PCA and Cluster Analysis for Criteria Mapping in Landfill Siting

Kaoutar BENNIS [#1], Lahcen BAHI [#2]
[#] Laboratory of Applied Geophysics, Geotechnical Engineering, Geology Engineering and Environment,
EMI, University Mohammed V Agdal, Rabat, Morocco
[1] benniskaoutar@gmail.com
[2] bahi@emi.ac.ma

Abstract— Landfill siting is of primordial interest in waste management. As the size of modern cities grows along with the shifts in demographics and composition of solid wastes, it has become important to choose the location of a waste disposal area in a way that insures long-term usability, environmental impact minimization and other considerations; these considerations can vary greatly among cases and from it follows that the criteria involved in siting a new landfill location should be subject to a conscientious and thorough choosing process.

This work is a literature review focusing on criteria used in the siting of new landfills, it took advantage of the statistical methods of data mining in order to establish patterns; it can be viewed as a guide. Starting with collection of scoop-compatible articles, an extraction of criteria from each article is done which opens the door for analysis by Principal Component Analysis (PCA), Bivariate Correlation (BC) and Cluster Analysis (CA). The corroborate use of this methods and the definition of the conceptual layers of category and supercategory fortifies the statistical significances of the results.

Keyword- Landfill, Siting, Criteria, Data Mining

I. Introduction

The landfill of waste always was one of the most used methods to put out residuals; this is true everywhere in the world [1] even if intensive efforts were directed toward recycling and recovery of solid wastes [2]. The secured landfill is the mode of waste disposal which guaranties to best account for environmental impacts, the public health problems and socio-economic issues associated with landfilling [3], it was therefore necessary in our days to choose adequately the site of landfill, a choice which have become an important task in waste management [4].

The process of choosing a site for landfill has always been complicated because of the implicated conflicts among parameters of choosing [5, 6], in fact, the selection of a new disposal site is a critical question in the urban planning because of the enormous impact on the economy, ecology and the health of environment [7], such a selection should resort to a maximum of information and insure availability in order to output acceptable results by the stakeholders [2]. The problem of the location of undesirable facilities was studied extensively in the literature; the mathematical models to solve this problem have evolved from monocriteria to multicriteria models capable of hosting conflictual criteria [8]. The methods of multicriteria decision analysis (MCDA) can, in such cases, be used, especially since they incorporate both quantitative and qualitative criteria [9]. According to [10], the selection of a site for waste disposal can be divided into two steps: the identification of potential sites through preliminary screening then the evaluation of these sites by evaluation of their adequacy on the basis of environmental impact assessment.

In evaluating a site as a possible location for public landfilling, many factors can be considered [11, 4].

Choosing the criteria is a complicated task, Al-Hanbali et al. [12] notes a lack in the detailed standards or specifications for the management of solid wastes along with specific criteria to select the appropriate location of the landfill, this lack in Jordan legislation was compensated by resorting to the criteria utilized by the international organization US Environmental Protection Agency (US EPA) another countries legislations which were derived from the priorities and local requirements. Charnpratheep et al. [10] resorted to the proximity of geographical objects, slope and height but recommended to conduct a complementary investigation to choose the criteria for selection of the site of the landfill, such as socio-economic and hydrogeological criteria. During the evaluation of a site as a possible location for landfilling, many factors can be considered [11]. For Al-Jarrah & Abu-Qdais [2], making a successful choice needs to guaranty overcoming many significant environmental and political obstacles; many factors should be considered in the selection of a site, their presentation can differ but the best is the one that the community can understand. Aragones-Beltran et al. [13] has identified the criteria on which the deciders base their knowledge, he's carried out a literature review in order to know which are the principal criteria cited by authors. Chang et al. [7] juged that many factors and criteria for the choice of the site should be closely organized and analysed, he defines the suitability criteria by emphasizing on the minimization of the potential risk for the health and admits having criteria implicated in the selection of landfill site, a literature review enabled him to identify the most important ones. De Feo & De Gisi [14] adopted ten (10) evaluation criteria to make the decision makers task easier and clearer, the authors considered that too many complicated criteria can result in an incomprehension and difficulty of comparing criteria. Effat & Hegazy [15] have developed a set of criteria by combining an intense literature review and expertise; these criteria were based on national and international laws relative to the site selection of a landfill. Geneletti [16] have adopted a stakeholders analysis in order to identify the criteria which should be meet in determining a new site, a list of stakeholders were established and each stakeholder participated by a liste of criteria he sees important for the study, the lists were regrouped in constraints and factors.

Even if many methods were suggested for the selection of indicators (and for criteria, by analogy), Puig et al. [17] speaks of two principal approaches : the top-down and bottom-up approaches; the top-down is based on identification of indicators from literature review and reduction of the number of indicators until a final set is agreed upon. The bottom-up approach consists of compiling the final set of indicators from the suggestions of stakeholders on the basis of their perception of issues and signification.

In this work, we will conduct a literature review which can serve as a starting point for anyone desiring to choose criteria to model its problem of site location, the ultimate goal of this study is not to select the criteria but rather to draw the state of the art of criteria chosen by the authors and exhibit the tendencies which can guide toward a more adapted and fine criteria choosing operation.

II. MATERIAL AND METHODS

A. Data Source

To choose the articles which will be used in the extraction of the most recurrent criteria in the problem of choice of a new public landfill site, we've opted for articles referenced in the database Scopus, in particular those of "Science direct" whose scope is close to that of our study. During the research, we used the tagwords "landfill" & "Siting", "Waste disposal" & "Siting".

The results were twenty-six (26) articles which were read and from which were extracted criteria adopted by the authors of an article. The (26) articles retained are those in direct relation with the subject of landfill siting, a greater amount of articles were read in which any articles suspected to be in liaison with the subject was included, the sub-set of definitive admitted articles forms therefore a restrictive selection favoring a more confined yet trusted work.

B. Criteria Census

Before analyzing the gathered articles, we will establish a summary report concerning these articles.

Summarily, the journals where the articles were published are: Waste Management (13 articles), Journal of environmental management (3), Journal of hazardous materials (1), Journal of environemental science & technology (1), Environmental geology (1), EnvironmentAsia (1), Journal of geographic information system (1), The Egyptian journal of remote sensing and space sciences (1), Waste management & research (1), Engineering geology (1), Applied Geography (1), Sustainable cities and society (1).

Between the (26) articles, 349 criteria (with redundancy) were evoked. In average 13,42 criteria were adopted per article, the minimal number was (5) for [18], the maximum number was (39) for [14]. The standard deviation (square root of the variance) is 7,28.

For a more detail presentation of the articles we've established a table (table 1) enlisting articles, their titles, year of publication, the publishing review and the criteria evoked in the article.

