USING SPATIAL DATA MINING TECHNIQUES TO REVEAL VULNERABILITY OF PEOPLE AND PLACES DUE TO OIL TRANSPORTATION AND ACCIDENTS: A CASE STUDY OF ISTANBUL STRAIT

Abdulvahit Torun*, Şebnem Düzgün**

*General Command of Mapping, Cartography Dept., 06100 Cebeci, Ankara, Turkey
**Middle East Tech. Univ., Geod. & GI Tech. Dept, Ankara, Turkey -
atorun@hgk.mil.tr; duzgun@metu.edu.tr

Technical Commission II, WG II

KEY WORDS: Spatial Analysis, Spatial Statistics, Spatial Data Mining, Vulnerability, Spatial Data Warehousing

ABSTRACT:

Public and private organizations have legacy or operational spatial databases or non-spatial databases, which are also somehow linked to a spatial database or a spatial meaning. In addition to mission related databases, these organizations either have or access several databases comprising such as census, economic, security, image, multimedia, statistical information for planning, intelligence, decision and policy making. The extended size of these data sets makes it difficult to search for meaningful patterns or relationships among data. Data mining is a technique allows researchers to overcome obstacle and discover potentially interesting and useful patterns of information embedded in large databases. Spatial data mining (SDM) differs from non-spatial data mining because of the underlying spatial data where attributes of spatial objects are affected by attributes of spatial neighbours. There is a strong relationship between modern knowledge-management and decision-making, where information extraction from great size of data is currently an important issue. Owing to the recent transitions in social and environmental issues, public and governmental organizations try to satisfy a large set of new requirements against new threats regarding natural and non-natural disasters. One way of minimizing losses, making mitigation more effective, increasing responsiveness to different types of crises/disasters, and conserving resources, is increasing the efficiency of knowledge discovery, which plays an important part in decision and policy making and implementation in the age of information. In this study, after introducing spatial data mining techniques, use of these techniques for revealing vulnerability of people and places due to oil transportation and oil/gas fires at Istanbul Boğazı (Istanbul Strait). Data is structured as a spatial data warehouse after cleaning, selection, transformation, reduction and consolidation to accomplish SDM operations.

1. INTRODUCTION

1.1 Oil Transportation as Disaster Risk

The Turkish Strait System includes the Straits of Istanbul, Çanakkale and the Marmara Sea, connecting the Black Sea and the Mediterranean Sea. Due to unique characteristics and conditions of İstanbul Strait with narrowest point of 740 m, it poses considerable risks for maritime transportation. As, İstanbul Strait has rocky curves and sharp turns, surface and subsurface currents and counter currents, and sudden daily and seasonal changes in weather conditions, the route is difficult and dangerous (İstikbal, 2000).

More than 50.000 vessels, in average, annually use the Turkish Straits and this number has been increasing steadily. That means 15 passages out of 150 per day are tankers, half of which are 5% of which were tankers longer 200 m (one passage at every 10 minutes).

Accidents of shipping in the straits are examined under four categories: collision, grounding, fire and standing. When an accidents occurs accumulation of particles in the air and explosion during fire and heavy oil contamination formed on the surface of the sea and on the shores of Marmara Sea and the Istanbul Strait cause high risk of catastrophic disasters.

Because of being a mega-city, probably being at the most severe seismic hot spot world-wide, subject to ship/tanker accident at passing through the Strait, Istanbul is face-to-face a severe disaster every day. A single catastrophic disaster accident in the İstanbul Straits could cause fires, huge disasters on the coastal areas, environmental pollution, risk the lives of millions of inhabitants and destroy the historical heritage of many thousands of years minute (İstikbal, 2000; Erdik, 2002; Kötter, 2004; Erdik, 2005; Fernandez, 2005; Linnerooth-Bayer, 2005).

Global trends in population, urbanization, natural (earth quake, climate) and non-natural (terror, fire, land use, water use) disasters (among others) are imposing stresses and risks on societies and their environments. The impacts affect people differentially due to thier wealth, environment and location (Linnerooth-Bayer, 2005).

