On the Use of GIS for Health and Epidemiology Control*

Giuseppe Tradigo, Patrizia Vizza, Pietro H. Guzzi, and Pierangelo Veltri

Universit degli Studi Magna Graecia di Catanzaro
Viale Europa, 88100 Catanzaro, Italy
{gtradigo,vizzap,hguzzi,veltri}@unicz.it

Abstract. Monitoring the evolution and diffusion of diseases is an important task for health monitoring. Recent phenomena, such as the spread of Coronavirus (Covid-19), highlighted the relevance of adopting Geographical Information Systems (GIS) in epidemiology modeling. GIS may offer general views and indications which domain experts and governments could use to give immediate directions to social actors and operators.
We report on the possibility of using geographic database model instruments in order to acquire, store and manage health-related data. The reported case study is about an application able to correlate TSH (Thyroid-Stimulating Hormone) with environmental data. The reported example aims to show how to acquire, analyze and integrate clinical and geographical data to evaluate possible correlations for the prevention of chronic diseases, especially neoplasms, by means of mapping disease features with environmental factors.

Keywords: Geographic Information Systems · epidemiology · Virus · TSH.

1 Introduction

In the last years, GIS tools have become more useful in the public health sector, especially in the epidemiology field [1–3]. Epidemiology models represent an important tool to manage and assess environmental risk factors, such as potential cause of illness, by quantifying their impact on human health in population at risk [4]. Therefore, they allow to carry out a number of prevention activities, and also monitoring and evaluating the effectiveness of these activities.

GIS can also be useful in associating epidemiological and health data by defining how environmental factors affect the health status of group of individuals or entire populations living in a territory, in the onset of specific diseases [5–8,

* *Discussion Paper*, a preliminary version of this paper has been published at IEEE BIBM 2019 supported by POR PIH-GIS Regional Funds.

22]. GIS systems and analysis methodologies support domain experts in studying the overlap of clinical and environmental data and can help with the identification of possible relations between clinical features and diseases. These relations can be rendered with rich epidemiological maps which represent the present and future status (prediction) of the on-set of a disease and its human-to-human transmission. Correlations between environmental risk factors and clinical data or events can then be used to evaluate or foresee the impact on human health and are usually be analyzed by using spatial analysis techniques [9–11]. These techniques allow to study disease pathogenesis and etiology, identifying any links between them and the places where they develop [12].

In literature, many contributions exist which focus on spatial epidemiology [13–15]. In [16] authors report on the use of GIS in epidemiology, focusing on methodologies involving geocoding, distance estimation, residential mobility, record linkage and data integration, spatial clustering, small area estimation, and disease mapping. Authors in [17] study the correlation between climate and geographic distribution of tuberculosis by using GIS, aiming to identify high-risk population groups and their geographic areas. Contribution in [19] investigates environmental and geographical influences on epidemiology of acromegaly in Brazil. The authors validate a method to link an acromegaly registry with a GIS mapping module aiming to represent the spatial distribution of patients and to identify disease clusters.

We present an example of GIS application to study and evaluate the correlation between geographical and clinical geographical by using TSH (Thyroid-Stimulating Hormone) data. Shapefiles of environmental layers have been used in GIS to identify specific territorial aspects and their influence on clinical outcomes. A dataset has been used containing data of patients enrolled at Magna Graecia University of Catanzaro. A preliminary version of the paper has been previously published in [18].

2 GIS in health related applications

Environmental monitoring and the analysis of clinical data aim to prevent chronic diseases, especially the neoplasms and their correlation with environmental factors. In literature there exist many definitions of both correlation (e.g. Pearson correlation) and spatial correlation (*scorr* spatial correlation function between two variables over an X-Y domain). For the purposes of the present work, we adopt a loose definition of spatial correlation as follows. Let's consider a clinical event E in the input clinical dataset and a geographical entity G; we say that E and G are spatially correlated if there exists a spatial query Q for which the location of E is *contained* in the area subtended by G. In order to investigate on spatial correlations in our dataset, different environmental layers have been used. These are available on the Geoportal of the Calabria Region in shapefile format [20]. A shapefile is a standard format for vectorial spatial data. It has been developed by the Environmental Systems Research Institute (ESRI) to extend the interoperability between ESRI systems and the other GIS [21]. The shapefile is a non topological format and it is used in Geographic Information System to store positions, shapes and attributes of geographical features. It describes primitive geometric data such as points, lines, polygons and texts, called features, and their associated information, called attributes (e.g., a data describes a river and the associated attributes could be the name or temperature of the river). In such a way, more representations of geographic data can be created, which allow to influence the power and accuracy of geospatial analyses.