TABLE I
Articles used in the literature review

	Article	**Criteria**
1	Lin and Kao, 1998 [20]	Ground water protection areas | soil and geology | restricted zones | existing road network | land slope | population density | land ownership.
2	Leao et al., 2004 [18]	Soil type | Distance from water bodies and weltlands | Slope | Distance from urban areas | Distance from urban areas.
3	Kontos et al., 2005 [21]	Water permeability | Distance from water sources | Surface water | Sensitive ecosystems | Land cover | Urban areas | Cultural areas | Visibility | Land uses | Morphology | Wind orientation.
4	Salman and Gholamlifard, 2006 [22]	Water permeability | Depth of the underground water table | Distance from rivers | Distance from residential areas | Distance from roads | Slope.
5	Al-Jarrah and	Land slope | Soil hydraulic conductivity | Depth to the groundwater from landfill base

	Abu-Qdais, H., 2006 [2]	| Distance to surface water bodies | Distance to residential, religious and archeological sites | Land cost | Percentage of the highest price | Distance from highway | Distance from waste generation source | Distance from airport runway.
6	Gemitz et al., 2007 [23]	Residential areas | Land use | Highways and railways | Environmentally protected areas | Important aquifers | Surface water bodies | Springs and wells | Exceptional geological conditions | Distance from country borders and costlines | Hydrogeology | Distance from water bodies | Hydrology | Proximity to residential areas | Site access | Type of land use | Proximity to waste production centers | Site orientation | Slope of the landfill surface.
7	Delgado et al., 2008 [24]	Distance to water bodies | Distance to urban settlements | Distance to airport | Distance to storage plant | Land cover / land use | Slope | Distance to communication |Soil depth by landforms | Soil permeability | Active faults | Distance to roads.
8	Chang et al., 2008 [7]	Distance to wells | Distance to wetlands | Clay-loam soil | Distance to rivers | Distance to roads | Transportation issues | Environmental and ecological impact | Public nuisance | Economical impact | Historical markers.
9	Guiqin,et al., 2009 [25]	Man & animal habitats | Surface water (rivers, lakes) | Ground water | Distance from airfields | Agricultural land | Forest land | Special land | Land shape (slope, altitude) | Price of lands | Distance of transport (distance from waste production centers, distance from roads).
10	Sharifi et al., 2009 [25]	Physiography (DEM, slope and aspect) | Lithology (karst, aquifers) | Seismic activity (faults) | Underground water (depth, speed & aspect of underground water) | Wetlands (rivers, streamns, springs) | Meteorological analysis (isotherms, isorain) | Climate analysis (climate regimes) | Ecologically important areas (protected areas, hot spots) | Soil classification (land use) | Urban & rural areas (infrastructure, cities, villages).
11	Ekmekçioglu et al., 2010 [26]	Net cost per ton | Technical reliability | Feasibility | Air pollution control | Emission levels | Separation of waste materials |Surface water dispersed releases | Number of employees | Waste recovery | Energy recovery | Hydrology |Topography and soil | Adjacent land use | Climate | Flora and fauna | Site capacity | Road access | Cost.
12	Aragones-Beltran et al., 2010 [13]	Distance to E.D.A.R | Distance to another MSW plant | Distance to landfill | Municipalities and waste volume | Accesses | Water |Runoff and sewage systems | Power |Roads | Water sources | Visual impact | Community affected by smells | Topography | Cattle ways | Archaeological sites | Flood areas | Protected areas | Land planning | Facilities and infrastructures | Environmental issues | Nearby municipalities.
13	Sener et al., 2010 [27]	Roads | Slope | Height | Geology | Landuse | Settlement | Surface water | Aspect (wind) | Protected areas.
14	De Feo and De Gisi, 2010 [14]	Population density of the municipality | Distance from the motorway | Waste production barycentricity | Interference of the additional traffic with local roads | Accessibility | Capacity of traffic splitting | Cost of the area | Economic damages | Absence of areas of the highest value for natural habitats and species of plants and animals | Absence of heavy plants.
15	Moeinddini et al., 2010 [28]	Temperature | Rain | Wind direction & wind gust | Flooding over a 100 years | Permeability of soil | Depth of soil | Slope | Airport | Qanats, springs and wells | Roads and railroads | Residential areas | Faults | Industries power lines | Historical and tourism centers | Surface water | Ground water | Visibility from roads and railroads | Visibility from residential areas | Sensitive ecosystem | Land use and land cover.
16	Abessi and Mohsen, 2010 [29]	Topography | Soil and geology characteristics | Climate | Vegetation maps | Surface and ground water characteristics | Specific environmental zones | Residential zones | Accessibility | Distance to residential areas | Distance to water solutions | Applicability | Waste transport.
17	Geneletti, 2010 [16]	Distance from settlements | Elevation | Slope | Distance from water bodies | Soil permeability | Prime farmland | Ecological values | Dust | Visibility | Accessibility.

18	Al-Hanbali et al., 2011 [12]	Distance from urban areas | Distance from agricultural lands | Distance from roads | Aquifer media | Depth to water table | Distance from faults | Distance from wells | Distance from streams | Slope.
19	Tavares et al., 2011 [30]	Waste transportation cost | Distance from electrical grid | Distance from coast line | fly ash transportation cost | Potable water demand | Land orientation | Land cover | Distance from road network | Distance from coast line | Terrain slope | Terrain elevation | Distance from urban centers | Land use | Air pollution | Visibility impact.
20	Effat and Hegazy, 2012 [15]	Permeability | Ground water depth | Distance to sabkha | Distance to fault | Distance to shores | Distance to protected zones | Distance to high order stream | Distance to cities | Slope of the land | Distance to power supply | Accessibility | Distance to archeological sites | Distance to airport | Aspect.
21	Kara and Doratli, 2012 [31]	Distance from waste generation centers | Distance from roads | Slope | Distance from surface waters | Distance from ground water areas | Distance from environmentally sensitive area | Vegetation types | Soil productivity |Soil permeability | Distance from settlements | Distance from cultural sites | Distance from stones carries.
22	Gorsevki et al., 2012 [32]	Slope | Elevation | Distance from rivers | Distance from lakes | Distance from springs | Landuse | Hydrology | Distance from faults | Distance from urban & rural areas | Proximity to roads | Proximity to building materials | Proximity to dense population.
23	Ersoy et al., 2013 [33]	Legally excluded areas | Soil or rock properties | Topography | Groundwater protection areas | Land use | Surface water protection areas | Residential area | Landscape protection areas | Nature reserve.
24	Gbanie et al., 2013 [34]	Water bodies (perennial streams, intermittent streams) | Groundwater level | Landform/topography | Slope | Geology | Soil | Land cover | Urban | Airfield/helipad |Wind direction (aspect) | Community acceptance | Distance from generation point | Distance from major roads | Land price.
25	Demesouta et al., 2013 [35]	Aquifers vulnerability | Fault zones | Geothermal fields | Water reserve facilities | Seismic hazard assessment | Soil permeability | Slopes | Elevation | Soils texture | Protected areas | Surface water |Forests | Pluviometry | Wetlands | Temperature | Irrigated areas | Strong winds | Borders & coastlines |Land use |Residential areas | Industrial & commercial units | Historical/cultural sites |Transportation network | Water consumption source | Road network discriminations | Airport | Public utilities | Salinization zones.
26	Suthar and Sajuan, 2014 [19]	Site recommendation | Leaching to ground water | Waste transportation and distance from city | Impact on river catchment area | Distance from surface water body | Distance from agricultural land | Distance from forest area |Distance from urban settlement | Aesthetic sense of the city and site location | Land quality (barren/ fertile) | Flood in river catchment area | Effect on water table by diffuse leaching | Pollution impact on urban river | Rainwater drainage problem | Effect on surrounding health and soil quality | Effect on agricultural land | Effect on forest ecosystem | Effect on human health | Ecological risks of decomposing organic wastes (eutrophication, rainwater runoff quality, etc.) | GHG's emission and global warming potential | Risk of volatile compound from dumped municipal waste | Ecological risks of pesticides and other persistent chemical sprayed over landfill sites by Govt. authorities | CO2 storage and sink due to forest | Waste transportation and air pollution issues | Fulfillment of future land requirement for the extension of landfill site | Effect on wild life | Impact on human settled at landfill sites or nearby locations | Odor problems to people residing in nearby location of landfill | Noise problems due to transportation | Risk of disease due to dumping of hospital wastes | Problem due to dust | Problem due to hazardous volatile compounds | Land availability according to projected size of landfill | Land value in real estate market | Operating costs of dumping including manpower and transportation | Economical benefits to waste reuse/recycling community | Economic burden due to spraying pesticides over landfill sites to control epidemical issues | Cost of leachate collection and treatment | Construction cost for drainage system in the bottom of landfill.

