

Enhancing Geospatial Business Intelligence capabilities using Multi-Criteria Decision Making



Mohamed Hanine, Omar Boutkhoul, Abdessadek Tikniouine, Tarik Agouti
Cadi Ayyad University, Marrakesh
Morocco
m.hanine@uca.ma {o.boutkhoul@uca.ma} {tikniouine@uca.ma}, t.agouti@uca.ma

ABSTRACT: *The Geospatial Business Intelligence (Geo-BI) is a collection of technologies that combines Geographic Information System (GIS) and Business Intelligence (BI). Geo-BI provides an efficient tool for browsing and storing very large spatio-temporal databases. However, the process for evaluation and selection of the optimal decision in Geo-BI is very complex. Hence, the Multi-Criteria Decision Making (MCDM) methods are used as a tool to facilitate the decision making process. In this paper, we present a methodology called Multi-Criteria Geospatial Business Intelligence (MC-GeoBI) which aims to enhance Geo-BI applications by applying MCDM. The latter obviously offers several advantages in terms of decision making when one must take into account the multiple and contradictory aspects of the criteria. To illustrate this, an application model has been developed to select the most appropriate site for landfill of industrial wastes.*

Keywords: Multi-Criteria Geospatial Business Intelligence (MC-GeoBI), Geospatial Business Intelligence (Geo-BI), Multi-Criteria Decision Making (MCDM), Geographic Information System (GIS), Business Intelligence (BI), On-Line Analytical Processing (OLAP)

Received: 17 June 2015, Revised 18 July 2015, Accepted 24 July 2015

© 2015 DLINE. All Rights Reserved

1. Introduction

Business Intelligence (BI) is a technology dedicated to analyze business data and processes with the aim to obtain the appropriate decision of interest to keep them under control, and to realize their immediate display (dynamic dashboards, reports, historical systems, etc...). Business Intelligence applications are based on architecture of software that is usually composed of the following:

- An ETL (Extract, Transform and Load) tool that allows extracting data from different heterogeneous sources (databases, XML or flat files, etc...), transform (integration, data structure ...) and load the data in a data warehouse.
- A data warehouse which stores the organization's historical data for analysis purposes.
- An On line Analytical Processing (OLAP) server which enables the rapid and flexible exploration and analysis of large amounts of data stored in the data warehouse [1].

- Decision makers and managers that use many tools such as reporting tools, dashboards and different OLAP clients.

Additionally it is commonly recognized that about 80% of data has a spatial component [2], and that working with massive quantities of data is not always easy to represent. In this situation, we can use the spatial component to enhance the BI user experience with spatial analysis tools and map displays, such as Geographical Information Systems (GIS) to provide a better support for the analysis and decision making process [3]. Recently, Business Intelligence (BI) and GIS have become closely integrated and merging both tools provides an efficient browsing and storing of very large spatio-temporal databases and a powerful visualization system for incorporating data [3].

In this regard, we define Geospatial Business Intelligence (Geo-BI) an innovative powerful methodology that allows support of traditional and spatial data through merging GIS and BI technologies [3]. The Geo-BI approach adds to the business information that is extracted from data warehouse, the territorial information concerning the positioning of the quoted sizes within dashboards.

Despite these Geo-BI advantages, the decision making process can be a challenging task in the location classification problems [4]. However, to enable investors to explore, compare, and rank candidate locations against this rich data source, a Multi-Criteria Decision Making (MCDM) methodology is applied for the enhancement of Geo-BI process.

The main objective of the research is to present a Multi-Criteria Geospatial Business Intelligence (MC-GeoBI) approach to make better decisions and to enhance Geo-BI model by applying MCDM. The proposed approach has many criteria and alternatives that affect the decision making processes. We can define MC-GeoBI as a set of models, descriptions and analysis methodologies that exploit the availability of traditional and spatial data to retrieve information, knowledge and advice useful in supporting complex decision making processes according to many spatial and business-related factors and alternatives. This approach utilizes GIS application and employs techniques of BI such as data warehousing and OLAP in order to narrow a vast list of available alternatives. Subsequently, an AHP (Analytical Hierarchy Process) as a method of MCDM is used to rank in order the candidate alternatives for selecting the most appropriate ones. Finally, an example is presented to illustrate how this methodology was applied to solve a site selection problem for landfill of industrial wastes in region of Casablanca in Morocco.

This paper is composed of five sections. Besides the introduction, the second section gives a brief overview of the related works in Geo-BI and MCDM. Then a step-by-step MC-GeoBI methodology is presented in the third section. As regards the next section, an empirical study is illustrated for the application of the prototype for selecting an appropriate site for landfill of industrial wastes. Finally, concluding remarks and discussions follow.