The online Geoportal of Calabria Region shares a cartographic data collection containing a number of environmental layers with IODL 2.0 license. Among the available layers, we selected the ones listed below as the most useful for our investigation:

- urban area;
- extra-urban area;
- commercial and industrial settlements;
- public and private service companies;
- rural settlements;
- mining area;
- yards;
- landfill;
- scarp yards;
- local railways;
- airport areas and heliports;
- regional, provincial and district administrative range.

These layers have been used to verify and analyze possible correlations between clinical and territorial aspects. The application of shapefiles and the use of geographic layers are important to assess whether and how some pathologies or in any case clinical alterations of individuals can be directly or indirectly related to strictly territorial aspects.

Table 1. Clinical database extraction. TSH values are expressed in $\mu U/ml$.

Patient	Latitude	Longitude	Altitude	TSH
1	38.632...	16.072...	465.80	2.40
2	39.527...	15.924...	5.85	1.86
3	39.523...	15.962...	277.56	2.07
4	39.533...	15.989...	346.60	2.02
5	39.516...	15.943...	134.49	0.05
6	39.527...	15.935...	188.68	1.26
7	39.511...	15.948...	182.01	1.27
8	38.246...	16.139...	323.80	2.84
9	38.245...	16.138...	301.11	4.23
10	38.246...	16.142...	341.25	1.25
...				
2269	38.698...	15.990...	302.01	2.32

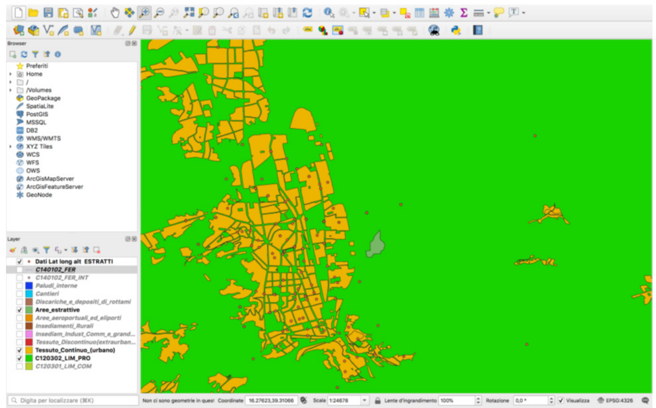

Fig. 1. Topological overlay between clinical data and extracted areas

As case study, we report the study of the correlation between the patterns of topological data with an example of thyroid's data regarding a south Italian region patients. A Health-GIS system has been implemented aiming to cross-check the TSH value with further environmental data through a topological overlay operation. The goal is to identify the geographical areas with the greatest presence of cases with specific pathology. Moreover, it is also interesting to calculate and trace the average distance covered by a patient for a specific health service.

3 Experiments

The health dataset regards single file with .xls extension in which each record represents a patient and contains:

- geo-referenced data (district, home address, latitude, longitude and altitude)
- clinical data (TSH level expressed in $\mu U/ml$)

The clinical dataset contains 2269 records and Table 1 reports an extract of the database we used for the investigation. The QGIS platform has been used to cross-check these clinical data with data relating to environmental factors. QGIS is a free and open source geographic information system to visualize and overlap vector and raster data in different formats and projections without any conversion into a common internal format. The clinical dataset has been converted in .cvs file and loaded into the QGIS system to represent spatial and clinical information together in a geographical map. Patients locations and other logistical data are geo-referenced and related with clinical data. For instance, red dots in Figure 1 refer to patients' location data extracted from the considered clinical dataset and projected on extractive areas polygons.

Starting from this representation, environmental layers listed above have been correlated with clinical dataset and some of these correlations are reported below.