Vulnerability is the concept that explains why, with a given level of physical exposure, people are more or less at risk and likely to experience harm as a result of exposure. In this study a limited definition of vulnerability is assumed. The closeness and accessibility to services such as health, security, fire and shelter stations by using main road network of Istanbul when there happens a ship accident.

This paper focuses on explorative spatial data analysis and clustering methods to find vulnerable areas and people, and pattern of ship accidents according to accident data for last 8 years. The resultant patterns and maps may help identifying areas where vulnerability or people is high for use of local decision makers and planners.

2. METHODOLOGY FOR VULNERABILITY

Natural disasters are result of sudden energy release of the three basic environmental phenomena: air, water and land. Non-natural disasters are results of fire, accidents, terror, war etc.

For risk assessment and risk mapping, vulnerability analysis for disaster-prone areas by incorporating information about past disasters, socio-economic conditions of people, and inventories of damaged major structures in the affected area are needed (Kötter, 2004; Muggenhuber et al., 2004; Thomas and Cutter, 2002).

A knowledge based data cleaning method is used to make the data consistent, complete and correct. Methods of defining vulnerability and a GIS methodology employing spatial multicriteria analysis, spatial data analysis and clustering methods to determine pattern of ship accidents are given in the rest of this part.

2.1 Istanbul Data: Census, Public Services and Ship Accident

Various spatial and spatially referenced data are needed during different phases of disaster. Data about spatial, environmental and hydrological characteristics, social and economic life around Strait, registered fire and accidents caused by navigating ships and tankers are needed for complete and accurate decision making.

The spatial extent of the study covers Boğaziçi part of İstanbul. The spatial data set contains boundaries of sub-provinces (ilce), districts (mahalle), parcel blocks (ada), transportation network (roads, streets), location of services (security such as police and fireman, schools etc.), ship accident data. Population of mahalle and ilce are due 1996 (Table 1).

The data were collected for management of disasters after 1999 Golcuk Earth Quake by İstanbul Metropolitan Municipality (IMM). All spatial and spatially referenced attribute and statistics data are structured as a geodatabase in ESRI ArcGIS in ED 50 datum and UTM projection.

2.2 Data Cleaning and Data Preparation

Data Cleaning Model: One of the bottlenecks of vulnerability assessment is not having enough, up-to-date, accurate, and complete data for modeling, validating and testing phases. Thus data cleansing is a crucial process to create necessary base providing no bias is injected to the data. Data populated in data warehouse from various external sources usually contains syntax, schema and semantics (S^3) differences in addition to errors due to correctness, completeness and consistency (C^3).

There are three types of approaches for analysis of data sets with missing values such that 'using only available complete data', 'finding a single estimate of missing value where the missing value is a drawing from an underlying probability model' and 'using model-distribution- based approach to missing-data prediction' (Haining, 2005). Being Simple and intuitive, the second approach is applicable to fill missing field values based on imputation.

Then, data is evaluated by means of national, regional and intra-statistics about each item considering feature type, category and its hierarchy among other features. The outliers are detected by comparing average, standard deviation and maximum of similar features. Data cleansing model is depicted in Figure 1.

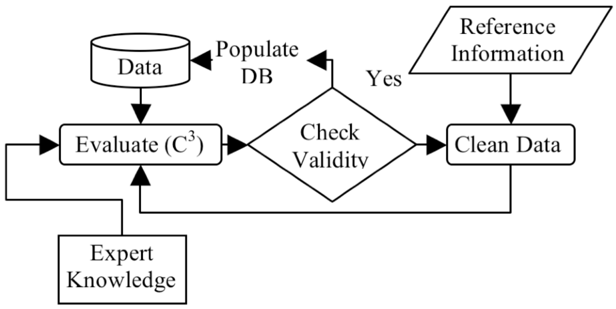

Figure 1. Data Cleaning Work Flow

2.2.1 Data Preparation

Census Data: Census data referring to boundaries of ilce and mahalle are converted into field model. A smooth transition census surface among neighbouring census tracts is accomplished by introducing a low pass filter which affects boundaries more than central part of tracks (Bonham-Carter, 1994).