2. Related Works

2.1 Geospatial Business Intelligence (Geo-BI)

The architecture of Business Intelligence (BI) containing many technologies such as Data Warehousing, Data Mining, data integration, OLAP etc. have been proposed in many fields [5-9]. It is used to establish decision support that prioritizes efficient querying of huge amounts of historical, clean and consolidated data [10]. However, a spatial perspective is needed for strategic analysis: the integration of spatial component in BI is called Geospatial Business Intelligence (Geo-BI) in order to reveal relationships based on the geographical positions associated with the information.

The main aim of this section is to review the use of BI technology (such as data warehouse or/and OLAP) with GIS in two ways, either by integrating spatial and warehousing information in a single framework or treating tools separately. There are previous studies in this area. Some approaches and research tools form a good basis for a multidimensional spatio-temporal decision support tool [11]. The Geomatics Center of Laval University in Canada provides Spatial OLAP (SOLAP) as a tool for support rapid and easy spatio-temporal analysis, exploration of multidimensional data and offers a set of visualization techniques such as maps, tables and diagrams [12-14]. The Informatics Center of the Federal Pernambuco University in Brazil has proposed GOLAPA (Geographic Online Analytic Processing Architecture) providing users with an abstraction of the complexity involved in querying both analytic and geographic data for decision support [15,16]. Among other tools that combine BI with GIS, we cite CubeView and Polaris proposed by Voss [17] to facilitate the observation of spatial patterns and temporal trends in large volumes of data. Furthermore, in [18], the authors have developed a robust, easily accessible and user-friendly Geo-BI solution for the local governance of infrastructure services. Moreover, in [19], the authors proposes an extensible GIS-OLAP integrated solution supporting geographical dimensions and measures, based on a formal multidimensional model named GeoCube. More

	Interesting patterns/ Knowledge Discovery	Large Complex Data sets	Multidimensional view/ Navigation	Statistical Analysis	Spatial Presentation	Visual Charts	Spatial Analysis	Numerical Spatial Problem Solving
Statistical software				X		X		
Data Mining	X			X		X		
GIS Software					X		X	
Relational DB		X						
OLAP		X	X			X		
OLAPGIS	X	X	X	X	X	X	X	X

Table 1. Traditional BI technology versus an OLAP-GIS system (source: [20])

than that, development of OLAP- GIS tool [20] was proposed by the University of Pittsburgh for numerical-spatial problem solving in community health assessment analysis. Table 1 summarizes the benefits of many technologies such as OLAP and GIS, and the added value in the combination of both tools proposed in the university study [20].

With regard to the researches discussed above, we shall now conclude that the approaches proposed are limited in the use of data warehouse models, as these models are composed solely of quantitative data, while the necessary data for decision-making in complex problems are often quantitative and qualitative in type. This causes a difficulty in assessing the quality of geospatial information. Therefore, it is normal to take into account several types of heterogeneous data.

2.2 Multi-Criteria Decision Making MCDM

In this section, we review the MCDM and choose the most suitable method that can be applied to the selection of the potential alternative resulting from decision-makers' preferences. The MCDM is a branch of operation research models and a well known field of decision support system (DSS) [21]. The state of the art is very rich by various multi-criteria decision making methods suggested in many studies for the selection problem or classification of which the most important are Analytic Hierarchy Process (AHP), Preference Ranking Organization METHod for Enrichment (PROMETHEE), ELimination and Choice Expressing REality (Electre), Technique for Order Preference by Similarity to Ideal Solution (Topsis), Analytical Network Processing (ANP), and Multi-Attribute Utility Theory (Maut), etc...

In a previous analysis, Forman and Gass [22] argue that AHP was the most suitable methodology to solve problems associated with different perspectives and location decision-making process problems. AHP has been successfully applied in all aspects of Decision Science processes and applications: selection, evaluation, allocations, planning and development, priority and ranking, and decision-making [23].

The AHP method that was developed by Saaty (1980) [24, 25] is a tool for multi-criteria decision-making. This method allows managers to structure complex problems they face in issuing judgments based on their experience and available data. The AHP method allows us to decompose a problem, following the arborescence of different criteria and sub-criteria decisions associated with this problem and to compare these criteria together, in pairs, using a weighting scale (Table 2), and the best alternative can be selected according to the highest rank among alternatives.

2.3 The Proposed Methodology for MC-GeoBI

The methodology MC-GeoBI system aims to evaluate and select a recommended decision to satisfy the sufficient requirement

Relative importance	Definition
1	Equal importance
3	Weak importance
5	Strong importance
7	Demonstrated importance over the other
9	Absolute importance

Table 2 . Scale of pair-wise comparison of AHP.