C. Data Mining

1) Definition of Data Mining

Data mining builds upon the electronic storage of data and automatic or augmented research which modern computers can carry, this mining serves at resolving problems by data analysis, especially those stored in the databases; Data Mining aims at discovering patterns which can give sense or insure an advantage (it is used in marketing, finance, economy and other fields for competitiveness). These patterns are stronger if they reveal predictions that are none-trivial [36].

2) Principal Component Analysis

The method of principal component analysis (PCA) is one of Data Mining's most used and known methods. It belongs to the descriptive methods, family of geometrical models and sub-family of factorial analysis. When p variables describing n individuals of a population are all numeric, each individual can be represented by a point in a p-dimensional space. The set of individuals is therefore a cloud of points in the p-dimension space. When $p \leq 2$, the distance between individuals can be visualized, in the case of p=3, it still can be visualized but with difficulty, as for $p > 3$, this visualization is impossible. PCA intervenes when we desire to realize a reduction of the dimension of the space while conserving a maximum of represented information, it's a geometrical and statistical approach where we search for a new space where independent axes explain the variability of data.

3) Bivariate Correlation

Bivariate correlation is used for its aptitude to show statistically significant correlation. Correlation is the measure of a relation (or dependence) between two variables, Pearson's correlation was adopted which measure the linear association. This method is not a data mining method; it is a statistical formula permitting to highlight in a simplified fashion relations between variables and allow for the possibility of confirming or denying the interpretations that were made on the basis of PCA's results.

4) Cluster Analysis

Cluster analysis (CA) belongs to the type of methods called descriptive, under the family of geometric models. CA is a method used when the purpose is to find homogeneous sub-sets by conducting statistical operation of grouping objects (individuals or variables) in a limited number of groups named "clusters" or "segments"; these groups share two main properties: firstly, they're not known in advance by the analyst and are not discovered before the operation is finished, secondly, the group is a combination of objects which have similar characteristics and which are separated from other objects that are different (or more different)[37].

CA distributes the objects on groups not in function of predefined criteria but only in function of similitude with other objects, this means that the number of clusters is not known in advance, it is then a descriptive method, not a predictive one [37]. CA, in this work, is based on grouping strategy of "nearest neighbor", with a metric of intervals by the Pearson correlation – a metric is the measure of the distance between similar and none-similar objects-.

D. Data Frequency

To statistically analyze the criteria, we need to measure their frequency of appearance (occurrence) in the articles, such a measure is impossible without resorting to categories; in fact, categories are a step toward elimination of the problem of none-normalization of the criteria's designations, a same criterion can have different designations which results in a dilution of the frequency numbers since if we consider each designation a criterion on its own, we will have occurrences divided between two designations that refer to a same criterion. Also, the effect of settling for the author-given designations is of accepting redundancies, some of the will-be-called criteria will be equivalent in their meaning but not in their names, an overabundance of criteria will therefore add-up to the dilution of the frequency numbers which will negatively impact the results of the analysis and probably produce irrelevant conclusions.

In fact, the frequencies represent measures of data on a discrete measure scale, a set of data is identified as discreet if the scale used isnon continuous, in this, we can, case resort to a frequency table where will be listed numerical measure of appearance of the variables (categories of criteria) in the individuals (articles).

First, we define the categories then we inventory the criteria into categories, then for every article we represent the occurrence of a category by the number one (1) and its absence by the number zero (0). A matrix is formed where the binary values can be summed on the columns to have a relative frequency of a category as compared with other categories; the sum of these binary values on the rows gives an indication on the number of categories cited per article. The matrix of frequencies is the basis for application of the data mining method, each line in it represents what is called a table of absolute frequencies, here the absolute values are for the categories per article.

III. RESULTS AND DISCUSSION

A. *Analysis of Categories*

1) *Creation of categories*

As explained in the data frequency part, categories should be established, tagwords are used to represent the enumerated criteria in the articles. For example, the category {"Agricultural"} contains the tagwords {"agriculture", "farmland", "irrigated areas"}.

TABLE II
Categories and tagwords

	Category	**Tagwords**
1	Agricultural	"agriculture", "farm", "irrigated areas"
2	Airport	"airport", "airfield", "helipad"
3	Meteorological & Climat	"meteorology", "climat", "rain", "pluviometry", "temperature"
4	Border & Coastlines	"border", "coastlines"
5	Cost	"cost"
6	Cultural	"cultural", "religious", "archelogical", "historical"
7	Dust	"dust"
8	Economic	"economic"
9	Electrical	"electrical", "power lines", "power supply", "electrical grid"
10	Ecology & Environment	"ecology", "environment", "ecosystem", "wild life", "flora", "habitat of man and animal", "natural habitat", "global warming", "vegetation"
11	Faults	"fault"
12	Flood	"flood"
13	Forest	"Forest"
14	Geologie	"Geologie"
15	Underground water	'groundwater', "underground water", "aquifer", "hydrogeology", "water table"
16	Health	"Health", "disease"
17	Industrial	"industrial"
18	Infrastructures (miscelleneous)	"facilities", "drainage system"
19	Land price	"land price", "real estate market"
20	Land use	"Land use (LU)", "Land cover (LC)"
21	Odor	"odor"
22	Population density	" population density"
23	Protected areas	"protected area", "nature reserve"
24	Restricted areas	"restricted areas", "legally excluded areas", "special land"
25	Rivers	"rivers"
26	Roads	"roads", "road network", "highway", "accessibility", "transportation network"
27	Separation of waste material	"separation of waste materials"
28	Site (miscelleneous)	"site capacity", "extension of landfill site", "land availability"
29	Slope	"slope", "topography", "physiography"

30	Soil	"soil", "permeability"
31	Urban	"Urban", "Urban areas", "residential areas/zones", "settlements",
32	Visibility	"visibility"
33	Visual	"aesthetic", "visual impact"
34	waste (miscelleneous)	"emission level", "waste water treatment plant", "energy recovery", "number of employees", "surface water dispersed releases", "feasability", "fiabilité technical reliability", "
35	Waste recovery	"waste recovery"
36	Waste transportation	"waste generation source", "waste generation centers", "waste transportation", "centre de génération de déchets", "waste production barycentricity"
37	Waste volume	"waste volume"
38	Water body	"water body", "lake", "dam"
39	Water source	"water sources"
40	Water (generic)	"wells", "hydrology", "water", "water demand", "surface water protection area", "wetland"
41	Wind direction	"wind direction", "orientation", "aspect", "wind gust'
42	Elevation	"physiography", "altitude", "height", "elevation",

