MultiMaps: a tool for decision-making support in the analyzes of multiple epidemics

Gesiel Rios Lopes
Alexandre C. B. Delbem
gesielrios@usp.br,acbd@icmc.usp.br
Institute of Mathematical and Computer Sciences
University of São Paulo
São Carlos, São Paulo, BRA

Roberto Fray da Silva
roberto.fray.silva@gmail.com
Institute of Advanced Studies
University of São Paulo
São Paulo, São Paulo, BRA

Cláudio Bielenki Júnior
Sérgio Henrique Vannucchi Leme de Mattos
bielenki@ufscar.br,sergiomattos@ufscar.br
Department of Hydrobiology
Federal University of São Carlos
São Carlos, São Paulo, BRA

Denise Scatolini
denise.scatolini@saocarlos.sp.gov.br
Center for Disease Control and Prevention
São Carlos Municipality
São Carlos, São Paulo, BRA

Filippo Ghiglieno
filippo.ghiglieno@df.ufscar.br
Physics Department
Federal University of São Carlos
São Carlos, São Paulo, BRA

Antonio Mauro Saraiva
saraiva@usp.br
Polytechnic School
University of São Paulo
São Paulo, São Paulo, BRA

ABSTRACT

The decision-making process for complex problems based on heterogeneous and multiple data sources requires structuring information with adequate representation for the phenomenon under analysis. This is specifically true for temporal and spatial problems, as those dimensions add complexity to data processing, information extraction, and interpretation of results. However, most available methodologies and tools were not developed considering multiple stakeholders and feedback loops. Therefore, a tool is needed to allow managers and researchers to gain deep insight into significant volumes of heterogeneous data to make better decisions. In this context, this paper introduces an open-source tool that consistently integrates data from different sources and contexts, optimizing and facilitating the analysis, management, and representation of space and the phenomena that occur in it. The tool uses techniques from multi-criteria analysis to define the influence of the data and map algebra to combine these data and build thematic maps from multiple heterogeneous sources. This allows users to perform geospatial statistical analysis, outlier detection, and evaluation of regions of interest without the need for specific knowledge in modelling the phenomena, supporting decision-making. The tool is then implemented on a case study for evaluating epidemics and infectious and neglected diseases in an urban environment in Brazil. The insights obtained allow the decision-maker to understand better the data inputs and the different possible results of the multi-criteria decision-making model, considering different weight combinations. It is vital to observe that the tool can be used in different contexts, areas, and data sources with different spatial and temporal aspects.

SpatialEpi '22, November 1, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9543-4/22/11…$15.00
https://doi.org/10.1145/3557995.3566119

CCS CONCEPTS

• **Applied computing → Multi-criterion optimization and decision-making**.

KEYWORDS

Multi-criteria decision-making, Integration of multiple sources, Spatio-Temporal data analysis, Map Visualization, Map algebra

ACM Reference Format:
Gesiel Rios Lopes, Alexandre C. B. Delbem, Roberto Fray da Silva, Cláudio Bielenki Júnior, Sérgio Henrique Vannucchi Leme de Mattos, Denise Scatolini, Filippo Ghiglieno, and Antonio Mauro Saraiva. 2022. MultiMaps: a tool for decision-making support in the analyzes of multiple epidemics. In *The 3rd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology (SpatialEpi '22) (SpatialEpi '22), November 1, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3557995.3566119

1 INTRODUCTION

Decision-making for complex problems based on heterogeneous and multiple data sources requires structuring information with adequate representation for the phenomenon under analysis. Dimensions such as temporal and spatial add complexity to data processing, information extraction, and interpretation of results [13].

Many works in the literature address the use of multi-criteria decision-making methods (MCDM), such as the Analytical Hierarchy Process (AHP), to overcome this significant challenge. In [5] the authors use the Analytical Hierarchy Process (AHP) method with map algebra to determine the environmental fragility of the Rio Brilhante watershed in Mato Grosso do Sul, Brazil, from the superposition of natural and anthropic factors. The authors use the AHP method to define a set of weights with the level of importance among the variables with the most significant influence on the definition of the environmental fragility of the basin from the paired analysis by an expert.

