H. and Y. Ge. 2014. A Method For Extracting Spatial Rules From Spatial Data Based On Rough Fuzzy Sets. *Knowledge-based Systems*. 57: 28-50.
- [12] Boulos, M. N. K. 2004. Towards Evidence-Based, GIS Driven National Spatial Health Information Infrastructure And Surveillance Services In The United Kingdom. *International Journal of Health Geographics*. 3(1): 1.
- [13] Stocks, P. 1936. Distribution Of Cancer In England And Wales At Various Sites. *British Empire Cancer Campaign*. 12: 239-280.
- [14] Walter, S. D. and S. E. Birnie. 1991. Mapping Mortality And Morbidity Patterns: An International Comparison. *International Journal Of Epidemiology*. 20(3): 678-689.
- [15] Miller, H. J. 2004. Tobler's First Law And Spatial Analysis. *Annual Association – America*. 94(2): 284-289.
- [16] Jensen, R and R. Shen. 2004. Semantics-Preserving Dimensionality Reduction: Rough And Fuzzy-Rough Based Approaches. *IEEE transactions on Knowledge and Data Engineering*. 16(12): 1457-1471.
- [17] Richards, T. B., C. M. Croner, G. Rushton, C. K. Brown, L. Fowler. 1999. Geographic Information Systems And Public Health: Mapping The Future. *Public Health Report*. 114(4): 359-73.
- [18] Rushton, G. 1998. Improving The Geographic Basis Of Health Surveillance Using GIS. In *Gatrell, A. and Loytonen M. (ed.). GIS and Health*. Philadelphia: Taylor and Francis. 63-80.
- [19] Gavin, E. 2002. *Geo-Information Supports Decision-Making in Africa – An EIS-AFRICA Position Paper*, [http://www.eis-africa.org/DOCS/A5-Engv7.pdf] Pretoria, South Africa: EISAFRICA].
- [20] Dunn, C.E., J. Woodhouse, R.S. Bhopal, S.D. Acquilla. 1995. Asthma And Factory Emissions In Northern England: Addressing Public Concern By Combining Geographical And Epidemiological Methods. *Journal of Epidemiological Community Health*. 49(4): 395-400.
- [21] Field, K., L. Beale, H. Heatlie, M. Frischer. 2001. Using a Geographic Information System to forecast the diffusion of drug misuse. *Proceedings of the 21st Annual ESRI International User Conference: 9–13 July 2001*. San Diego, California [http://gis.esri.com/library/userconf/proc01/professional/papers/pap331/p331.htm].
- [22] Higgs, G. and M. Gould. 2001. Is There A Role For GIS In The 'New NHS'? *Health Place*. 7(3): 247-59.
- [23] Lai, P.C., F.M. So, K.W. Chan. 2009. *Spatial Epidemiological Approaches in Disease Mapping and Analysis*. CRC Press: New York, USA. 2.
- [24] Deconinck, E., T. Hancock, D. Coomans, D.L. Massart, Y.V. Heyden. 2005. Classification Of Drugs In Absorption Classes Using Classification And Regression Trees Methodology. *Journal of Pharma and Biomedical Anal.* 39 (1-2): 91-103.
- [25] Deconinck, E., Q.S. Xu, R. Put, D. Coomans, D.L. Massart, Y.V. Heyden. 2005. Prediction Of Gastro-Intestinal Absorption Using Multi-Variate Adaptive Regression Splines. *Journal of Pharma and Biomedical Anal.* 39(5): 1021-1030.
- [26] Miyaki, K., I. Takei, K. Watanabe, H. Nakashima, K. Omae. 2002. Novel Statistical Classification Model Of Type 2 Diabetes Mellitus Patients For Tailor Made Prevention Using Data Mining Algorithm. *Journal of Epidemiology*. 12(3): 243-248.
- [27] York, T. P., and L. J. Eaves. 2001. Common Disease Analysis Using Multivariate Adaptive Regression Splines. *Genet Epidemiology*. 21(1): 649-654
- [28] Øhrn, A. 1999. Discernibility And Rough Sets In Medicine: Tools And Applications. Thesis (PhD), Norwegian University of Science & Technology.
- [29] Meng, B., J. Wang, J. Liu, J. Wu and E. Zhong. 2005. Understanding The Spatial Diffusion Process Of Severe

- Acute Respiratory Syndrome In Beijing. *Public Health*. 119(12): 1080-1087.
- [30] Wang, J. F. 2006. Spatial Dynamics Of An Epidemic Of Severe Acute Respiratory Syndrome In An Urban Area. *Bulletin of World Health Organization*, 2006. 84(12): 965-968.
- [31] Ulugtekin, N., S. Alkoy, and D. Z. Seker. 2007. Use Of A Geographic Information System In An Epidemiological Study Of Measles In Istanbul. *Journal Of International Medical Research*. 35(1): 150-154.
- [32] Slowiński, K., R. Słowiński and J. Stefanowski. 1988. Rough Sets Approach To Analysis Of Data From Peritoneal Lavage In Acute Pancreatitis. *Medical Informatics*. 13(3): 143-159.
- [33] Ohm, A., L. Ohno-Machado and T. Rowland. 1998. Building manageable rough set classifiers. In *Proceedings of the AMIA Symposium*. American Medical Informatics Association. 543.
- [34] Vinterbo, S. and A. Øhrn. 2000. Minimal Approximate Hitting Sets and Rule Templates. *International Journal of Approximate Reasoning*. 25(2): 123-143.