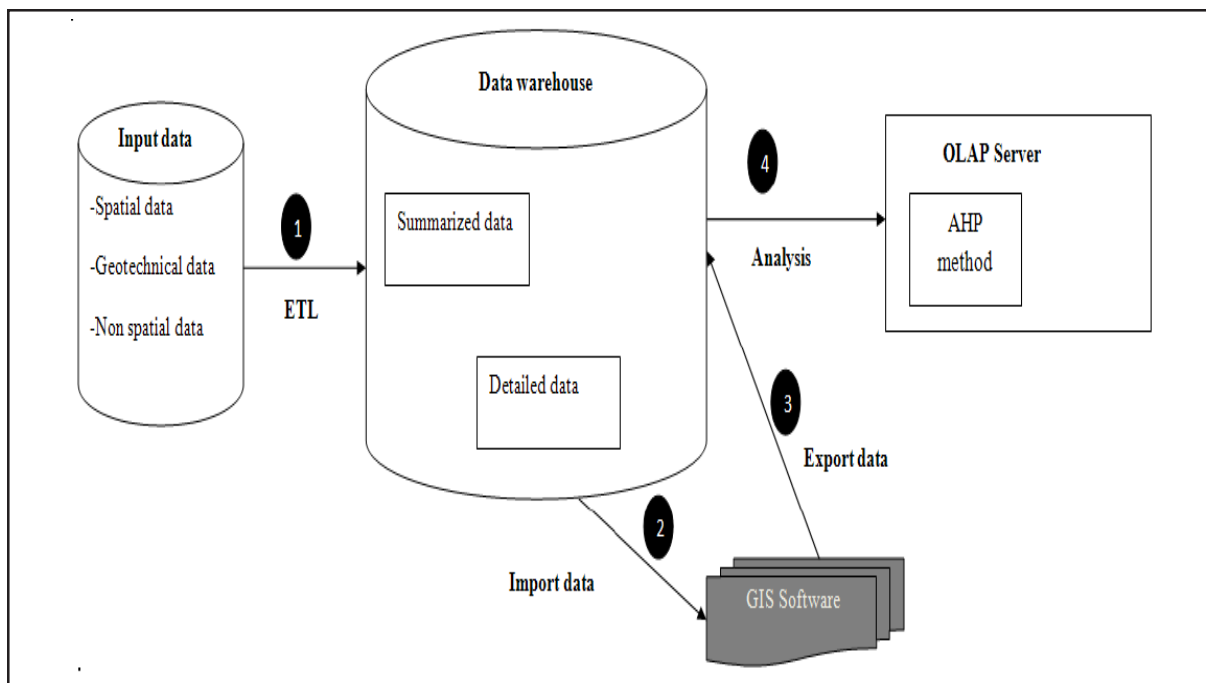


Figure 1. Conceptual model of the proposed MC-GeoBI

of decision makers. For example, classification or/and researching a site among several available sites. This method is shown in Figure 1. The following steps, as highlighted in black circles explain how it works.

1) In this step, the decision maker uses an ETL (Extract, Transform and Load) framework to extract data from a source (i.e., the data provided by the user or by existing data files used in transactional database systems). The source data is converted into target data, this latter will be stored in the data warehouse schemas. The objective of this conversion is to delete duplicate data and convert different data formats.

2) The data stored in data warehouse is imported to GIS software and spatial analysis is performed. The spatial analysis (buffer analysis) reduces the list of available sites to a few candidate sites that validate a set of user-defined spatial criteria.

3) The GIS software exports the list of candidate sites into data warehouse.

4) The data is imported in the form of an excel spreadsheet from data warehouse into a JAVA application to perform OLAP analysis using the AHP method in order to rank the candidate sites according to business related factors.

4. Application

In order to test the validity of the proposed approach, we develop a prototype to select the most appropriate site for the implementation of a landfill of industrial wastes in the region of Casablanca in Morocco. The main goal in this case study is selecting a recommended solution to satisfy the sufficient requirements of decision makers by using MC-GeoBI methodology. The site selection process development for landfill of industrial wastes is a complex decision because it depends on a number of factors that can be categorized as spatial (such as proximity to residential areas, geographical and geotechnical characteristics of sites, etc.) and business-related factors (such as size of site, land owning, proximity to sources of wastes, etc.). The analysis of the list of the spatial and business related factors has significant impacts on the process of site selection for landfill of industrial wastes. As explained earlier, spatial factors are used to select the feasible sites from a vast list of available sites. Then the business-related factors are used in rank order of the candidates' sites from which the most appropriate site is selected [26]. The results demonstrate the capabilities of the system in a real-life but hypothetical application scenario. This process of application is described and discussed in the following steps:

4.1 Data Input

Firstly, the prototype applied in this research requires a compilation of spatial and non spatial data that are imported into the system.