2) Analysis of Relative Frequencies

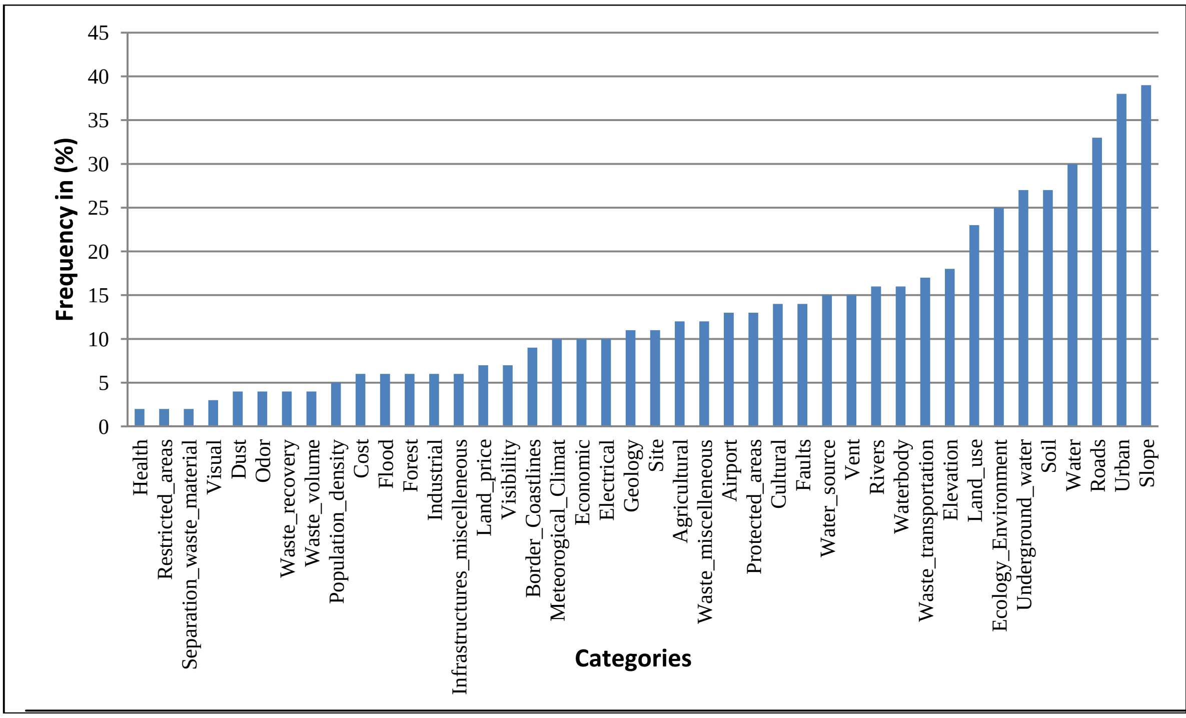

Fig. 1. Histogram of relative frequencies

The relative frequencies allow comparing frequencies of appearance of the categories and forming an idea about the occurrence of a criterion in the literature; the histogram of the relative frequencies shows frequencies under the form of percentages calculated by dividing each absolute frequency on the sum of absolute frequencies. For example, the relative frequency of {agriculture} is 2%, the categories {restricted areas}, {health} and {separation of waste material} have the lowest relative frequencies of 0,3367%, {slope}'s relative frequency is the highest with 7,41%. The analysis of the relative frequencies for the (42) categories reveals that (16) categories have relative frequencies that are superior to the average (2,38%), these categories are {cultural}, {elevation}, {ecology & environment}, {underground water}, {land use}, {river}, {road}, {slope}, {urban}, {waste transportation}, {waterbody}, {water sources} and {water}.

3) Analysis by PCA

Extraction of the principal components

The software IBM SPSS permits to apply the PCA, thanks to this software and the matrix of frequency tables for the (42) categories, we can obtain the table of total explained variance which shows that 75% of the information is explained by the ten (10) first axes resulting for the PCA, (15) axes are necessary before the eigenvalues drops under one (1).

TABLE III
Selected components with the percentages of variance they explain

Components	Initial eigenvalue		
	Total	**% of Variance**	**Cumulative %**
1	6,507	15,493	15,493
2	4,123	9,817	25,310
3	3,691	8,788	34,098
4	3,566	8,491	42,589
5	2,937	6,992	49,582
6	2,518	5,996	55,578
7	2,389	5,688	61,266
8	2,207	5,255	66,521
9	1,976	4,706	71,227
10	1,632	3,885	75,112

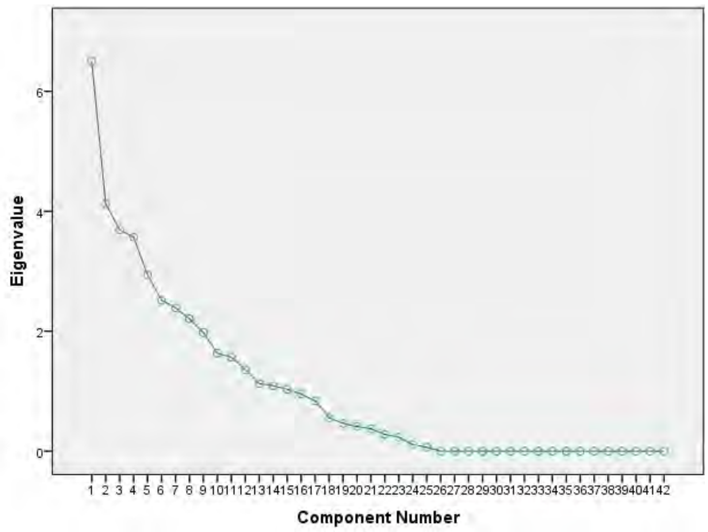

Fig. 2. Eigenvalues for the PCA selected axes

The second table of importance is the matrix of components which shows the correlations of the initial variables with the newly extraced axes. Indeed, it is with the help of these correlations that we can understand the meaning of a new axe but, more interestingly, the correlation of several initial variables with an axe stipulates that they appeared in parallel in the articles.

TABLE IV
Matrix of components

	Components									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Agricultural	,266	-,201	,365	,498	,145	,266	,044	,028	-,183	-,164
Airport	-,318	-,002	,658	,061	,366	-,086	,240	-,128	,167	,234
Meteorological & Climat	-,052	,651	-,031	,037	,392	,459	,160	-,053	-,099	-,078
Border & Coastlines	-,242	,264	,194	,180	,093	-,337	-,120	,191	,286	-,068
Cost	,638	,199	-,269	,174	,441	-,262	-,066	,045	,152	-,094
Cultural	-,041	,320	,505	-,220	-,202	,062	,130	-,485	-,029	,034
Dust	,579	-,105	-,037	,650	-,137	-,063	-,061	-,097	-,151	,148
Economic	,402	-,431	-,060	-,015	-,071	,345	-,427	-,264	,140	-,146
Electrical	,113	,214	,562	-,196	,131	-,448	,123	-,137	,111	-,029
Elevation	-,315	,291	-,212	,288	,051	,063	-,323	,298	-,137	-,064
Ecology & Enviroment	,461	,224	,020	-,007	,059	,140	,034	-,026	-,597	-,014
Faults	-,359	,239	,260	,226	,005	,221	,036	-,232	,526	-,234
Flood	,629	,347	,408	,028	-,165	-,018	-,092	-,209	,101	,128
Forest	,348	-,074	,543	,410	,379	,286	-,024	,245	,056	-,036
Geologie	-,326	,217	-,046	,025	,109	,209	-,247	,566	-,026	,477
Underground water	-,190	,010	,390	,314	,034	,336	,269	,260	-,135	,306
Health	,723	-,077	,103	,536	-,040	,080	-,165	,079	,132	,108
Industrial	-,078	,235	,337	-,104	,344	,400	-,373	-,414	,062	,230
Infrastructures (miscelleneous)	,040	,563	,334	-,142	-,236	,401	,056	,200	,122	-,172
Land price	-,050	-,507	,340	-,291	,324	,033	,026	,054	-,185	,270
Land use	-,343	,523	-,200	,069	,359	-,125	-,283	,079	,159	-,095
Odor	,811	,177	,288	,019	-,355	-,013	-,066	,145	,135	,035
Population density	,002	-,392	-,251	-,276	-,005	,251	-,343	-,092	,304	,321
Protected areas	-,108	,303	-,043	-,139	-,445	-,076	,116	,468	,177	,315
Restricted areas	,039	-,418	,458	-,027	,422	-,034	,270	,247	-,194	-,173
Rivers	-,323	-,383	,054	,124	-,120	,148	,353	,015	,382	-,374
Roads	-,101	-,155	,180	-,579	,364	,045	-,197	,022	,286	,027
Separation of waste material	,328	,252	-,491	-,210	,535	,009	,429	-,015	,116	-,067
Site (miscelleneous)	,613	,154	-,129	,102	-,161	-,129	,392	-,013	,225	,399
Slope	-,685	,125	-,144	,277	,282	-,249	,279	,134	-,130	-,010
Soil	-,313	,239	-,202	,147	-,052	,248	,501	-,332	-,281	,242
Urban	-,233	,382	,182	,435	-,405	-,294	-,178	,052	-,062	-,156
Visibility	-,047	,303	-,020	,213	,102	-,434	-,212	-,567	-,338	,009
Visual	,255	,116	,391	-,512	-,357	-,189	,188	,062	-,080	,075
waste (miscelleneous)	,718	,218	-,147	-,272	,173	-,050	-,232	,099	-,078	-,015
Waste recovery	,759	,126	-,280	,236	,357	,065	,190	,046	,179	,030
Waste transportation	,132	-,174	,353	-,496	,090	-,259	-,233	,248	-,372	-,106
Waste volume	,526	,415	-,141	-,520	,061	-,065	,363	,077	,123	-,091
Water body	,118	-,416	,262	,365	-,039	-,418	,140	,119	,173	-,060

Water source	-,093	,425	,130	-,133	-,334	,400	-,058	,183	-,141	-,349
Water (generic)	,293	,210	,208	-,127	,259	-,092	-,170	,243	-,105	-,274
Wind direction	-,394	,483	,215	,151	,218	-,316	-,298	-,033	,137	,255

The axes and the categories with which they are strongly correlated are shown in the table 5.