However, the works in the literature do not provide a friendly and intuitive interface where users can integrate their data sources and define the importance of these sources through multi-criteria methods. In this context, this work presents MultiMaps, an open-source tool that consistently integrates data from different sources and contexts, optimizing and facilitating the analysis, management, or representation of space and the phenomena that occur in it, enabling users to perform geospatial statistics, detection of outliers and evaluation of regions of interest, supporting decision-making.

The main components of MultiMaps are: (1) Load Data from Multiple Sources anonymously; (2) Geolocate and Aggregate data, considering the spatial granularity of the study area; (3) Define thematic layers with data properties and spatial representation of interest related to the phenomenon; and (4) Provide a friendly web interface where the user can define the influence of each thematic layer through MCDM methods to create a thematic choropleth map with the spatial pattern of the phenomenon.

The current implementation presents a case study for improving decision-making on controlling epidemics and infectious and neglected diseases in an urban environment in Brazil. The insights obtained allow the decision maker to understand better both the data inputs and the different possible outcomes of the multi-criteria decision-making model, considering different combinations of weights. The tool can be used in different contexts, areas, and data sources with different spatial and temporal aspects.

2 SYSTEM ARCHITECTURE

In general, decision-making carried out by managers for addressing and controlling epidemics and diseases [13] requires a combination of information, such as estimates of cases, deaths, and their concentrations in city regions, among others.

These data sources have two main obstacles [10]: (i) Data access, as some of them are confidential, sensitive, or lack professionals to collect and pre-process them before publication; and (ii) Consistency among different variables in terms of temporal and spatial granularity. When data are available, they usually correspond to reports of relatively large periods and regions, with inadequate data representation for forecasting the occurrence or potential impacts of epidemics.

The MultiMaps tool is a proposal that aims to overcome these barriers. It can handle open access data anonymously, use different sources provided by various organizations, and work with different spatial and temporal granularities. Finally, the accuracy and reliability of predictions can be improved as data sources and their sizes increase.

Figure 1 presents the flow of data ingestion from different sources, the geolocation of these data, the data aggregation considering the spatial granularity of the study area, and the definition of thematic layers with the properties of the data with the spatial representation of interest related to the phenomenon under analysis.

After data ingestion, aggregation, and definition of thematic layers with their respective properties, the user, via a web interface, uses an MCDM technique to define the influence of each thematic

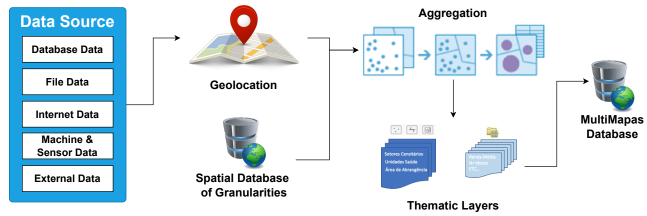

Figure 1: Data acquisition, aggregation, and creation of thematic layers.

layer. Via map algebra, a weighted combination of each thematic layer with its respective weight is made, producing a final choropleth map with the spatial pattern of the phenomenon under study, as shown in Figure 2.

MultiMaps's interface web was developed using Python as the programming language with the Django framework in version 3.1.7. The data are stored in the PostgreSQL 12 database with the PostGIS extension. MultiMaps also uses MongoDB and Redis to cache data and thus improve usability. The frontend is created using a combination of JavaScript, HTML and CSS. The interactive maps are built using Leaflet version 1.8.0.

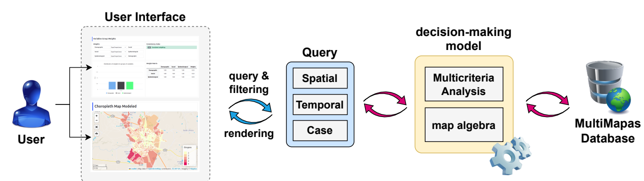

Figure 2: Visual inspection engine overview.

We considered both tools/methods from spatial statistics [4, 7, 17] and MCDM [8, 19] to develop the MultiMaps tool. For a thorough review of the AHP method [16], we refer the reader to the works by [3, 11]. For examples of uses of AHP and MCDM for the healthcare sector, we refer the reader to the works by [6, 18].