- Spatial data, like environmental and spatial data in Casablanca region (.dbf format) showing Governorates layer, Rivers layer, big roads layer, zone of housing layer, Borders layer and water table elevation layer. All of these layers are called base map layers.
- Non spatial data, like wastes indicators, industrial zone characteristics, etc ...

Accordingly, Pentaho Data Integration as open-source ETL package was adopted for migrating data.

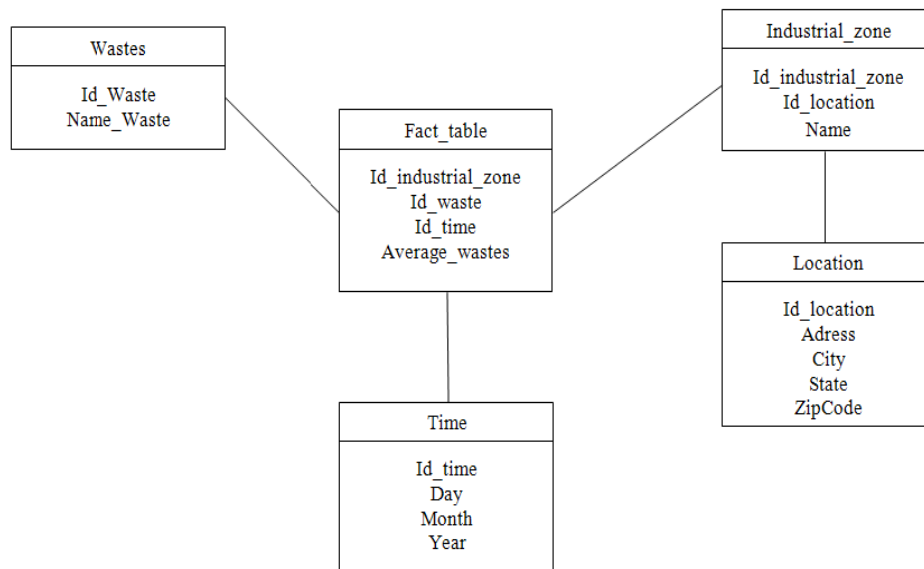


Figure 2. Snowflake schema for site selection

4.2 Implementation of Data Warehouse and OLAP

There are many approaches for designing a data warehouse: Star schema, Snowflake schema or Hybrid schema. The snowflake schema is adopted in this research for rapid and easy migration of data from transactional database to data warehouse [27]. Figure 2 illustrates the snowflake proposed for elaborating the data warehouse in this research. In this schema, there is one fact table, three dimension tables and one sub-dimension table.

Several software choices are available to build a data warehouse such as MS SQL server, Oracle, Postgre SQL or any other database management software supporting the multidimensional concept. Postgre SQL is selected in this study to provide a fast, stable and standards compliant database with an excellent geospatial extension (PostGIS).

