

The relation between population density and fire / rescue service incident density in urban areas

Jukka Matthias Krisp¹, Věra Karasová¹

¹ Helsinki University of Technology,
Department of Surveying, Cartography and Geoinformatics
PO Box 1200, FIN-02015 HUT, Finland
jukka.krisp@hut.fi
vera.karasova@hut.fi

Abstract. Within this paper the evaluation of the variables for an emergency risk assessment is enhanced by a numerical value describing their spatial relation. The application of a cross tabulation table indicates the relation between two variables. As a sample case we use as inputs the population density raster and its relation to the incident density raster in the administrative area of Helsinki. Improvements in the identification of high or low risk areas can assist the emergency preparedness planning and resource evaluation.

1 Introduction

Within an ongoing project, the Institute of Cartography and Geoinformatics (ICG) at the Helsinki University of Technology (HUT) has been closely cooperating with the Finnish fire and rescue services to explore possibilities to enhance existing risk analysis procedures. Based on the expertise of its employees, the Rescue Office of the city of Espoo has developed a model for risk analysis. The existing model, documented by Ihamäki (Ihamäki, 2000) uses three variables, while in reality the occurrence of accidents is dependent on many more factors. Therefore, the possibility of a more sophisticated model should be investigated.

Visual spatial analysis methods, explored by Krisp (Krisp et al., 2005) among others, can be applied to enhance the finding and representation of an emergency risk assessment. Improvements in the identification of high or low risk areas can assist the emergency preparedness planning and resource evaluation (e.g. for finding the best location for a new fire station). The exploration of the relations between variables can be assisted by a visual analysis that considers the spatial distribution of the phenomena and its relevance to the emergency services (e.g. age distribution, building types, workspace distribution etc.).

If we want to go further in our analysis of finding other variables methods of spatial data mining can be used. Data mining is the process of identifying or discovering useful and as yet undiscovered structure in the data (Fayyad and Grinstein, 2003). Spatial data mining, researched by Koperski, Han, Josselin and Ester among others (Koperski and Han, 1999, Josselin, 2003, Ester et al., 1999) is in the

development stage, so few commercial tools are available. Based on the research made (Karasova and Krisp, 2005) and the first implementations of software prototypes we can see that for example the methods like association rules can be useful. To evaluate the significance of this potential variables we want assess how they match spatially (and perhaps temporally) with the occurrence of incidents, relevant to the rescue services. Statistical and spatial analysis methods can assist this process. They have been investigated in a wide-range concerning the analysis of spatial data by Cressie (Cressie, 1993), Bailey (Bailey and Gatrell, 1995) and more recently by Zhang and Goodchild (Zhang and Goodchild, 2002) among others. Research in this field can distinguish between methods applied to point datasets and methods for raster datasets. Our input data does not allow to apply spatial point data analysis directly, because the population information is an attribute to the individual point buildings (more in 1.1 datasets). Therefore we apply a cross tabulation method to a raster dataset derived from the point data. Within the framework of this paper Figure 1 gives an overview of how the method of comparing two raster datasets by a cross tabulation matrix can be implemented in the process of creating a risk map.