Public Services: Public services used in analysis are given in Table 1. Rating values are assigned to each feature according to its service capacity among total capacity.

Feature type: Poly	Normalized rating (weight), interval [0,1]	Data Content
School Egitim	service area	Shelter: Schools (primary to high)
Health Saglik	Personnel + service area + mobile service	Health Centers: Hospitals, local clinics
Fire Itfaiye	personnel x vehicle	Fire stations and fire groups
Security Guvenlik	personnel x vehicle	Security: Gendarme and police stations

Table 1. Feature classes and indexes used in analysis

Ship Accidents: There is a database of 'Navigation and Accidents' comprising two main relations namely 'Ship Information Table-SIT' and 'Accident Information Table-AIT' disseminated by Directorate of Secure Ship Transits, Turkey. Seas of Turkey are partitioned into 6 sub-regions namely İstanbul Strait (1), Çanakkale Strait (2), Marmara Sea (3), Karadeniz (4), Ege Denizi (5), Akdeniz (7). SIT comprises information about Accident ID, Name and Flag, Type and carry, weight, width and length, draft, navigation direction, number of passengers. AIT contains items of Accident ID, Region ID, location (verbal descriptive), time, hour, cause of accident, loss of life and goods, environmental pollution, meteorological conditions, informing institution, time of information, brief information, conclusion and action taken.

As the information in SIT table is mostly not complete, this table is discarded. Time dimension is not considered in this study. More than 150 cases are analyzed since 1998 at Istanbul Strait. Attribute values for 'cause of accident' are assigned with following values according to the description and information given in AIT table. 'missing', 'mul-function or bad weather', 'leaning one side, wrong navigation', 'aground', 'hitting', 'colliding', 'sinking' and 'fire'. Domain of 'Loss of Life and Goods' contains unique values of 'danger', 'wounded', 'lost', 'lost due to sink' and 'death'. Attribute, 'Environmental pollution' takes values of 'none', 'unknown' and 'exists'. Instance of these three items are cleaned and populated with rating values which are ordinal sets having 8, 5 and 3 members respectively. The higher number for rating represents the more harmful loss. Location of the accidents are given in descriptive statements such as ‘in the front of shore …, offshore …, … miles from west of …’. These spatial information in verbal format are located on a map by the help of an expert on İstanbul Strait.

2.3 Vulnerability Model: Multicriteria Analysis

Many concepts, rules, and principles associated with vulnerability are employed in methodologies designed for urban vulnerability analysis. There is no unique, identifiable, widely accepted model that generates objectively optimal solution which works with minimal set of spatial, socioeconomic, environmental and disaster data (Rashed and Weeks, 2003).

Vulnerability is a vector field that specifies a geographic location that is to be damaged higher than a certain level because of a disaster. A location is assigned by an index which is calculated as a function – usually a summation function – of affecting factors representing the ability of reaching/by reached public services and closeness to danger (Cutter et al., 2000). Each factor is weighted by its relative significance in affecting vulnerability of people and areas. The higher the index shows the greater the people are open to danger.

Each service/factor such as security, hospitals etc. is assigned a rating and weight in the range of 0.1-1. In the first run, cost distance of each service is calculated based on cost surface of road network (Figure 2). The index, a measure of vulnerability of people in administrative units, is computed by summation of the products of rating and weights of each factor as follows:

$$Vul\_indx = D(pop)\sum_{srv} cD(srv)Dw(srv) \qquad (1)$$

Where:

$D(pop)$: density of population (ratio of population according to area)

$cD(srv)$: Accumulative cost of each cell to service locations that are given in Table 1.