TABLE V
PCA's axes in their correlation with categories

Axe	**Categories**
1	Coast, Dust, Flood, Health, Odor, Site (miscellaneous), Waste (miscellaneous), Waste recovery, Waste volume.
-1	Slope
2	Climat, Infrastructures (miscellaneous), Land use.
-2	Land price.
3	Airoport, Cultural, Forest , Electrical.
4	Health
-4	Roads, Visual, Waste volume.
5	Separation of waste material
6	-
7	Soil
8	Geology
-8	Visibility
9	Faults
-9	Ecology & environment
10	-

The presence of categories with weak frequencies have strongly influenced the PCA process, in fact, PCA is a linear technic based on quadratic optimization; it does not account for non linear relations and is very sensitive to extreme values, the weaker frequency are among these extremes.

To avoid this problem and enhance the statistical validity of the numbers, we resort to grouping the categories in supercategories.

B. Analysis of Supercategories

1) Creation of Supercategories

The results obtained by the PCA analysis on categories have shown limitations, by regrouping the categories in supercategories, we expect strengthening correlations and gain in statistical meaningfulness.

TABLE VI
Supercategories and categories

	Supercategory	Categories included
1	Hydrology	Water, Water source, Waterbody, Coastlines & borderlines, Flood.
2	Hydrogeology	Underground water
3	Wastes	Separation of waste material , Waste (miscellaneous), Waste recovery, Waste volume, waste transportation.
4	Geology & faults	Geologie, Faults
5	Topography	Elevation, Slope
6	Climat	Meteorological & Climat, Wind direction
7	Infrastructure	Electrical, Infrastructures (miscelleneous), Roads
8	Land use	Agricultural, Airoport, Forest, Land use, Protected zones, Restricted areas, Cultural
9	Economical	Cost, Economical, Industrial, Land price
10	Lithology	Soil
11	Social	Health, Odors, Population density
12	Technical	Site, Visibility, Visual,
13	Ecology & environment	Ecology & Environment

2) *PCA Analysis*

Extraction of principal components

This analysis reveals that five (5) axes are capable of explaining the data; they all have eigenvalues greater than one (1) , the threshold of 75% on explained variance is not maintained unless we conserve six (6) axes not five (5), which give us a cumulative of 78,63%.

TABLE VII
Rotated component matrix

		Component				
		1	**2**	**3**	**4**	**5**
1	Hydrology	,146	,204	-,002	,034	,824
2	Hydrogeology	,154	,564	-,510	,418	-,091
3	Wastes	,479	,017	,638	-,222	-,266
4	Geology_Faults	-,322	,663	-,366	-,258	,161
5	Topography	-,662	,287	-,176	,041	-,122
6	Climat	-,388	,777	,163	-,050	-,065
7	Infrastructure	,124	,094	-,143	-,819	-,231
8	Land use	-,154	,760	-,004	,066	,350
9	Economical	,763	-,231	-,107	-,152	-,147
10	Lithology	-,449	,024	-,113	,624	-,287
11	Social	,758	-,108	,120	-,047	,178
12	Technical	-,001	-,060	,796	,179	,115
13	Ecology_Environment	,399	,227	,476	,470	-,308

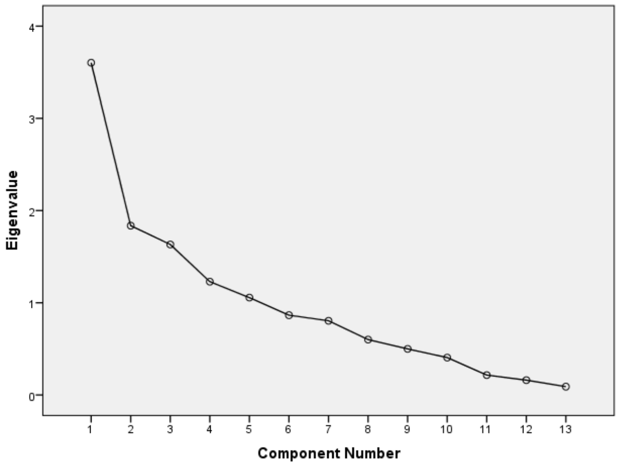

Fig. 3. Eigenvalues retained for the axes of PCA

Interpretation of the PCA results

To interpret the factorial axes obtained with PCA, we can either use the matrix of components which links the intial variables (here, the supercategories) with the new axes or use the space projections, in this case, this means projecting the five-dimensional space (5 new axes) into two-dimensional spaces. These projections can help gaining an understanding between the supercategories and new axes by visual reading; still, the matrix of components is a much concise way of developing such an understanding.