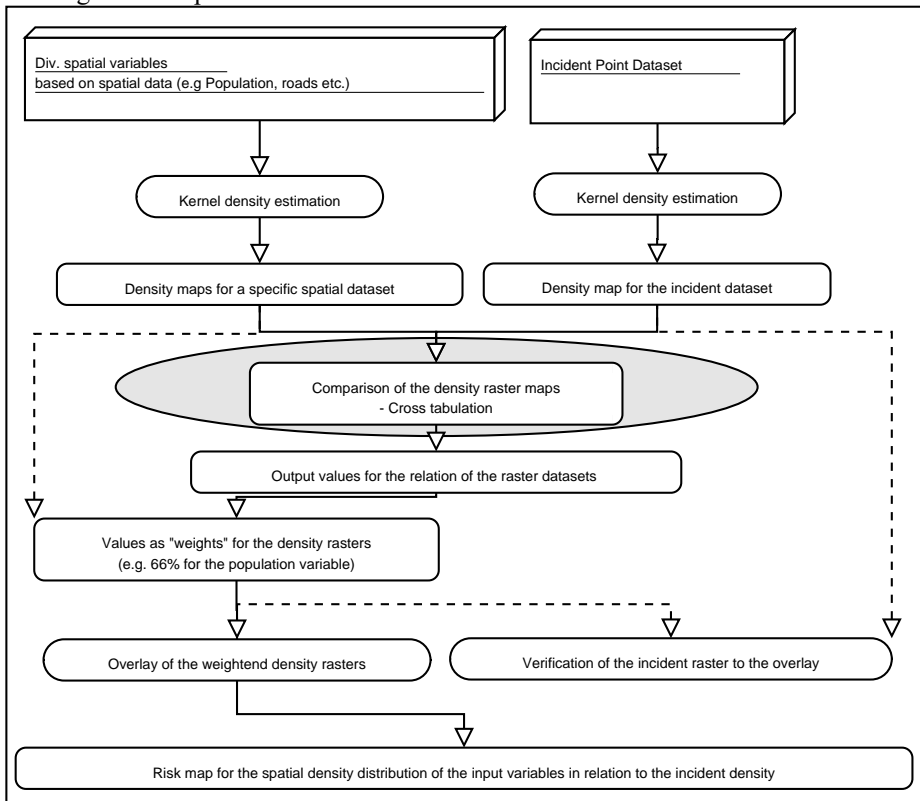


Figure 1. The use of cross tabulation in the risk analysis process

The cross tabulation method can be applied to the raster density maps for different spatial datasets (e.g. population density, traffic density etc.) and incident density

raster. The output value can be used as a weight indicator to distinguish the importance of each specific input variable.

1.1 Population and incident datasets

The Helsinki Fire & Rescue department has provided sample datasets, which contain all the fire alarm, rescue missions and also automated fire alarm systems missions within Helsinki city area for the years 2000 - 2003. The simplified material, illustrated in Figure 2, includes selected attribute information. That makes it possible to make different selections e.g. building block apartment fire missions, in certain time periods, in certain areas. The data is not publicly available.

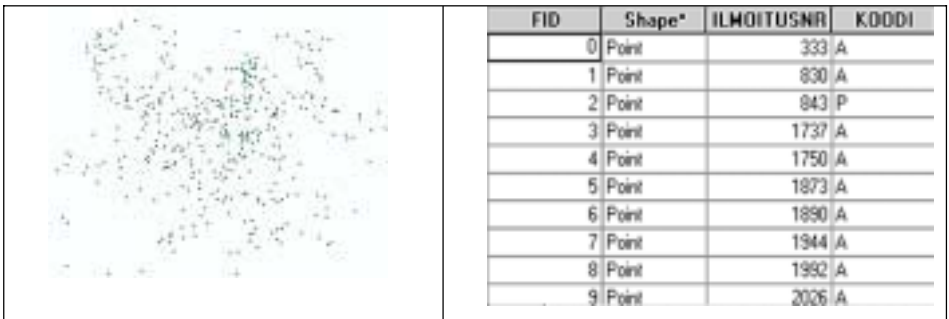


Figure 2. Accident dataset illustration original point file; points for individual incident locations and the data table

The population information in SeutuCD, shown in Figure 3, originates from the population register centre. Their Population Information System contains information for the whole Finland on Finnish citizens and foreigners who are permanently residing in Finland. The population information is stored as a data table, which relates to the individual housing centre points.

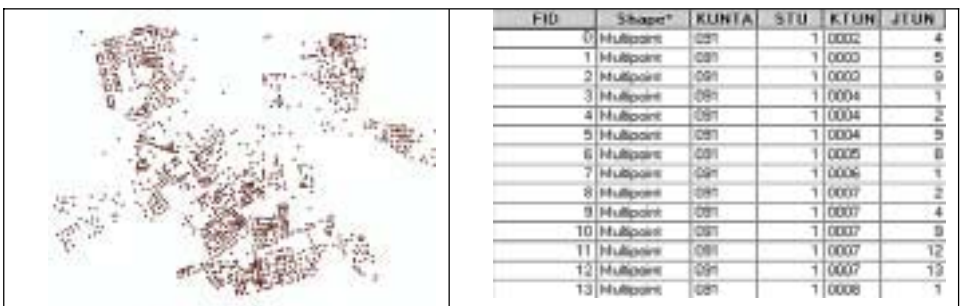
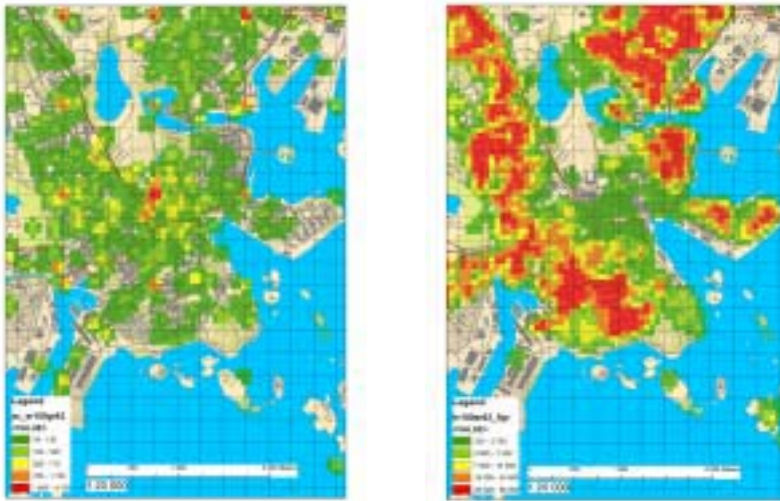