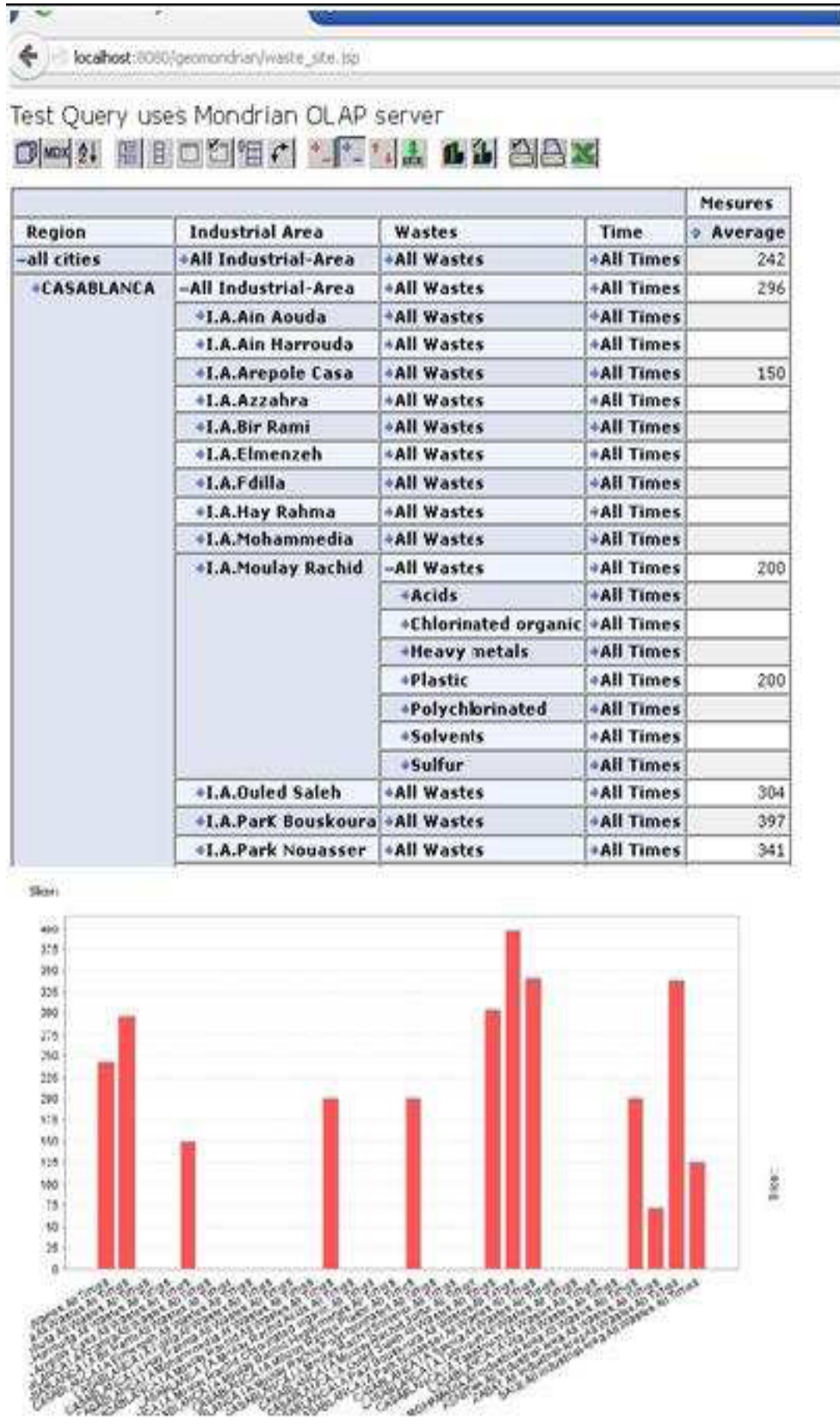


Figure 3. Visual interface of the Mondrian server.

It is important to note that there are many OLAP commercial software packages such as SAS OLAP Server or Microsoft Analysis Services that are also available in the market. Moreover, our intention is not to develop a commercial software product, we utilize rather a Mondrian Server OLAP (open source software). The advantage of this server is that the users can use different graphical wizards to easily write their queries and display results in different graphical formats. An example of this application applied in our case study is shown in Figure 3 which allows us to answer very complex queries, for example: “*What is the average value for each type of waste for each industrial area and for Casablanca city in any time?*”.

4.3 Spatial Analysis

The spatial factors require geographical analysis of the available sites. Therefore, it was decided to use commercial GIS software. In this case, MapInfo was the selected GIS software for this research and the elaborate description of GIS techniques and its use through MapInfo can be found in the referenced source [28].

Moreover, the system was run for exported data to GIS software. After performing the spatial analysis, the list of available sites was narrowed to five candidate sites that satisfied the following criteria: (Figure 4 illustrates the location of these five candidate sites along with other details)

- The soils with good aptitude for agriculture are excluded.
- The coast with a protection zone of 5 km is excluded.
- 2 km around residential areas are excluded.
- 1 km around rivers, streams, lagoons and lakes are excluded.

The spatial analysis concerns a:

- Buffers Analysis to narrow the list of sites according to defined criteria.
- Themes to separate layers of map data and descriptive information. For example, in a region’s database, different themes could be roads, industrial zones, rivers, etc

It is important to note that there are many interfaces designed and written with Java to import and export data for the GIS software and the Data warehouse.