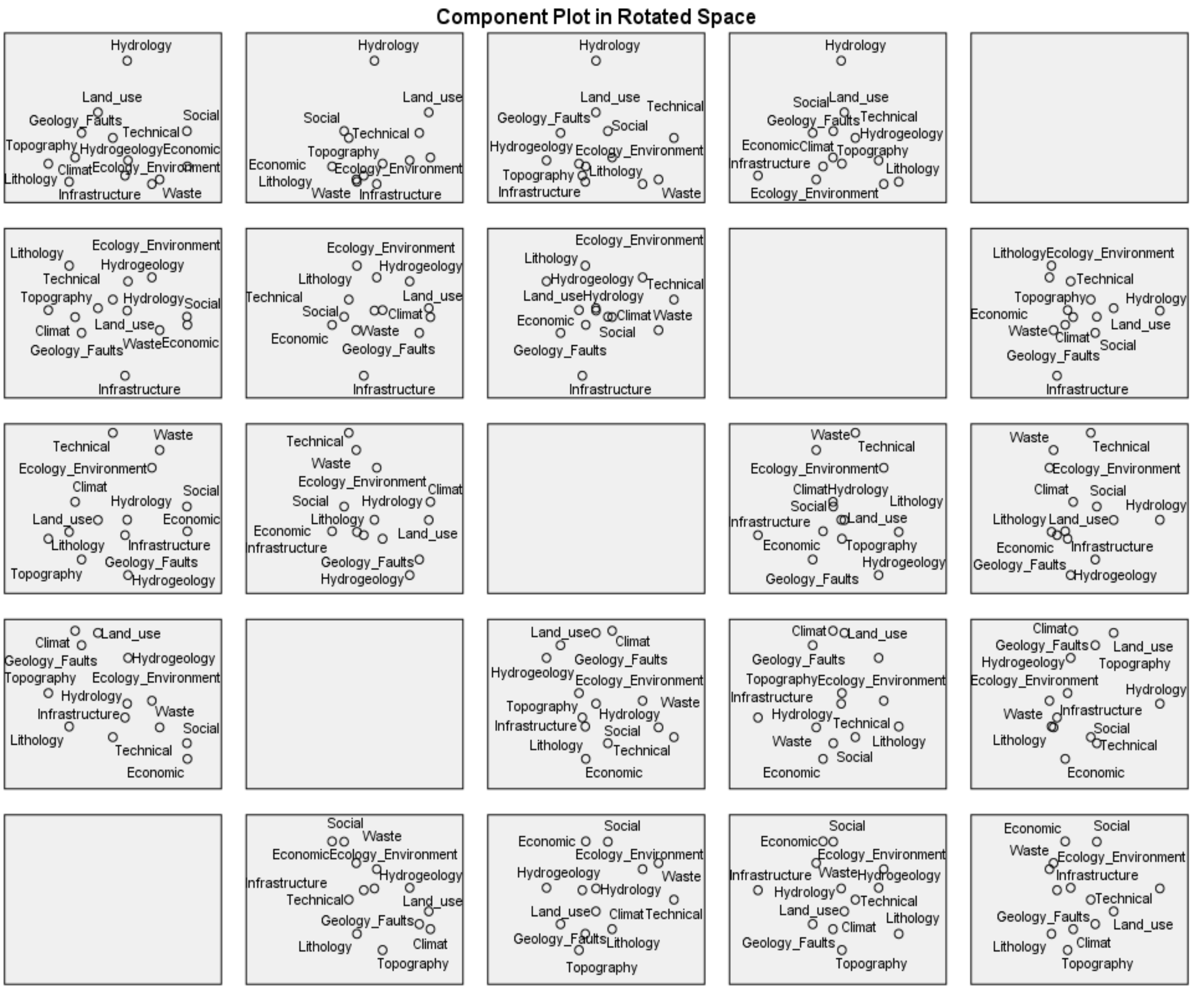

Fig. 4. Bidimensional projections of the PCA space, the entitled points represent the initial variables

TABLE VIII
Axes and their correlations with supercategories

Axe	Strongly correlated supercategories
1	Economic, Social
-1	Topography
2	Geologie_Faults, Climat, Land use
3	Wastes, Technical
4	Lithology
-4	Infrastructure
5	Hydrology

Analysis by bivariate correlation

Bivariate correlation measures the correlation pair-wise and verifies their statistical validity by a Pearson test, the result is a square matrix of a dimension (13) (the number of supercategories), we only take notice of the statistically significant correlations (table 9). We resort to this analysis in order to confirm or deny the patterns extracted from the PCA analysis.

TABLE IX
Pairs of supercategories for which the Pearson correlations are significant

Pairs de or correlated supercategories	Correlation (Pearson's)
Wastes-Ecology&Environment	0,425
Geology&Faults – Topography	0,445
Geology – Climat	0,582
Geolog&/Faults – Land use	0,485
Topography – Climat	0,400
Topography – Economic	-0,429
Topography – Social	-0,465
Climat – Land use	0,534
Economic – Topography	-0,429
Economic – Social	0,563
Lithology – Social	-0,422
Social – Topography	-0,465

By representing these correlations in a graph where the arcs represent either a positive or negative correlation (i.e: a correlation or an anti-correlation), this graph reveals some cluster of supercategories linked by positive and statistically significant correlations.

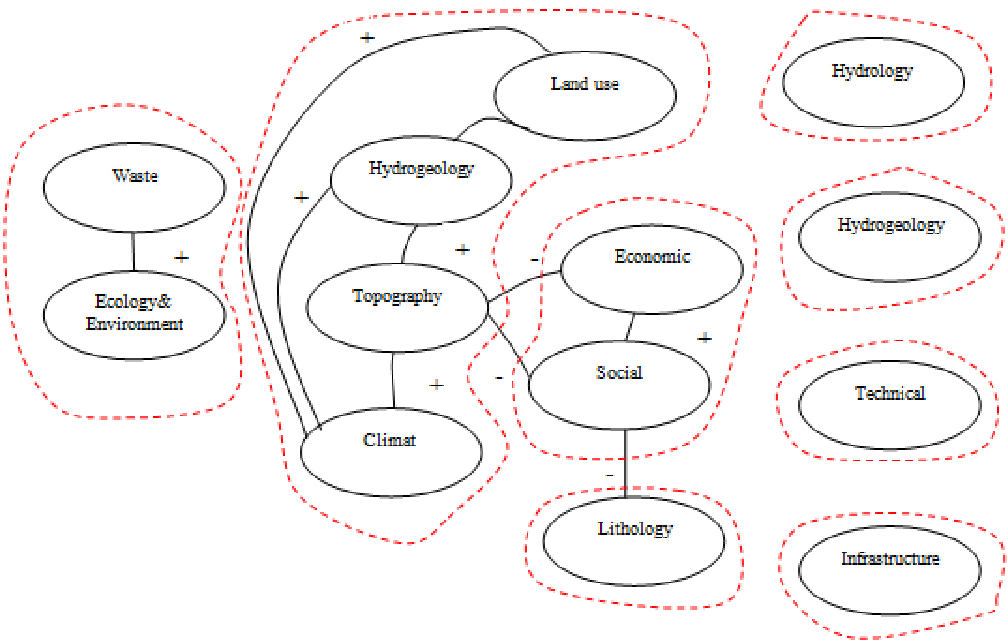

Fig. 5. Map of supercategorie and their correlations.

The arcs with a "+" sign represent positive correlations, the arcs with "-" sign represent negative correlations. The clusters extracted by bivariate analysis are shown in the table 10.

TABLE X
Pairs of supercategories for which the Pearson correlations are significant

Cluster	**Included supercategories**
1	Waste, Ecology_Enviroment.
2	Land use, Hydrogeology, Topography, Climat
3	Economic, Social
4	Lithology
5	Hydrology
6	Hydrogeology
7	Technical
8	Infrastructure

Cluster Analysis

An analysis by Cluster Analysis (CA) is conducted with the "nearest neighboor" strategy and an interval measure by Preason's correlation, the dendrogram of the method is shown in figure 6.

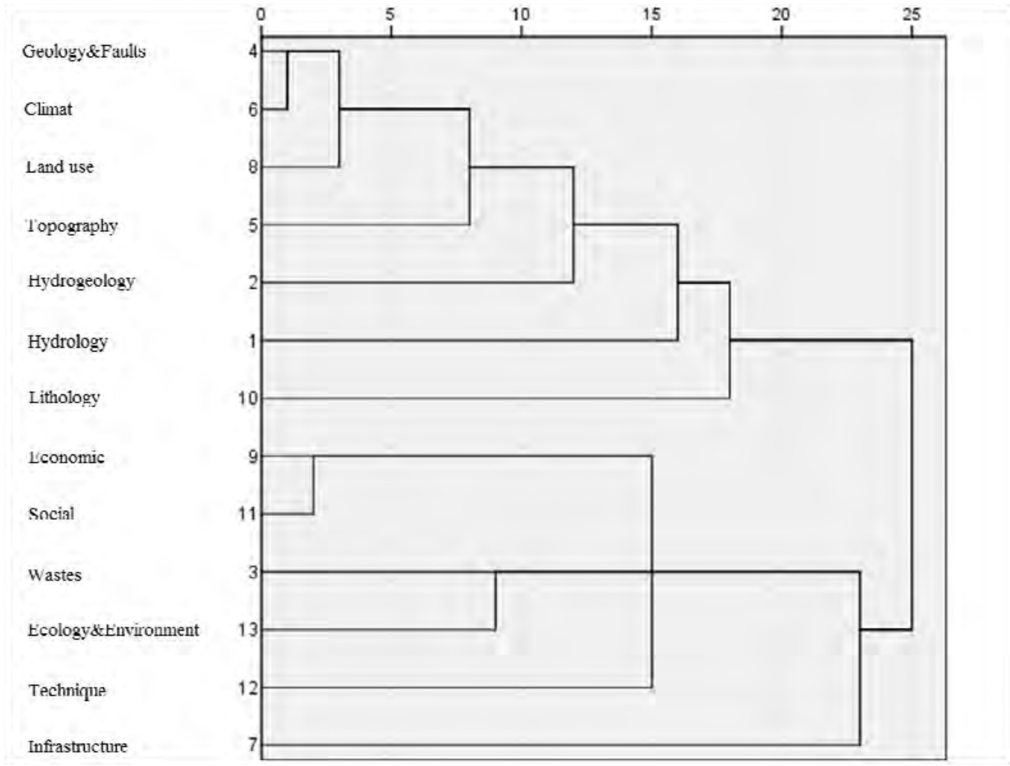

Fig. 6. Dendrogram of the CA method

Interpretations

The dendrogram of the CA method reveals that the distance between {Topography} and {Geology&Faults, Climat; Land use} is significant, in fact, bivariate correlation (BC) favors the cluster {Topography; Geology&Faults; Climat; Land use} but the PCA analysis favors breaking up this cluster into two cluster {Topography} and {Geology&Faults; Climat; Land use}. Hence, it seems that retaining the two distinct clusters is the most justifiable choice.