Figure 3. Population dataset original point file; points for buildings with attribute information about the amount of inhabitants in an additional data table

This data is updated five times a week by the population register centre. Completeness omission of the data is said to be 0 %, which we doubt is true. SeutuCD used in this research is updated in 2003 by YTV but the temporal validity of the data has not been tested. Both datasets have been translated into ESRI's shapefile format.

1.2 Density maps for population and incidents

Previous research investigates the application of a density analysis for these datasets, which takes known quantities of some phenomena and spreads it across the landscape based on the quantity that is measured at each location and the spatial relationship of the locations of the measured quantities (ESRI, 2004). In this particular case a kernel density for Finnish data has been applied. The kernel density considers the density of features in a search radius around those features. It can be calculated for both point and linear features. Conceptually, a smooth curved surface is fitted over each point in kernel density for point features. The surface value is highest at the location of the point, and diminishes with increasing distance from the point, reaching zero at the search radius distance from the point. The volume under the surface equals the Population field value for the point, or one if none is specified. The density at each output raster cell is calculated by adding the values of all the kernel surfaces where they overlay the raster cell centre. The kernel function is based on the quadratic kernel function described in Silverman (Silverman, 1986). The population and incident densities are calculated for the Helsinki city centre area. The map has a scale of 1:20.000. Figure 4 shows the population density for Helsinki in a colour coding from green, indicating a low density, to yellow, orange and red, representing a high density.



a. Incident density

b. Population density

Figure 4. a. Incident density b. Population density in Helsinki center area

By comparing the two representations visually we can identify that the areas with a high population density and hot spots for incidents. The visual comparison of the two

datasets indicates that areas with a high population density do in fact not correspond with areas with a high density of reported incidents (Krisp et al., 2005).

1.3 Scope of this paper

The visual comparison between the population density data and the incident data shows that the connection between the population density and the incidents reported by fire and rescue services are not as strong as assumed. Within this paper we apply a cross tabulation table to get a numerical value that describes the relation between the two variables concerning its spatial distribution. As a sample case we use the same inputs (population density raster and the incident density raster) in the whole administrative area of Helsinki.

2 Method – Comparing two raster maps by cross tabulation

To acquire a numerical value we calculate cross tabulation matrix (sometimes referred to as a confusion matrix) to determine the relationship between the density values for all the cells in the two input raster. Generally the method is used to compare the same phenomena in two datasets. In our case we use this method to compare two different variables and the relation of their spatial distribution. Figure 5a and 5b illustrate the method by a nine-cell density map with three classes, A, B, and C.

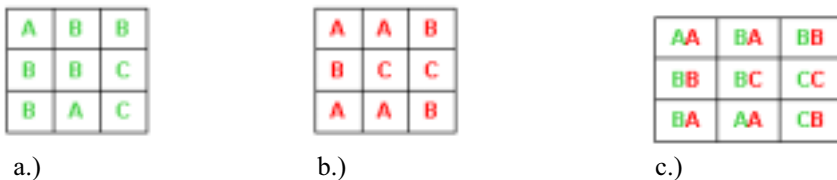


Figure 5. a,b datasets, c Cross tabulation table for three classes

The resulting cross tabulation provides a pixel, by pixel comparison of the interpreted classification types shown in Figure 5c. The upper right cell, the data in the dataset a) is B, and the dataset b) it is also B. Therefore this is a match between the two datasets. The result can be quantified into a cross tabulation (confusion matrix) as shown below in Figure 6a.