Concepts of spatial and spatio-temporal proximity are closely linked to the transmission of infectious diseases, as the transmission is more likely if individuals at risk are close in a spatial and temporal sense. Regarding non-communicable diseases, the proximity to environmental risk factors can be vital to analyze. Epidemiological analyses, therefore, must take into account space and time, having as a fundamental principle to examine the dependence between observations in relation to these two dimensions [15].

MultiMaps uses the global Moran index (Moran's I) [9, 14] to verify whether aggregations performed at different spatial granularities represent a phenomenon from a spatial point of view. The global Moran's I is bounded between -1 and +1, where -1 is strong negative SA (dispersion), 0 is random, and +1 is strong positive SA (clustering of data values) [12].

Once the global dependency is verified, MultiMaps also calculates the Local Indicators of Spatial Association (LISA) [2]. LISA is a decomposition of global Moran's I, used to identify cases in which the value of an observation and the average of its surroundings is either more similar (high-high or low-low) or dissimilar (high-low or low-high) than we would expect from pure chance.

3 CASE STUDY: EVALUATING MULTIPLE EPIDEMICS IN AN URBAN ENVIRONMENT

The study of the geographical distribution of diseases and their relationships with socio-environmental factors constitutes the object of what is called Spatial Epidemiology. Understanding and mapping the risks and difficulties of coping with diseases considering these socio-environmental factors is complex and involves variables with different characteristics and from diffuse sources [1].

One of the issues for decision-making in the health area is the consistent integration of information from different sources [10]. For example, Health Surveillance deals with data from patients at different levels of care by health units, transmission vectors, socio-economic aspects, multimorbidities (which can generate diagnostic confusion and aggravations), as well as the resources of the Health System distributed in the regions of the city. Frequently, to better represent epidemics, information must be aggregated according to a specific spatial granularity, which may differ between regions of social and epidemiological data and even between data from different epidemics. Time granularity is also essential, as epidemiological cycles can differ in addition to seasonal aspects [1].

This study case illustrates the four components of the Multimaps tool. In order to synthesize the available variables, a group of multidisciplinary specialists was consulted, with the presence of medical professionals, nurses, epidemiologists, and statisticians, among others. This was essential to define a set of possible input variables that a modeling system would need to perform the mapping of critical regions in relation to epidemics.

Based on interviews with a multidisciplinary group of specialists and data availability, a set of 8 variables was mapped and grouped into 3 groups according to demographic, social, and epidemiological aspects. Three variables encompass demographic aspects. One variable considers criteria of social vulnerability. Four epidemiological variables contain the count of confirmed cases of three epidemics with different characteristics and contagion (dengue, COVID-19, and tuberculosis) and the influence of the distribution of health units.

The case study was carried out in the municipality of São Carlos (Figure 3), a medium-sized city in the interior of the state of São Paulo, Southeast Brazil, located in the central-eastern region of the state. According to the 2010 census, it has a total territorial area of 1,136,907 km^2, a population density of 195.15 inhabitants/km², a population of 221,950 inhabitants, a Gini index of 0.63, human development index (HDI) of 0.805, and gross domestic product of R$ 6,712,498.00. The spatial granularity adopted was the census sectors [1] of the urban area of the municipality.

After defining the hierarchical contribution of each variable through the AHP method (Figure 4), the user generates a choropleth map, as illustrated in Figure 5. The user can then compare his model map (map to the left of Figure 5) with an expert's view (map to the right of Figure 5), in addition to viewing all maps used as input.

[1]The census sector is the territorial unit established for cadastral control purposes, formed by a continuous area, located in a single urban framework or rural, with a size and number of households that allow a survey by a census taker.

Figure 3: Location of the municipality and urban perimeter of São Carlos – SP.

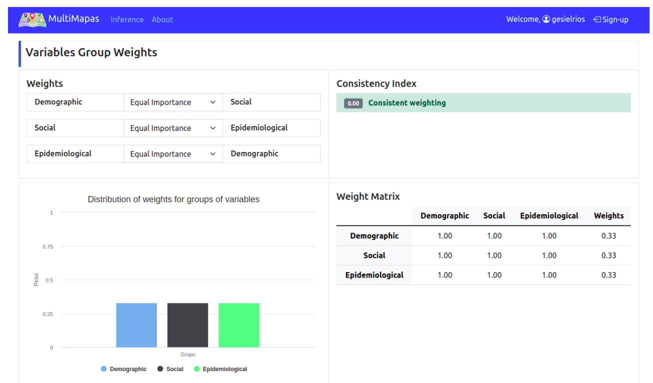

Figure 4: Hierarchical contribution of each variable.