$Dw(srv)$: Normalized rating values (weights) given in Table 1 for services

$srv = \{School, Health, Fire, Security\}$

Categorical/nominal data is mapped on a categorical numeric scale for numeric analysis. Qualitative data such as road quality in road network are mapped into an ordinal scale starting from 1 to 10.

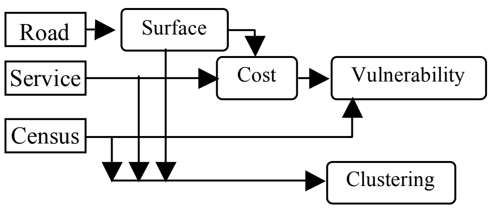

Figure 2. Vulnerability analysis and clustering

2.4 Spatial Data Mining Techniques to determine Distribution of Ship Accidents

The accidents happened in İstanbul Strait are analysed by means of spatial data analysis and data mining techniques.

Exploratory Spatial Data Analysis Techniques: Spatial data analysis comprises a broad spectrum of techniques that deals with both spatial and non-spatial characteristics of spatial objects. Exploratory techniques allow to investigate first and second order effects of the data. First order variation informs about large scale global trend of phenomena which is spatially distributed through the area. The second order variation defines dependence between observations.

Global Autocorrelation (Moran’s I): Moran’s I is a measure of global spatial autocorrelation. Global or local autocorrelation reveal feature similarity based location and attribute values to explore the pattern whether it is clustered, dispersed, or random (Longley et al., 2003).

Hot Spots (Getis-Ord): The G-statistic is often used to identify whether hot spots or cold spots exist based on so-called distance statistics. Hot spots are regions that stand out compared with the overall behaviour prevalent in the space. Hot spots can be detected by visualizing the distribution in format of choropleth or isarithmic maps (Baily et al., 1996; Shekhar et al., 2003; Haining, 2005).

Local Autocorrelation (Anselin’s Local Moran I): The local Moran statistic is used as a local indicator of spatial association which is calculated for individual zones around each observation within a defined neighbourhood to identify similar or different pattern in nearby. Because the distribution of the statistic is not known, high positive or high negative standardized scores of I_i are taken as indicators of similarity or dissimilarity respectively (Haining, 2005).

Density (Kernel): Kernel density estimation is a nonparametric unsupervised learning procedure (classifier). Kernel, k is bivarate probability density function which is symmetric around the origin (Hastie et al., 2001).

2.4.1 Spatial Clustering: Finding Patterns

Assigning each member of a large data set into homogeneous clusters is a fundamental operation in data mining. Clustering is the process of creating a group of data organized on some similarity among the members of a dataset. Each cluster consists of members that are similar in-between the cluster, however dis-similar to members of other clusters. There are four major clustering approaches. Partitional clustering algorithms construct k clusters -usually defined in advance by the user- due to an evaluation criterion. Hierarchical clustering algorithms create a hierarchical decomposition using some criterion which can be represented as dendograms. Density-based partitioning algorithms search for regions which are denser than a given threshold to form clusters from these dense regions by using connectivity and density functions. Grid-based algorithms are based on a multiple-level granularity by quantizing the search space into a finite number of cells (Shekhar et al., 2003; Han et al., 2001a; Han and Kamber 2001b; Chawla et al., 2001).

In this study, two partitioning algorithms namely k-means and ISODATA are used for clustering ship accidents and factors of vulnerability respectively.

k-Means : K-means represents an attempt to define an optimal number of k locations where the sum of the distance from every point to each of the k centers is minimized what is called global optimization. In practice, (1) making initial guesses about the k locations and (2) local optimization for cluster locations in relation to the nearby points is implemented. Thus, two k-means procedures might not produce the same results, even if k is identical because of several underlying local optimization method.