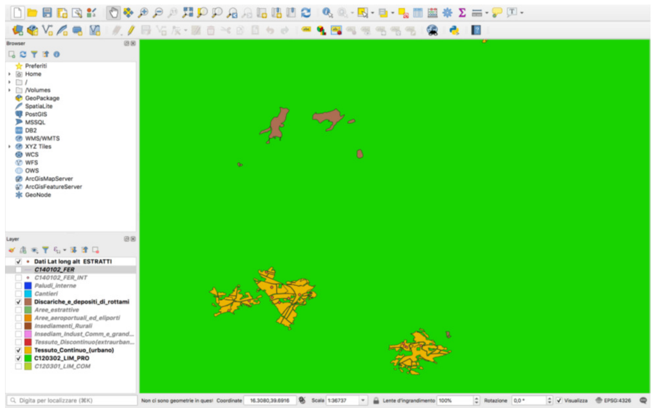

Fig. 2. Topological overlay between clinical data and landfills and scrap yards.

Since several diseases have a close link with overexposure to toxic substances, it is relevant to measure the occurrence of patients locations with potential toxic areas. Air pollution is one of the major environmental risks and in 2015 it has been recognized by the World Health Assembly (WHA) as one of the world's major public health problems. Pollutants are elements and chemical compounds deriving mostly from human activities such as, for example, industrial and commercial production. For this reason, the onset of several cancer cases in a specific geographic area has led to the urge to investigate and monitor industrial activities insisting on the same area.

Figure 1 shows the topological overlay between clinical data (points in red) and extractive areas (dark green areas) also reporting urban areas (areas in yellow). The extraction areas generate a series of pollutants whose exposure has been associated with a series of harmful effects on human health, especially for the respiratory and cardiovascular system and for neoplasms. Pollutants are mainly responsible for a substantial impoverishment of the soil (e.g., caused by deforestation and intensive agriculture), a continuous deterioration of water quality (e.g., caused especially by industrial discharges), and a growing contamination of the air (e.g., caused from the presence of fine dust and particulate in smog). Many substances such as, for example, nitrogen oxide, sulfur oxide, ozone and dust act as irritants; pollen and other agents present in the air are responsible for allergies such as asthma and seasonal colds; some toxic agents such as benzene and polycyclic aromatic hydrocarbons are carcinogenic; carbon monoxide compromises the transport of oxygen by the blood with serious effects on the brain. Other metals come into contact with various organs and tissues causing biological alterations and causing damage to the heart and the apparatus.

Figure 2 shows the topological overlay of clinical data with the level of landfills and scrap deposits (areas in brown), always bringing the urban area (areas in yellow) as reference. The economic growth of the most industrialized countries has generated a substantial increase in consumption and urbanization with a consequent increasing in waste production and in its difficult regular disposal. Waste has negative consequences on the environment and on human health. Asbestos is harmful to human health; breathing dust containing asbestos fibers can cause serious diseases, pleural tumors (pleural mesothelioma) and lung cancer.

4 Future Works

Spatial information in clinical data is crucial for prevention and monitoring evolution of global epidemiology. In this paper we analyze environmental data related with health related ones. We are working to complete a model and general purpose system able to relate clinical, environmental and epidemiology related information for real time data analysis and simulation. The aim is to support prevention in a large scale context. Our framework is general enough to support the loading and preprocessing of generic datasets in standard format (e.g. csv). We are planning to integrate user-friendly functionalities to import different datasets. We are also working on consolidating the representation of time as an analysis dimension and refining its use in the data analysis pipeline by adopting time series analysis techniques.

Acknowledgment

This research has been supported by POR CALABRIA FESR-FSE 2014-2020 PIH-GIS project CUP *J88C17000320006*. The authors thank Giuseppe Brescia for his contribution to a preliminary version of the paper.

References

1. M. Craglia and R. Maheswaran, GIS in public health practice. CRC press, 2016.
2. G. Tradigo, P. Vizza, P. Veltri, P. H. Guzzi, An Information System to Track data and processes for food quality and bacterial pathologies prevention. Italian Symposium on Advanced Database Systems (SEBD), 2019.
3. G. Tradigo, C. Pagliaro, G. Canino, P. H. Guzzi and P. Veltri, GIS for the analysis and monitoring of environmental issues related to life quality. Italian Symposium on Advanced Database Systems (SEBD). 320-326, 2015.
4. A. B. Lawson, S. Banerjee, R. P. Haining and M. D. Ugarte, Handbook of spatial epidemiology. CRC Press, 2016.
5. S. M. Fletcher-Lartey and G. Caprarelli, Application of GIS technology in public health: successes and challenges. Parasitology, 143(4): 401-415, 2016.
6. G. Tradigo, P. Veltri, O. Marasco, G. Scozzafava, G. Parlato and S. Greco, Studying neonatal TSH distribution by using GIS. In Proceedings of the First ACM SIGSPATIAL International Workshop on Use of GIS in Public Health, 36-39, 2012.
7. G. Ambrogio, R. Conte, L. De Napoli, G. Fragomeni, F. Gagliardi, Forming approaches comparison for high customised skull manufacturing. Key Engineering Materials, 651, 925-931, 2015.