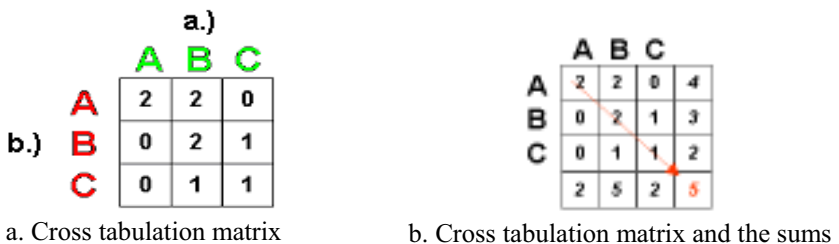
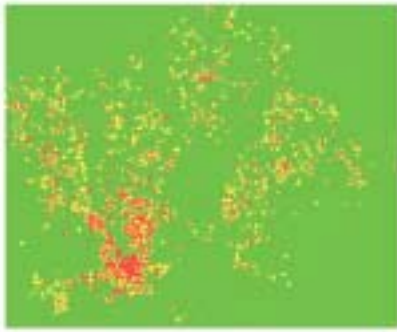


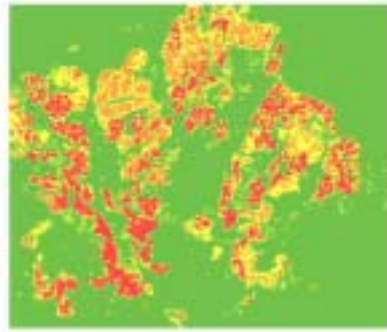
Figure 6. a. Cross tabulation matrix, b. cross tabulation with the sums

The cross tabulation matrix shows the comparison of the two data sets. Features, that are A in the dataset a), and in the dataset b), represent the upper left cell in Figure 6a with the value two. The diagonal, indicated red in Figure 6b represents areas that have not changed. Adding up the diagonal cells indicates, that five cells match. The similarities is a function of: the total cells on the diagonal / total number of cells, $(2 + 2 + 1) / (2 + 2 + 0 + 0 + 2 + 1 + 0 + 1 + 1) = 5/9 = .55\%$ agreement in the sample case above.

The input density raster datasets for the Helsinki area are both classified by using a natural breaks classification. Classes are based on natural groupings inherent in the data. The method identifies break points by picking the class breaks that best group similar values and maximize the differences between classes. The features are divided into classes whose boundaries are set where there are relatively big jumps in the data values. Figure 7 shows the classified dataset for the greater Helsinki area.



a. Incident distribution in Helsinki



b. Population distribution in Helsinki

Figure 7. Grids with the density classification a. incidents b. population

The scale of the sample area is 1:100,000. The cell size of the grid is 100 by 100 units. Classes are displayed in different colours from dark green representing low densities to light green, yellow, orange and red displaying high densities (hotspots) in the values. The actual values do not matter, because we compare only the spatial distribution of the phenomenon in the classified data.

A tabulate area tool is used, which is available as an add-on for the ArcGIS (ESRI, 2002) and calculates cross-tabulated areas between two datasets. Zones are created by identical integer values, so the input is must be must contain integer values. Each record in the output table is stored as the area of each class within each zone. The areas are divided by the cell size to get the count for the individual raster cells. The scripting syntax is: "TabulateArea_sa (in_zone_data, zone_field, in_class_data, class_field, out_table, processing_cell_size)".

3 Results

The output results of the tabulate area function for the tow sample data inputs are shown in table 1. In the calculation the area units and not raster cells are used.

Table 1. Tabulate area results

		Incident Density Classes					
		VALUE	VALUE_1	VALUE_2	VALUE_3	VALUE_4	VALUE_5
Population Density Classes	1	197650000.000	1560000.000	1050000.000	1390000.000	2430000.000	
	2	21390000.000	1170000.000	1300000.000	1060000.000	1750000.000	
	3	21250000.000	1780000.000	1560000.000	1170000.000	900000.000	
	4	19990000.000	2240000.000	1890000.000	1530000.000	1020000.000	
	5	13500000.000	2940000.000	3470000.000	3900000.000	2840000.000	

One cell in the input raster is a 100x100 unit, which equals 10.000 units. To acquire the number of actual raster cells we divide all output values by 10.000. The number of raster cell is shown in table 2. Additionally it shows the cross tabulation sum results for the diagonal cells (highlighted in yellow) and its sum divided by the total number of cells, indicated in red.

Table 2. Cross calculation results

		Incident Density Classes					Percent	
		1	2	3	4	5		
Population Density Classes	1	19765	156	105	139	243	20408	65.7
	2	2139	117	130	106	175	2667	8.6
	3	2125	178	156	117	90	2666	8.6
	4	1999	224	189	153	102	2667	8.6
	5	1350	294	347	390	284	2665	8.6
Percent		88.1	3.1	3.0	2.9	2.9	0.66	

The number of similarities in the two input raster datasets is a function of the total cells on the diagonal divided by the total number of cells. The output for the diagonal cell sum shown in table 3 is 0.658932, which is rounded to a value of 66% an agreement in table 2. The percentages for the summed values are given to show the significance of class 1. The influence of class 1 is determining the result, because of the very high amount of cells in this class.