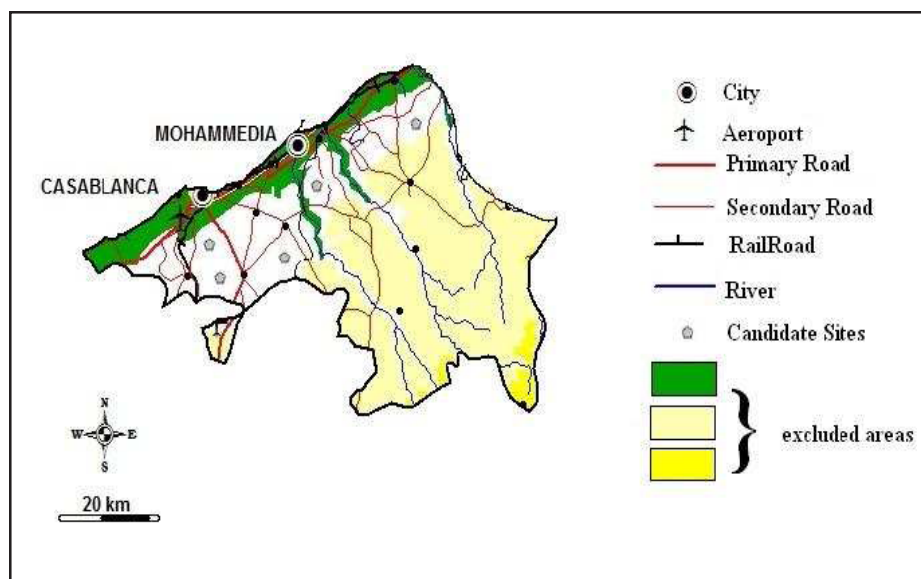


Figure 4. Map of candidate sites

4.4 OLAP Analysis:

In the previous step, we got five sites although they are selected in the same category. These have different characteristics. The choice and classification of a site from the best to the less favorable is performed here by OLAP analysis using Analytical Hierarchy Process method (AHP). The AHP method and its implementation in software developed by JAVA are described in the following steps:

- 1) We initially develop a hierarchy model of the problem based on the criteria and alternatives. As shown in Figure 5, the top level and the lowest level of the hierarchy denote the overall objective (selecting the most appropriate site for landfill of industrial wastes) and the proposed sites candidates (Site1, Site2, Site3, Site4, Site5) respectively. The four main criteria, namely size of land, proximity to wastes sources, land owning, and proximity to road network were included in the second level.
- 2) Secondly, the decision makers performed the pair-wise comparisons between all criteria the matrix is obtained in Figure 6. The comparisons are constructed by using Saaty’s scale (table 2). For each factor (e.g., size of land), a matrix is developed to compare the given sites for this factor as shown in Figure 7. Then these matrices are normalized and checked for consistency.
- 3) The weight of each candidate site is calculated according the priorities of these matrices previously. Figure 8 shows the classification of the candidate sites.

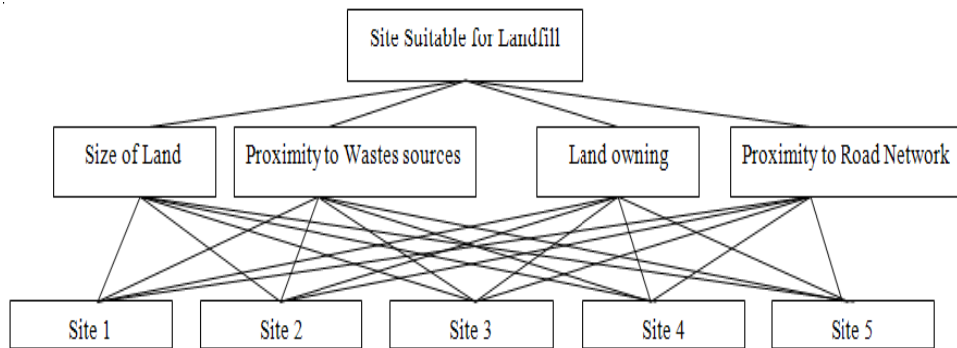


Figure 5. The hierarchy model of a suitable site selection

The recommended solution is the most optimal and suitable choice to satisfy spatial, business related factors and alternatives. At this study case, the recommended site is Site 3.

This demonstration is clear and shows that the methodology MC-GeoBI can aid decision makers in selecting the most appropriate site for landfill of industrial wastes. The land selection business related factors and spatial criteria in the approach can be customized according to user requirements. Thus, the system could be applied for any other purposes where the problem is the classification of location.

	Size of Land	Proximity to wastes	Land owning	Proximity to road	Weight
Size of Land	1	3	3	5	0.45
Proximity to wastes sources	1/3	1	5	3	0.35
Land owning	1/3	1/5	1	1	0.10
Proximity to road	1/5	1/3	1	1	0.10