Figure 6 shows the heatmap with the health care cases, allowing a visual inspection of the areas that concentrate the most significant number of cases, filtering the cases by notification date and week epidemiological. The heatmap is generated from the Gaussian kernel density estimation with a radius of influence of 10 meters. Figure 7 shows the evolution of cases over time in an accumulated manner or with sliding windows.

4 CONCLUSIONS AND FUTURE WORKS

Decision-makers require methods and interfaces that allow them to use heterogeneous data at different scales of time and space, visually compare the results and also require a dashboard summarizing the main estimates found. In this context, this work presents Multimaps, an open-source tool that consistently integrates data from different sources and contexts, optimizing and facilitating the analysis, management, or representation of space and the phenomena that occur in it by professionals from different areas.

The tool uses multi-criteria analysis techniques to define data influence and map algebra to combine these data and build thematic maps from multiple sources. The current implementation presents a case study for analyzing and controlling multiple epidemics and infectious and neglected diseases in an urban environment in Brazil. The insights obtained allow the decision-maker to understand better the data inputs and the different possible outcomes of MCDM model, considering different combinations of weights. It is vital to observe

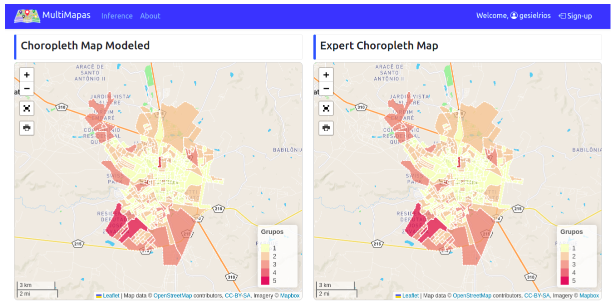

Figure 5: Choroplethic map modeled by the user and by the healthcare expert.

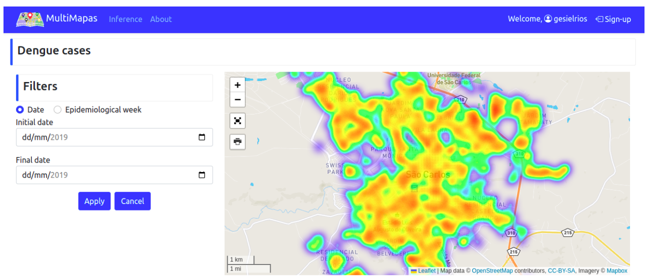

Figure 6: Heatmap with healthcare cases.

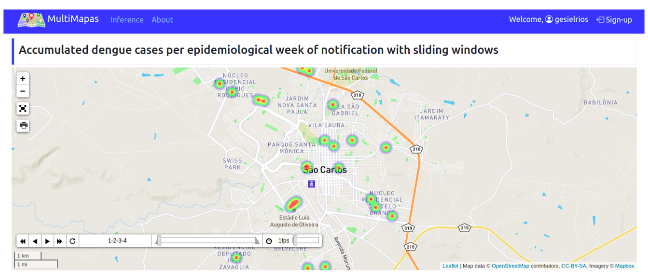

Figure 7: Heatmap with time with healthcare cases.

that the tool can be used for complex problems in different areas, with different MCDM models and inputs.

Future works will include: (i) inserting additional models for weighting the variables; (ii) adding the possibility of using unsupervised and supervised learning models to improve the quality of the final maps; (iii) improving the feedback loops, allowing the user to compare the different versions of the maps generated on the same screen; and (iv) application to other healthcare sector issues.

ACKNOWLEDGMENTS

The authors of this work would like to thank the Center for Artificial Intelligence (C4A-IUSP) and the support from the São Paulo Research Foundation (FAPESP grant 2019/07665-4, 2020/16578-5) and from the IBM Corporation. We would like also to thank the Artificial Intelligence group of the Institute of Mathematics and Computer Science of the University of São Paulo (CEMEAI); the support of José Galizia Tundisi from the Science, Technology and Economic Development Department from São Carlos Municipality.