The k-means algorithm is built upon four basic operations: (1) selection of the initial k means for k clusters, (2) calculation of the dissimilarity between an object and the mean of a cluster, (3) allocation of an object to the cluster whose mean is nearest to the object, (4) Re-calculation of the mean of a cluster from the objects allocated to it so that the intra cluster dissimilarity is minimised. Except for the first operation, the other three operations are repeatedly performed in the algorithm until the algorithm converges (until no points change clusters). The essence of the algorithm is to minimise the cost function which is a function of dissimilarity measure between each observation with mean of cluster. Dissimilarity is usually modelled as Euclidean Distance in k-means. The cost function is as follows;

$$Minimize \quad \sum_{j=1}^{n} \sum_{k=1}^{k} a_j d_{jk} z_{jk} \qquad (2)$$

where;

j , k denotes total number of observations and clusters

a_j denotes weight of observation j ,

d_{jk} denotes distance between observation j and centre of cluster k , and

$z_{jk} = \begin{cases} 1 & \text{if observation } j \text{ is in cluster } k \\ 0 & \text{otherwise} \end{cases}$ is an indicator of belonging of an observation to a cluster, providing that z_{jk} takes value *1* only once for assigned cluster, and value *0* for other clusters.

ISODATA (Iterative Self-Organizing Data Analysis Techniques) algorithm: As being a variety of k-means, the ISODATA clustering method uses the minimum feature distance to form clusters to identify statistical patterns in the data. It begins with either arbitrary cluster means or means of an existing signature set, and each time the clustering repeats, the means of these clusters are shifted. The new cluster means are used for the next iteration. The clustering of the data until either a maximum number of iterations is performed, or a maximum percentage of unchanged pixel assignments have been reached between two iterations. The optimal number of classes to specify is usually unknown. The resulting values of centers and their covariance matrix are used to do supervised classification. In his step any of classification methods such as decision trees (dendograms) or ML is used (Mather, 2004; Jensen, 1996).

In case of no solid vulnerability model, clustering gives initial knowledge on distribution of factors/dimensions and estimates of the representative of the class centers. Then, the resultant clusters can be associated with vulnerability classes if prior knowledge is existing such as a risk or vulnerability map.

3. APPLICATION AND RESULTS

Vulnerability and clusters of frequent accident ship locations at Istanbul Strait are explored on ESRI ArcGIS 9.1 and Workstation GRID platform.

The Moran's I and Local Moran's I analysis gives information about global and local autocorrelation between ship accidents with respect to their location and overall accident index. The G statistic and kernel analysis designate hot spots and distribution density of accidents. Vulnerability of places due to accessibility to service locations and weighting this index with population are handled in vulnerability analysis. The factors affecting vulnerability in this study are clustered to reveal patterns by means of clustering methods.

3.1 Completing Missing Data

In warehousing phase; resolving formats, data integration, data cleansing, consistency-checking tasks are accomplished. Significant amount of time is spent on *data cleaning*, the task of detecting and correcting errors in data. Data is validated and corrected by using 'mean' and 'hot deck' imputation methods. A common technique is validating input data by using expert knowledge which provides information about domain of each feature class and item. Following are the decision methods according to measurement levels of data.

- Categorical (Nominal) Data: Set of candidate values and its probability
- Ordinal Data: Range of ordered set and probability of each value
- Interval and Ratio Data: minimum, maximum and mean (median) of previous clean data.

3.2 Calculating Vulnerability of Reaching Services

The proposed vulnerability model suggests the overlap of hazard zone, accessibility to public services and population density. As there is no reliable statistics, available at the present time on losses from disasters at province level caused by each factor, all factors are treated equally in contributions to overall vulnerability.

Firstly, distance surface is calculated by using kernel density considering both closeness and weight of roads within the given radius of each grid. This gives a smoother output grid having 50 m for cell width and 500 m for radius are selected by considering average parcel area and average walking distance to a major road. Secondly, cost of reaching services or secure locations via least-accumulative-cost distance over normalized kernel density surface is calculated. Finally Vulnerability is calculated by weighting the total cost with population density and ratio of each service out of the total.