A proximity between {Hydrology} and {Lithology} can be viewed on the dendrogram, this proximity is not though maintained in the BC analysis neither is it in the PCA analysis, the supercategories {Hydrologie}, {Lithology}, {Hydrogeology}, {Technical}, {Infrastructure} should form each a cluster, still, there is a clearer proximity among {Hydrogeology}, {Hydrology} and {Lithologie} than between these and the rest of five cited supercategories.

{Economic} and {Social} are to be included into a single englobing cluster {Economic; Social}, all the methods agree to that. {Ecology&Environment} and {Waste} can form a cluster by CB and CA, PCA brings closer {Waste} and {Technical} which can also be justifiable by CA but not the point of envisaging an englobing cluster of {Waste; Ecology&Environement; Technical}, we then retain two clusters {Waste; Ecology&Environment} and {Technical}.

Finaly, we retain the following clusters: {Economical; Social}, {Waste; Ecology&Environment}, {Topography}, {Geology&Faults; Climat; Land use}, {Hydrogeology}, {Lithology}, {Hydrogeology}, {Technical} and {Infrastructure}.

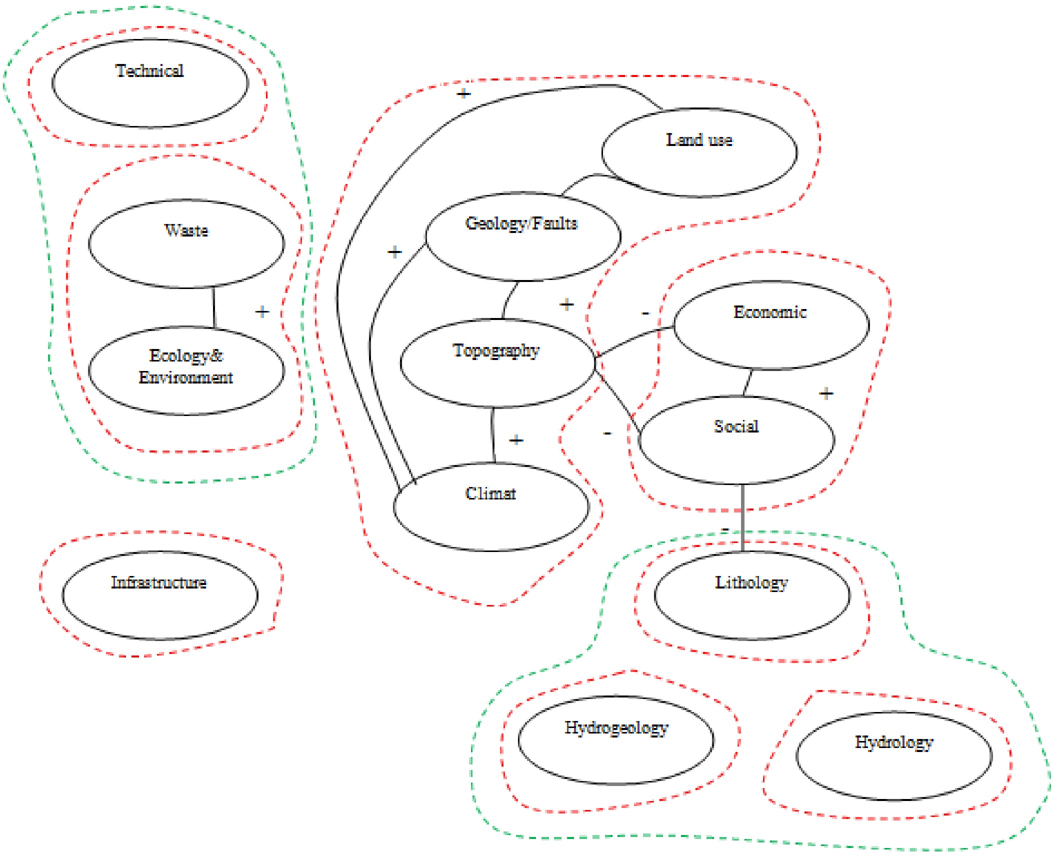

Fig. 7. The cluster obtained by comparison of the PCA, BC, CA results;

In red doted lines: confirmed clusters, in green doted lines: less confirmed clusters.

IV. CONCLUSION

The goal of this review was to establish a state of the art of the criteria presented by the literature for site selection; while collecting the articles, only the ones appearing in recognized journals and related to the subject were admitted, in fact, to verify the conformity of these articles with the theme, the articles were read and criteria used in each were collected.

While collecting the criteria, the need arose to eliminate redundancy, introducing categories served that purpose, categories gather criteria having a unique semantic meaning but different naming; in fact, a category tries to express one criterion, each criterion is recognizable by many tagwords.

These categories have helped regulate the nomenclature given to the criteria; nonetheless, the first PCA analysis exposed a weakness of the setting: some of the categories had minor frequency numbers which have isolated them; PCA is known for being unable to handle extremes and thus it was necessary to adopt larger categories which we called supercategories. If each category represented criteria, a supercategory represented a group of criteria. By grouping categories into supercategories, the minor frequencies will vanish into larger groups and it is expected that the significance of the statistics will be enhanced. PCA, BC and CA aided in studying correlations between supercategories in a conclusive fashion, thus, clusters of criteria were isolated with each one grouping criteria that have frequently appeared together, this allows to answer questions such as "what are the supercategories the authors tended to use simultaneously ?", "Given a supercategory X, what other criteria should accompany it as far as literature is concerned ?". Still, this information is only indicative of the relations between supercategories not between categories/criteria. The supercategories may be approached as broad indications for the taxonomy of criteria a researcher should consider for his specific siting problem. In preparing the criteria for siting problem, one should consider the following clusters : {Technical, Waste, Ecology&Environment}, {Infrastructure}, {Land use, Geology/Faults, Topography, Climat}, {Economic, Social}, {Lithology, Hydrogeology, Hydrology}. For example, if the researcher opts for criteria relevant to topography he should study the integration of criteria related to land use, geology and faults, climat; going backward, from supercategories to categories to tagwords, criteria can be discovered which might be of interest to the his study.

As mentioned in the introduction, the literature review is only a step in studying the criteria involved in the landfill siting problem, it can be completed by stakeholder analysis; stakeholders contribution is particularly interesting since it confines even greater importance to the locality and regionality of a site selecting process.

REFERENCES

[1] D. P., Komilis, R. K., Ham, R., Stegmann, " The effect of municipal solid waste pretreatment on landfill behavior: a literature review", Waste Management and Research, vol. 17, no. 1, pp. 10-19, 2002.

[2] O., Al-Jarrah, H., Abu-Qdais, "Municpal solid waste landfill siting using intelligent system", Waste Management, vol. 26, pp. 299-306, 2006.