Figure 6. AHP criteria matrix

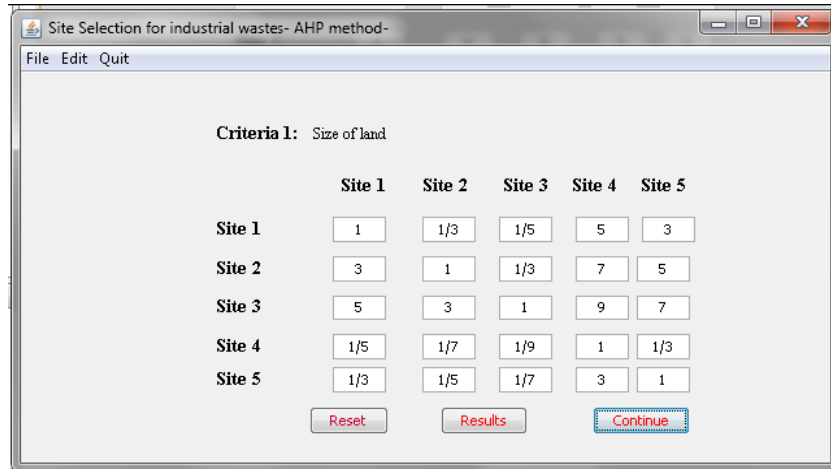


Figure 7. Comparison matrix candidate site of each factor (i.e. Size of land)

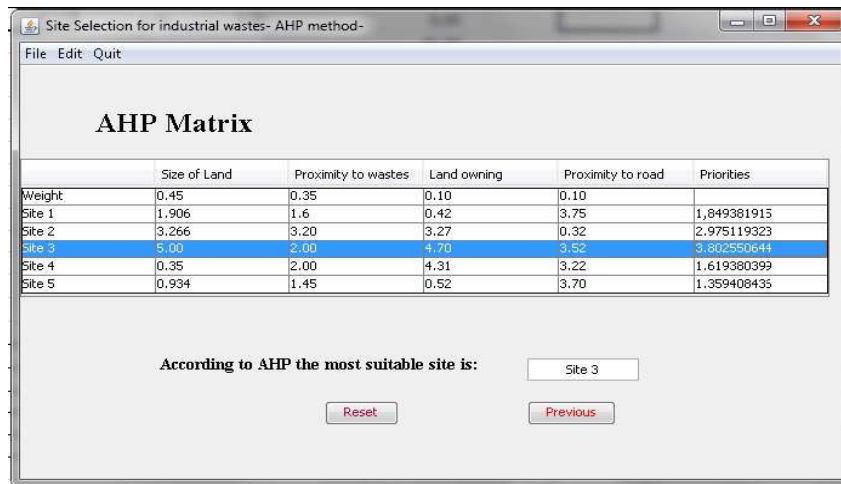


Figure 8. Final AHP analysis indicating the most appropriate site for landfill of industrial wastes

5. Conclusion

This paper proposes MC-GeoBI approach, which uses Multi-Criteria Decision Making (MCDM) to enhance Geospatial Business Intelligence (Geo-BI). The methodology proposed provides a wider context for decision making process to decision makers as it aggregates the following: a ETL for migrating data to a specific data model for decision support (Data warehouse), the functionalities of the visualization and spatial analysis (GIS), and OLAP for analyzing a great volume of data by using AHP method for ranking in order the candidate alternatives. For illustration, a relevant case study concerns a localization of sites for the landfill of industrial wastes was developed using MC-GeoBI methodology. We believe that the use of MCDM methods with Geo-BI system could be an efficient tool to deal with spatial situations. In the future we hope to increase the effectiveness of MC-GeoBI approach by developing a single framework integrating the features of Geo-BI with other MCDM methods.

References

- [1] Badard, T., Dubé, E. (2009). Enabling geospatial Business Intelligence. Special issue on Business Intelligence for the OSBR.
- [2] Franklin, C. (1992). An Introduction to Geographic Information Systems: Linking Maps to Databases. Database, April, p. 13-21.
- [3] Angelaccio, M., Basili, A., Buttarazzi, B., Liguori, W. (2012). Using geo-business intelligence to improve quality of life. IEEE

1st AESS European Conference on Satellite Telecommunications, Rome, 02–05 October.