The locations having very high or high ranks are considered as more vulnerable. The sites with lower values in the scale should not be evaluated as not vulnerable but relatively less susceptible compared to the sites with higher values. Most of the areas around İstanbul Strait are subject to high population vulnerability. Those areas far from Straits have lower vulnerability. The high vulnerable areas are mainly residential areas. However, areas having land use type of industry are around hot spots near by ports. Eventually, the results should be evaluated and reviewed by a domain specialist.

As the population density data is created from boundary polygons, the data is abruptly changes at polygon boundaries due to difference of population densities of census tracks, which is given in turquoise circle in Figure 6.

3.3 Patterns of Accidents and Vulnerability Factors

Moran's I: The calculated and expected Moran's Index value are 0.051833 and -0.006494 respectively, which is not a strong indication of clustering. In Moran Correlogram with 300 m lag distance, after the first bin the autocorrelation value decreases gradually starting from 0.091 to 0.078 and so forth by loosing the slope.

Local Moran's I and Getis-Ord : The locations having darker colours indicate clustering. Clustering occurs at sharp turning points. The zone having higher absolute Local Moran statistic is depicted in Figure 5. There are two locations where hot spot – red color – and one location cold spot – green color – are detected (Figure 5).

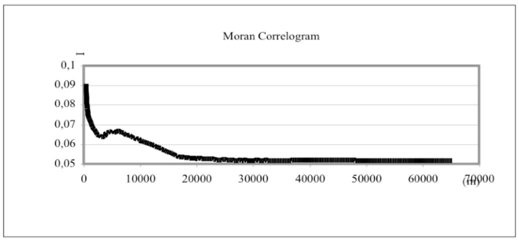

Figure 3. Moran Corelogram at 300 m lags

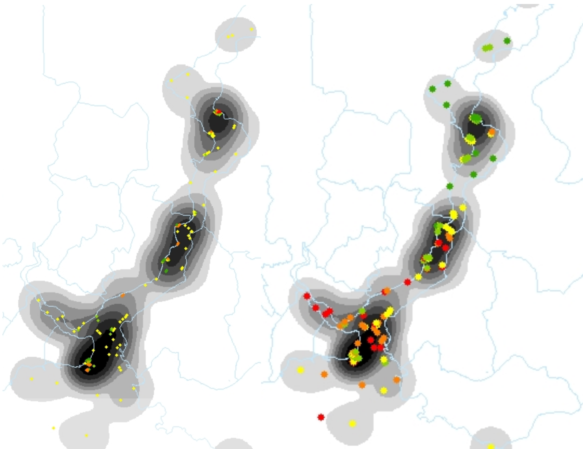

Figure 4. (left)Local Moran's I on kernel classification; Getis-Ord G statistic on kernel classification

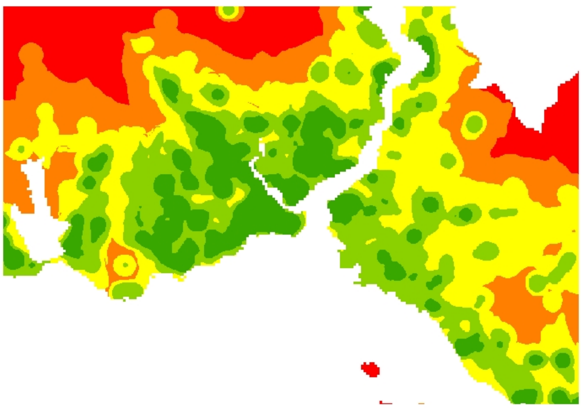

Figure 5. The patterns after clustering factors of vulnerability

Kernel: Selecting kernel width causes a natural bias-variance trade-off as a change on width of the average window. If the window is narrow, its variance will be relatively large. If the window is wide, the variance will be small because of the effect of averaging. Kernel classification of accident index are depicted in Figure 4 and overlapped with vulnerability surface in Figure 6.

k-means: The variables used for clustering ship accidents are location and normalized overall index value. The resultant clusters are visualized in choropleth maps and ellipses around clusters having frequency of 71, 55, 13, 10 and 6. The ellipses are abstractions of the clusters for visual evaluation (Figure 6). Determining number of the clusters is both a strength of k-means technique as well as a weakness. By changing the number, k-means can be used as an exploratory tool to identify possible different partitioning and hot spots.