The cross tabulation values in table 2 can be used in further analysis to determine the relation more specifically, e.g. between high population density cells (class 5) and the low incident areas (class 1) shown in green. The percentage in this case is 4.34 percent. This indicates a stronger relation than the case of a high incident density (class 5) and low population density (class 1), with 0.78 percent shown in blue.

Table 3 summarizes the output information from the input and the cross tabulation process.

Table 3. Input raster information and calculation results

Number of cells (total) - 31073	31073
Sum of cells diagonal (matching," got it right") - 20475	20475
Sum of cells not in diagonal (not matching," got it not right") - 10598	10598
Diagonal cell sum divided by the total number of cells =	0.658932

4 Conclusion and further research

The visual comparison between the population density data and the incident data in previous research had shown, that the relation between the population density hotspots and the incidents hotspots, reported by fire and rescue services, is not as strong as assumed.

It is difficult to confirm this by a cross tabulation matrix, because all raster cells (not only the hot spot areas) are considered in this calculation. The approach shows a relation of 66 percent for all raster cells. This is determined by the large amount of cells in class 1, representing low densities. Further research has to run this process for different classifications to show the relation. We assume a more general classification will give a stronger relation, while more classes will result in a lower relational value. Using different variables and the same classification methods for them into five classes, should give a general indication if the spatial distribution of one variable is stronger or weaker than another. Therefore the case study has shown that via cross tabulation analysis it is to some degree possible to review the significance of a variable in its spatial context for the risk analysis. Furthermore the result of this outcome of a calculation can be compared with a dataset using the same classification distribution, but a completely random spatial distribution.

Population density is still considered a key variable for the vulnerability model. We do not want to give up on the hypothesis that people cause the main risk of an incident and that population is the most valuable variable in a risk analysis. Risk analysis methods are based on the user-defined variables and up to this point we cannot be sure that our assumptions about the variables are correct. Applying analysis methods to other variables might help to determine, which might be more (or less) important in the risk analysis. Generally we can conclude that spatial data analysis is possible through statistical methods like cross tabulation. The ArcGIS software may be used to prepare data for statistical analysis and add-on functionalities provide the functions to run cross tabulation analysis.

Further research has to consider individual incidents recorded in the data and the time of each incident in relation to the time of the population density calculations. The integration of a time variable (daytimes / nighttimes) seems to be essential. Special events (like a hockey game) may result a temporary high risk in a certain area. The spatial development over time may help to develop scenarios for the future (e.g.

changing population density). Furthermore the relationships of objects interact and influence each other. In some cases we might be able to measure this influence. To model and visualize the influence of each spatial object on its environment is a challenge.

2 References

- Bailey, T. C. and Gatrell, A. C. (1995) *Interactive Spatial Data Analysis*, Addison-Wesley, Reading.
- Cressie, N. (1993) *Statistics for Spatial Data*, Wiley, New York.
- ESRI (2002) *ArcGIS 8.1 Online Manual*, Environmental Systems Research Institute Inc.
- ESRI (2004) *ArcGIS 9.0 Manual*.
- Ester, M., Kriegel, H.-P. and Sander, J. (1999) Knowledge Discovery in Spatial Database, in *Proceedings Invited Paper at 23rd German Conference on Artificial Intelligence* (Ed. Institute for Computer Science, U. o. M.) Bonn, pp. 13.
- Fayyad, U. and Grinstein, G. G. (2003) Information Visualization in Data Mining and Knowledge Discovery, in *San Fransisco*, pp. 1-17.
- Ihamäki, V.-P. (2000) Geographic information in the planning of rescue services, *The Emergency Services College, Espoo*.
- Josselin, D. (2003) Spatial Data Exploratory Analysis and Usability, *Data Science Journal*, 2, 100-117.
- Karasova, V. and Krisp, J. (2005) Application of spatial association rules for development of a risk model for fire and rescue services, in *Proceedings ScanGIS Stockholm*.
- Koperski, K. and Han, J. (1999) *Data Mining Methods for the Analysis of Large Geographic Databases*, Simon Fraser University, Burnaby, pp. 4.
- Krisp, J. M., Jolma, A. and Virrantaus, K. (2005) Using explorative spatial analysis to improve fire and rescue services in Helsinki, Finland, in *Proceedings First International Symposium on Geo-information for Disaster Management* Springer, Delft, The Netherlands, pp. 1282-1296.
- Silverman, B. W. (1986) *Density estimation for statistics and data analysis*, Chapman and Hall, London.
- Zhang, J. and Goodchild, M. F. (2002) *Uncertainty in Geographical Information*, Taylor & Francis, New York.