8. M. V. Caruso, V. Gramigna, G. F. Serraino, A. Renzulli, G. Fragomeni, Influence of aortic outflow cannula orientation on epiaortic flow pattern during pulsed cardiopulmonary bypass. Journal of Medical and Biological Engineering, 35(4):455-463, 2015.
9. T. Olsson, L. F. Barcellos and L. Alfredsson, Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. Nature Reviews Neurology, 13(1): 25, 2017.
10. M. Chin-Chan, J. Navarro-Yepes and B. Quintanilla-Vega, Environmental pollutants as risk factors for neurodegenerative disorders: Alzheimer and Parkinson diseases. Frontiers in cellular neuroscience, 9: 124, 2015.
11. G. Fragomeni, R. Iannelli, G. Falvo D'Urso Labate, M. Schwentenwein, G. Catapano, Validation of a novel 3D flow model for the optimization of construct perfusion in radial-flow packed-bed bioreactors (rPBBs) for long-bone tissue engineering. New Biotechnology, 52:110-120, 2019.
12. E. L. MacQuillan, A. B. Curtis, K. M. Baker, R. Paul and Y. O. Back, Using GIS mapping to target public health interventions: examining birth outcomes across GIS techniques. Journal of community health, 42(4): 633-638, 2017.
13. L. N. Carroll, A. P. Au, L. T. Detwiler, T. Fu, I. S. Painter and N. F. Abernethy, Visualization and analytics tools for infectious disease epidemiology: A systematic review. Journal of Biomedical Informatics, 51: 287-298, 2014.
14. G. Canino, P. H. Guzzi, G. Tradigo, A. Zhang and P. Veltri, A system for Geoanalysis of Clinical and Geographical Data. In Proceedings of the Third ACM SIGSPATIAL International Workshop on the Use of GIS in Public Health, 57-62, 2014.
15. G. Tradigo, C. Pagliaro, G. Canino, F. Casalinuovo, C. Graziani and P. Veltri, A model for the Geographical Analysis and monitoring of agricultural areas example and tests in south Italian regions. ACM SIGSPATIAL HealthGIS, 2014.
16. R. S. Kirby, E. Delmelle and Jan M. Eberth, Advances in spatial epidemiology and geographic information systems. Annals of Epidemiology, 27(1): 1-9, 2017.
17. R. Beiranvand, A. Karimi, A. Delpishes, K. Sayehmiri, S. Soleimani and S. Ghalavandi, Correlation Assessment of Climate and Geographic Distribution of Tuberculosis Using Geographical Information System (GIS). Iranian Journal of Public Health, 45(1): 86-93, 2016.
18. G. Tradigo, P. Vizza, G. Brescia, P. H. Guzzi, P. Veltri, A geographical patients based health information system. IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019.
19. L. A. Naves, L. B. Porto, J. W. Corra Rosa, L. A. Casulari, J. W. Corra Rosa, Geographical information system (GIS) as a new tool to evaluate epidemiology based on spatial analysis and clinical outcomes in acromegaly. Pituitary, 18(1): 8-15, 2015.
20. Geoportale Nazionale, available online at http://www.pcn.minambiente.it/mattm/.
21. ESRI Shapefile Technical Description. An ESRI White PaperJuly, available online at https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf, 1998.
22. Y. Aliakbarpoor, S. Comai, G. Pozzi, Designing a HL7 compatible personal health record for mobile devices, IEEE 3rd International Forum on Research and Technologies for Society and Industry, (RTSI), 1-6, 2017.
23. Caruso, M.V., Gramigna, V., Serraino, G.F., Renzulli, A., Fragomeni, G., Influence of aortic outflow cannula orientation on epiaortic flow pattern during pulsed cardiopulmonary bypass, Journal of Medical and Biological Engineering, 35(4): 455-463, 2015.