- [35] Buscema, M., E. Grossi, A. Bronstein, W. Lodwick, M. Asadi-Zeydabadi, R. Benzi and F. Newman. 2013. A New Algorithm For Identifying Possible Epidemic Sources With Application To The German Escherichia Coli Outbreak. *ISPRS International Journal of Geo-Information*. 2(1): 155-200.
- [36] Shi, W. 2005. Principle Of Modeling Uncertainties In Spatial Data And Analysis. Science. Beijing.
- [37] Dubois, D. and H. Prade. 1990. Rough Fuzzy Sets And Fuzzy Rough Sets. *International Journal of General Systems*. 17(2): 191-209.
- [38] Wright Willis, A., Evanoff, B. A., Lian, M., Criswell, S. R. and Racette, B. A. 2010. Geographic And Ethnic Variation In Parkinson Disease: A Population-Based Study Of US Medicare Beneficiaries. *Neuroepidemiology*. 34(3): 143-151.
- [39] Kugeler, K. J., G. M. Farley, J. D. Forrester and P. S. Mead. 2015. Geographic Distribution And Expansion Of Human Lyme Disease, United States. *Emerging Infectious Diseases*. 21: 1455-1457.
- [40] European Centre for Disease Prevention And Control. 2015. Geographic Distribution Of Areas With A High Prevalence Of HTLV-1 Infection. *Stockholm: ECDC*.
- [41] Song, C. and M. Kuldorf. Power Evaluation Of Disease Clustering Tests. 2003. *International Journal of Health Geographics*. 2: 9.
- [42] Dunn, C. E., J. Woodhouse, R. S. Bhopal and S. D. Acquilla. 1995. Asthma And Factory Emissions In Northern England: Addressing Public Concern By Combining Geographical And Epidemiological Methods. *Journal of Epidemiology and Community Health*. 49(4): 395-400.
- [43] Briggs, D. J. 2000. Environmental Health Hazard Mapping For Africa (p.140). Harare, Zimbabwe: WHO-AFRO.
- [44] Oleg, M. 1995. The Integration Of GIS, Remote Sensing, Expert Systems And Adaptive Co-Kriging For Environment Habitat Modeling Of Highland Haggis Using Object Oriented, Fuzzy Logic And Neural Network Techniques. *Computers and Geosciences*. 22(5): 585-588.
- [45] Jeff, B. and C.V. Deutsch. 2010. Programs For Kriging And Gaussian Simulation With Locally Varying Anisotropy Using Non-Euclidean Distances. *Computers and Geosciences*. 7(2011): 495-510
- [46] Pierre, G. 2006. Geostatistical Analysis Of Disease Data: Accounting For Spatial Support And Population Density In The Isoleth Mapping Of Cancer Mortality Risk Using Area-To-Point Poisson Kriging. *International Journal of Health Geographics*. 5: 52
- [47] Kristina, W. W., E. Symanski, D. Lai and A.L. Coker. 2011. Kriged And Modeled Ambient Air Levels Of Benzene In An Urban Environment: An Exposure Assessment Study. *International Journal of Health Geographics*. 10: 21.
- [48] Mohammad, A., P. Goovaerts, N. Nazia, M. Z. Haq, M. Yunus and M. Emch. 2006. Application Of Poisson Kriging To The Mapping Of Cholera And Dysentery Incidence In An Endemic Area Of Bangladesh. *International Journal of Health Geographics*. 5: 45
- [49] Benjamin, G. J., N. D. Burkett-Cadena, J. C. Luvall, S. H. Parcak, C. J. W. McClure, L. K. Estep, G. E. Hill, E. W. Cupp, R. J. Novak and T. R. Unnasch. 2010. Developing GIS-based Eastern Equine Encephalitis Vector-Host Models In Tuskegee, Alabama. *International Journal of Health Geographics*. 9: 12.
- [50] Ta-Chien, C., M. Chen, I. Lin, C. Lee, P. Chiang, D. Wang and J. Chuang. 2009. Spatiotemporal Analysis Of Air Pollution And Asthma Patient Visits In Taipei, Taiwan. *International Journal of Health Geographics*. 8: 26
- [51] Luis, R. B. 2004. Spatial Access To Health care in Costa Rica And Its Equity: A GIS-Based Study. *Social Science & Medicine*. 58(2004): 1271-1284.
- [52] Jiaxi, Z., W. Wang, Z. Tan, Q. Wu, W. Xiao, L. Shang, Y. Zhang, J. Peng and D. Miao. 2014. Spatial Analysis Of Schizotypal Personality Traits In Chinese Male Youths: Evidence From A GIS-Based Analysis Of Sichuan. *International Journal of Mental Health Systems*. 8: 3.
- [53] Sviatlana, P., S. E. Häberg, G. Aamodt, S. J. London, H. Stigum, W. Nystad and P. Nafstad. 2016. Association Between Pregnancy Exposure To Air Pollution And Birth Weight In Selected Areas Of Norway. *Archives of Public Health*. 74: 26.
- [54] Galati, A. and M. N. Avraamides. 2012. Collaborating In Spatial Tasks: Partners Adapt The Perspective Of Their Descriptions, Coordination Strategies, And Memory Representations. In C. Stachniss, K. Schill, & D. Uttal (Eds.). *Spatial Cognition. Lecture Notes In Artificial Intelligence*. Heidelberg, Germany: Springer. 7463: 182-195.
- [55] Badel, M., S. Angorani, and M. S. Panahi. 2011. The Application Of Median Indicator Kriging And Neural Network In Modeling Mixed Population In An Iron Ore Deposit. *Computers and Geosciences*. 37: 530-540.