ISODATA: Factors (census, cost of roads and distance to service locations) are partitioned into 5 clusters and visualized after classification due to class centers and their covariance matrix (Figure 5).

3.4 Results

Figure 6 depicts overlapping areas of vvulnerability of people and frequent ship accident regions. Three accident out of five locations are detected by using spatial data analysis (Local Moran's I, Getis-Ord statistic and kernel density) and clustering (k-means). These are sharp turns of the Strait. Two regions in south are hot spot where severe accidents are populated. Higher vulnerable people and places are surrounding all three locations. Vulnerability of districts (people) are overlaid with density distribution of accidents to reveal most severe locations during a possible accident where cultural heritage is lying.

This model is based on closeness of each cell to service locations by using cost surface of road network. The degree to which populations are vulnerable to hazards, however, is not solely dependent upon proximity to the source of the threat or the physical nature of the hazard – social factors also play a significant role in determining vulnerability (Cutter et al., 2000).

As the population density is derived from a thematic map with administrative unit boundaries, the resultant raster map is not representing the continuous characteristics of density of population distribution. Because of this, abrupt vulnerability changes are calculated at boundaries.

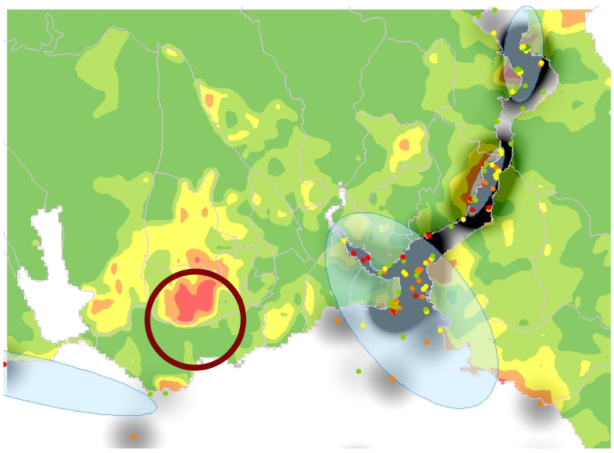

Figure 6. Vulnerability of people is overlapped with kernel density and hot spot analysis

Data quality and data cleaning: In order to create reliable results, high quality data (C^3) is needed. In Figure 6 the area in the circle is used as airport.

The road network for public use does not contain the roads in airports. Therefore these take lower values among the others despite having high accessibility by high quality roads. Although the airport is sparsely populated, because there are rather densely populated residential areas in the same district, this field takes high density values.

Domain knowledge and absence of purpose specific data: Most of the spatial data are collected for general purposes or for some specific purposes. Preparing a data set for analysis requires domain knowledge and other auxiliary knowledge to enrich data. For sparse areas having low road network density takes higher costs such as large undetermined areas and airport. To improve the model restricted cost calculations and precise knowledge is needed.

Accuracy Assessment: No accuracy assessment model is implemented in this study, yet. Use of spatial data mining methods in vulnerability analysis was proposed by several authors. However, there is a need for associating these methods with vulnerability and risk calculation mostly modeled as multicriteria analysis.

4. CONCLUSION

A linear model of vulnerability based on proximity analysis is implemented. Those locations close to main roads and service locations are calculated as low vulnerable areas. Some of high vulnerable areas are airport or parks. Further knowledge and quality data are needed for realistic modeling and reliable results. Social and economic data such as population should be given as parcel based or in iso-arithmic format. Further study will comprise two issues; (1) employing accuracy analysis and introducing social vulnerability factors in the model. After finding vulnerability the risk may be calculated more accurately by using IMM hazard data of building damage and deaths (Erdik, 2002).