[3] O., Abessi, S., Mohsen, "Hazardous waste landfill siting using GIS technique and analytical hierarchy process", EnviromentAsia, vol. 3, no. 2, pp. 69-78, 2010.

[4] G., Tchobanoglous, H., Theisen, S., Vigil, Integrated solid waste management: engineering principles and management issues, New York, U.S.A, McGraw Hill, 1993.

[5] E. A., McBean, F. A., Rovers, G. J., Farquhar, Solid waste landfill engineering and design, Australia, Englewood Cliffs, N. J., Prentice Hall, 1995.

[6] T. D., Kontos, D. P., Komilis, C. P., Halvadakis, "Siting MSW landfills on Lesvos island with a GIS based methodology", Waste Management & Research, vol. 21, pp. 262-277, 2003.

[7] N. B., Chang, G., Parvathinathan, J. B., Breeden, "Combining GIS with fuzzy multicriteria decision-making landfill siting in a fast growing urban region", Journal of environmental management, vol. 87, pp. 139-153, 2008.

[8] T. S., Hale, C. R., Moberg, "Location science research: A review", Annals of Operations Research, vol. 123, no. 1-4, pp. 21-35, 2003.

[9] S., Cheng, C., Chan, G., Huang, "Using multiple criteria decision analysis for supporting decisions of solid waste management", J. Environ. Sci. Health A, vol. 37, pp. 975-990, 2002.

[10] K., Charnpratheep, Q, Zhou, B., Garner, "Preliminary landfill site screening using fuzzy geographical information systems", Waste Manage. Res., vol. 15, pp. 197-215, 1997.

[11] G. M., Savage, L. F., Diaz, G. C. Golueke, Guidance for land filling waste in economically developing countries, Cincinnati, USA, United States Environmental Protection Agency, National Risk Management Research Laboratory, 1998.

[12] A., Al-Hanbali, B., Alsaaideh, A., Kondoh, "Using GIS-based weighted linear combination analysis and remote sensing techniques to select optimum solid waste disposal withing Mafraq city, Jordan", Journal of geographic information system, vol. 3, pp. 267-278, 2011.

[13] P., Aragones-Beltran, J. P., Pastor-Ferrando, F., Garcia-Garcia, A., Pascual-Agallo, "An analytic network process approach for siting a municipal solid waste plant in the metropolitan area of Valencia (Spain)", Journal of environmental management, vol. 91, pp. 1071-1086, 2010.

[14] G., De Feo, S., De Gisi, "Using an innovative criteria weighting tool for stakeholders involvement to rank MSW facility sites with AHP", Waste Management, vol. 30, pp. 2370-2382, 2010.

[15] H. A., Effat, M. N., Hegazy, "Mapping potential landfill sites for north Sinai using spatial multicriteria evaluation", The Egyptian Journal of Remote Sensing and Space Sciences, vol. 15, pp. 125-133, 2012.

[16] D., Geneletti, "Combining stakeholder analysis and spatial multicriteria evaluation to select and rank inert landfill sites", Waste Management, vol. 30, pp. 328-337, 2010.

[17] M., Puig, C., Woodridge, R. M., Darbra, "Identification and selection of environmental performance indicators for sustainable port development", Marine Pollution Bulletin, vol. 81, pp. 124-130, 2014.

[18] S., Leão, I., Bishop, D., Evans, "Spatial-temporal and spatial multicriteria evaluation to select and rank inert landfill sites", Waste Management, vol. 30, pp. 328-337, 2004.

[19] S., Suthar, A., Sajuan, "Rapid Impact assessment matri (RIAM) analysis as decision tool to select new site for municipal solid waste disposal: a case study of Dehradun city, India", Sustainable Cities and Society, vol. 13, pp. 12-19, 2014.

[20] H., Lin, J. J., Kao, "A vector-based spatial model for landfill siting", Journal of Hazardous Materials, vol. 58, pp. 3-14, 1998.

[21] T. D., Kontos, D. P., Komilis, C. P., Halvadakis, "Siting MSW landfills with a spatial multiple criteria analysis methodology", Waste Management, vol. 25, pp. 818-832, 2005.

[22] A., Salman, M., Gholamlifard, "Siting MSW landfills with a weighted linear combination methodology in a GIS environment", International Journal of environmental science & technology, vol. 3, no. 4, pp. 435-445, 2006.

[23] A., Gemitzi, V. A., Tsihrintzis, E., Voudrias, C., Petalas, G., Stravodimos, "Combining geographic information system multicriteria evaluation techniques and fuzzy logic in siting MSW landfills", Environmental geology, vol. 51, pp. 797-811, 2007.

[24] O. B., Delgado, M., Mendoza, E. L., Granado, D., Geneletti, "Analysis of land suitability for the siting of inter-municpal landfills in the Cuitzeo lake basin, Mexico", Waste Management, vol. 28, pp. 1137-1146, 2008.

[25] M., Sharifi, M., Hadidi, E., Vessali, P., Mostakhani, K., Taheri, S., Shahoie, M., Khodamoradpour, "Integrating multi-criteria decision analysis for GIS-based hazardous waste landfill sitting in Kurdistan Province, western Iran", Waste Management, vol. 29, pp. 2740-2758, 2009.

[26] M., Ekmekçioglu, T., Kaya, C., Kahraman, "Fuzzy multicriteria disposal method and site selection for municipal solid waste", Waste Management, vol. 30, pp. 1729-1736, 2010.

[27] S., Sener, E., Sener, B., Nas, R., Karagüzel, "Combining AHP with GIS for landfill site selection: a case study in the lake Beysehir catchment area (Konya, Turkey)", Waste Management, vol. 30, pp. 2037-2046, 2010.

[28] M., Moeindddini, N., Khorasani, A., Danekhar, A. A., Darvishsefat, M., Zenalyan, "Siting MSW landfill using weighted linear combination and analytic hierarchy process (AHP) methodology in GIS environment (case study: Karaj)", Waste Management, vol. 3.0, pp. 912-920, 2010.

[29] O., Abessi, S., Mohsen, "Hazardous waste landfill siting using GIS technique and analytical hierarchy process", EnvironmentAsia, vol. 3, no. 2, pp. 69-78, 2010.

[30] G., Tavares, Z., Zsigraiova, V., Semiao, "Multi-criteria GIS-based siting of an incineration plant for municipal solid waste", Waste Management, vol. 31, pp. 1960-1972, 2011.

[31] C., Kara, N., Doratli, "Application of GIS/AHP in siting sanitary landfill: a case study in northern Cyprus", Waste Management & Research, vol. 30, no. 9, pp. 966-980, 2012.

[32] P. V., Gorsevki, K. R., Donevska, C. D., Mitrovski, J. P., Frizado, "Integrating multicriteria evaluation techniques with geographic information systems for landfill site selection: a case study using ordered weighted average", Waste Management, vol. 32, pp. 287-296, 2011.

[33] H., Ersoy, F., Bulut, M., Berkün, "Landfill site requirements on the rock environment: a case study", Engineering Geology, vol. 154, pp. 20-35, 2013.

[34] S. P., Gbanie, P. B., Tengbe, J. S., Momoh, J., Medo, V. T. S., Kabba, "Modelling landfill location using geographic information systems (GIS) and multicriteria decision analysis (MCDA): case study Bo, Southern Sierra Leone", Applied Geography, vol. 36, pp. 3-12, 2013.

[35] O. E., Demesouta, A. P., Vavatsikos, K. P., Anagnostopoulos, "Suitability analysis for siting MSW landfills and multicriteria spatial decision support system: Method, implementation and case study", Waste Management, vol. 33, pp. 1190-1206, 2013.

[36] I. H., Witten, E., Frank, M. A., Hall, Data Mining: Practical Machine Learning Tools and Techniques, Burlington, Morgan Kaufmann, 2011.

[37] S. Tufféry, Data Mining and Statistics for Decision Making, Chichester, West Sussex Wiley, 2011.