- [4] Weber, P., Chapman, D. (2011). Location Intelligence: An Innovative Approach to Business Location Decision-making. *Transactions in GIS*, 15 (3) 309–328.
- [5] Chaudhuri, S., Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26 (1) 65-74.
- [6] Inmon, W. H. (2005). *Building the data warehouse* (4th ed.). New York: John Willey and Sons.
- [7] Kimball, R., Ross, M. (2002). *Data Warehouse Toolkit* (2nd ed.). New York: John Willey and Sons.
- [8] Song, I. (2009). Data Warehouse. Liu L., Özsu M. T. (Ed.), *Encyclopedia of Database Systems* (1, p. 2731- 2735). Berlin: Springer.
- [9] Abelló, A., Romero, O. (2009). On-Line Analytical Processing. Liu L., Özsu M. T. (Ed.), *Encyclopedia of Database Systems* (20, p. 2731-2735). Berlin: Springer.
- [10] Ferraz, V.R.T., Santos, M.T.P. (2010). GlobeOLAP: Improving the geospatial realism in multidimensional analysis environment. *Proceedings of the 12th International Conference on Enterprise Information Systems*, 5, HCI, Funchal, Madeira, Portugal, June 8 - 12.
- [11] Shekhar, S., Lu, C.T., Tan, X., Chawla, S., Vatsavai, R. (2001). Map cube: A visualization tool for spatial data warehouse. As Chapter of *Geographic Data Mining and Knowledge Discovery*. Harvey J. Miller and Jiawei Han (eds.), Taylor and Francis.
- [12] Rivest, S., Bédard, Y., Marchand, P. (1997). Towards better support for spatial decision-making: Defining the characteristics of Spatial On-Line Analytical Processing (SOLAP). *Geomatica, The Journal of the Canadian Institute of Geomatics*, 55 (4) 539-555
- [13] Rivest, S., Bédard, Y., Proulx, M.J., Nadeau, M. (2003). SOLAP: a new type of user interface to support spatio-temporal multidimensional data exploration and analysis. *International Society of Photogrammetry and Remote Sensing Joint Workshop of WG II/5, II/6, IV/1 and IV/2 on Spatial, Temporal and Multi- Dimensional Data Modelling and Analysis*, Québec, CANADA, October 2-3.
- [14] Rivest, S., Bédard, Y., Proulx, M.-J., Nadeau, M., Hubert, F., Pastor, J. (2005). SOLAP technology: Merging Business Intelligence with Geospatial Technology for Interactive Spatio-Temporal Exploration and Analysis of Data. *Journal of International Society for Photogrammetry and Remote Sensing (ISPRS) “Advances in spatio-temporal analysis and representation*, 60(1) p. 17-33.
- [15] Fidalgo, R., Silva, J., Times, V., Sousa, F., Barros, R. (2003). GMLA: A XML Schema for Integration and Exchange of Multidimensional Geographical Data. *GEOINFO In: Proceedings of the 5th Annual Brazilian Symposium on GeoInformatics*.
- [16] Fidalgo R. N., Times, V. C., Silva, J., Souza F.F., Salgado, A.C.(2004). Providing multidimensional and geographical integration based on a gdw and metamodels. *Brazilian Symposium on Databases (SBBD)*.
- [17] Voss, A., Hernandez, V., Voss, H., Scheider, S. (2004). Interactive Visual Exploration of Multidimensional Data: Requirements for CommonGIS with OLAP. *Database and Expert Systems Applications*, 15th International Workshop on (DEXA'04) , p. 883-887, California, IEEE.
- [18] Wickramasuriya, R., Jun Ma, Berryman, M., Perez, P.(2013). Using geospatial business intelligence to support regional infrastructure governance. *Knowledge-Based Systems*, 53. 80–89.
- [19] Bimonte, S., Tchounikine, A., Miquel, M. (2007). Spatial OLAP: Open Issues and a Web Based Prototype. 10th AGILE International Conference on Geographic Information Science Aalborg University, Denmark
- [20] Scotch, M. L. (2006). An OLAP-GIS system for numerical-spatial problem solving in community health assessment analysis. University of Pittsburgh.
- [21] Malczewski, J. (1999). *GIS and Multicriteria Decision Analysis*. Wiley, New York, NY, 392p.
- [22] Forman, E.H., Gass, S. I. (2001). The analytic hierarchy process: An exposition. *Operations Research* , 49. 469–86.

- [23] Vaidya, O., Kumar, S. (2006). Analytic hierarchy process: An overview of applications. *European Journal of Operational Research* 169. 1–29.
- [24] Saaty, T.L., Vargas, L. (2001). Models, Methods, Concepts and Applications of the Analytic Hierarchy Process. International Series in Operations Research and Management Science, Kluwer Academic Publishers.
- [25] Saaty, T.L. (2008). Decision Making With The Analytic Hierarchy Process. *Int. J. Services Sciences*, 1 (1) 83.
- [26] Irtishad, A., Salman, A., Pranas, L. (2004). Development of a decision support system using data warehousing to assist builders/developers in site selection. *Automation in Construction* 13. 525– 542.
- [27] Chau, K.W., Cao, Y., Anson, M., Zhang, J. (2003). Application of data warehouse and decision support system in construction management. *Automation in Construction* 12 (2) 213– 224.
- [28] Documentation of MapInfo software: <http://www.pbinsight.com/support/productdocumentation/details/mapinfo-professional>