Social, statistical and economic data are mainly based on administrative units such as province, sub-province or district which are represented in object models. Conversion of these data to continuous field data always create discrete changes at boundaries. Unit of social data should be small enough to represent continuous characteristics of distribution. It is intuitively known that if population density gets higher the vulnerability increases more than linear equation. Temporal characteristics and uncertainty of data and analysis of ship accidents will be handled in further studies to detect changing characteristics of the accidents and their location in time. On the other hand, vulnerability model for places and people will be enhanced by contribution of social, census and economic having more granularity in addition to hazard maps of IMM to create risk patterns.

REFERENCES

References from Journals:

Cutter, S.L, J.T. Mitchel, M.S. Scott et al. (2000): Revealing the Vulnerability of People and Places: A Case Study of Georgetown County, South Carolina, *Annals of the Assoc. of American Geographers*, 90(4), 2000, p. 713-737

Rashed, T., J. Weeks (2003): Assessing vulnerability to earthquake hazards through spatial multicriteria analysis of urban areas, *Int. J. of Geographical Information Science*

References from Books:

Baily, C.B., A.C. Gatrell (1996): *Interactive Spatial Data Analysis*, Addison Wesley Longman

Bonham-Carter (1994): *Geographic Information Systems for Geoscientists: Modelling with GIS*, Elsevier Science Ltd.

Haining, R. (2005): *Spatial Data Analysis: Theory and Practice*, Cambridge U. Press

Han, J., M. Kamber, A.K.H. Tung (2001-a): *Spatial Clustering Methods in Data Mining, Geographical Data Mining and Knowledge Discovery*, Eds. H.J. Miller, J. Han, Taylor and Francis

Han, J., M. Kamber (2001-b): *Data Mining: Concepts and Techniques*, Morgan Kaufman

Hastie, T., R. Tibshirani, J. Friedman (2001): *The Elements of Statistical Learning: Data Mining*, Inferece and Prediction, Springer-Verlag

Jensen, J.R. (1996): Introductory Digital Image Processing: A Remote Sensing Perspective, Prentice Hall

Longley, P.A, M.F. Goodchild, D.J. Maguire, D.W. Rhind (2003): *Geographic Information Systems and Science*, John Wiley & Sons

Mather, P.M. (1989): *Computer Processing of Remotely-Sensed Images*, John Wiley & Sons

Chawla, S. et al. (2001): Geographical Data Mining and Knowledge Discovery, Eds.: Miller, H.J., J. Han, Taylor & Francis

References from other Literature:

Erdik, M. (2002): Earthquake Vulnerability of Buildings and a Mitigation Strategy: Case of Istanbul, IMM

Erdik M. (2005): Istanbul Master Plan, Bogazici University

Fernandez, J. (2005): Earthquake Master Plan for Istanbul, executive summary, Metropolitan Municipality of Istanbul, 2005

İstikbal, C. (2000): Oil Transportation Through The Turkish Straits:A Great Challenge To Safety, Turkish Maritime Pilots' Association (TUMPA), Annual meeting of MOIG (Mediterranean Oil Industries Group), İstanbul

Kötter, T. (2004): Risks and Opportunities of Urbanisation and Megacities, FIG Working Week 2004, Athens, Greece, May 22-27, 2004

Linnerooth-Bayer, J. (2005): Risk and Vulnerability Program, Research Plan 2006-2010

Muggenhuber, G., R. Mansberger (2004): Spatial Information for Risk Management, 3rd FIG Regional Conference, Jakarta, Indonesia, October 3-7, 2004

Thomas, D.S.K., S.L. Cutter et al. (2002): Use of Spatial Data and Geographic Technologies in Response to the September 11th Terrorist Attack on the World Trade Center

References from websites:

Navulur, K.C.S, Bernard A. Engel (2005): Predicting Spatial Distributions of Vulnerability of Indiana State Aquifer Systems to Nitrate Leaching using a GIS