REFERENCES

[1] Ana Lúcia Andrade, Antônio Miguel Vieira Monteiro, Christovam Barcellos, Eugênio Lisboa, Lisiane Morelia Weide Acosta, Maria Cristina de Mattos Almeida, Maria Regina Varnieri Brito, Marília Sá Carvalho, Marina Atanaka dos Santos, Oswaldo Cruz, et al. 2007. Introdução à estatística espacial para a saúde pública. (2007).

[2] Luc Anselin. 1995. Local indicators of spatial association—LISA. *Geographical analysis* 27, 2 (1995), 93–115.

[3] Rosaria de FSM Russo and Roberto Camanho. 2015. Criteria in AHP: a systematic review of literature. *Procedia Computer Science* 55 (2015), 1123–1132.

[4] Peter J Diggle and Emanuele Giorgi. 2019. *Model-based geostatistics for global public health: methods and applications.* Chapman and Hall/CRC.

[5] Patricia Silva Ferreira and Charlei Aparecido da Silva. 2020. O método AHP e a Álgebra de Mapas para determinar a fragilidade ambiental da bacia hidrográfica do Rio Brilhante (Mato Grosso do Sul/Brasil), proposições para a gestão do território. *Confins. Revue franco-brésilienne de géographie/Revista franco-brasilera de geografia* 46 (2020).

[6] Zekun Gao, Yutong Jiang, Junyu He, Jiaping Wu, Jian Xu, and George Christakos. 2022. An AHP-based regional COVID-19 vulnerability model and its application in China. *Modeling earth systems and environment* 8, 2 (2022), 2525–2538.

[7] GBM Heuvelink and FM van Egmond. 2010. Space–time geostatistics for precision agriculture: a case study of NDVI mapping for a Dutch potato field. In *Geostatistical applications for precision agriculture.* Springer, 117–137.

[8] Abhishek Kumar, Bikash Sah, Arvind R Singh, Yan Deng, Xiangning He, Praveen Kumar, and RC Bansal. 2017. A review of multi criteria decision making (MCDM) towards sustainable renewable energy development. *Renewable and Sustainable Energy Reviews* 69 (2017), 596–609.

[9] Hongfei Li, Catherine A Calder, and Noel Cressie. 2007. Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. *Geographical analysis* 39, 4 (2007), 357–375.

[10] Weiqiao Liu and Jianjun Zhu. 2021. A multistage decision-making method for multi-source information with Shapley optimization based on normal cloud models. *Applied Soft Computing* 111 (2021), 107716.

[11] Yan Liu, Claudia M Eckert, and Christopher Earl. 2020. A review of fuzzy AHP methods for decision-making with subjective judgements. *Expert Systems with Applications* 161 (2020), 113738.

[12] Antônio Miguel Vieira Monteiro, G Câmara, MS Carvalho, and S Druck. 2004. Análise espacial de dados geográficos. *Brasília: Embrapa* (2004).

[13] RONEI M MORAES, JORDANA A NOGUEIRA, and ANA CA SOUSA. 2014. A new architecture for a spatio-temporal decision support system for epidemiological purposes. In *Decision Making and Soft Computing: Proceedings of the 11th International FLINS Conference.* World Scientific, 17–23.

[14] Patrick AP Moran. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37, 1/2 (1950), 17–23.

[15] Dirk U Pfeiffer, Timothy P Robinson, Mark Stevenson, Kim B Stevens, David J Rogers, and Archie CA Clements. 2008. Spatial analysis in epidemiology. (2008).

[16] Thomas L Saaty. 1978. Modeling unstructured decision problems—the theory of analytical hierarchies. *Mathematics and computers in simulation* 20, 3 (1978), 147–158.

[17] Richard Webster and Margaret A Oliver. 2007. *Geostatistics for environmental scientists.* John Wiley & Sons.

[18] Ming Yang, Li Jia, Wanyu Xie, and Tilei Gao. 2021. Research on risk assessment model of epidemic diseases in a certain region based on markov chain and AHP. *IEEE Access* 9 (2021), 75826–75839.

[19] Edmundas Kazimieras Zavadskas, Zenonas Turskis, and Simona Kildienė. 2014. State of art surveys of overviews on MCDM/MADM methods. *Technological and economic development of economy* 20, 1 (2014), 165–179.