CO-452

GEOVISUAL ANALYSIS OF MASS DATA FOR CIVIL SECURITY PURPOSES

WOLFF M., GONSCHOREK J.
University of Potsdam, POTSDAM, GERMANY

BACKGROUND AND OBJECTIVES

Caused by the spreading of geospatial data into daily life-routines this particular type of date becomes somewhat ubiquitous (e.g. mobile navigation or Google Maps). The increasing availability of (mass-) data and the need for specific information demand complex computational analysis tools and techniques. For a wide range of purposes highly dimensional data needs to be analysed rapidly to discover relationships, clusters and trends. To efficiently discover and visualise such spatial and spatio-temporal patterns, methods from computer science and knowledge discovery are combined with those from cartography and geovisualisation.

This paper presents therefore selected methods and techniques for detecting spatio-temporal clusters in large geospatial datasets. Using the example of emergency services of a major German city's fire department, geostatistical methods are used to detect and map spatial and temporal service clusters. For visualisation and interactive exploration these results are integrated into a two- and three-dimensional geovirtual environment.

APPROACH AND METHODS

Knowing the spatial distribution of emergencies is of significant importance for decision makers e.g. engaged in security- or planning agencies. Literature describes diverse methods to identify and map spatial hotspots of certain events. The easiest method to quickly visualise a set of events is a dot-map. However, considering the limitations given by the number of events and the map's scale, those maps are situated only for a first impression of point distributions. To gain deeper insights into distribution patterns, more sophisticated methods should be applied. Particularly suitable techniques to visualize point densities can be identified within the domain of geostatistics.

The fire department of the city of Cologne records all incoming emergency calls and archive them in a database. Most previous analytical efforts rely on knowledge discovery for strategical decisions like, e.g., defining the position of fire stations, assigning emergency service personnel and upgrading fire control operators. Questions concerning the frequency or spatial and temporal distribution of emergency services in the city of Cologne couldn't be answered.

Applying principles of Geovisual Analytics this work uses methods of GIScience to identify spatio-temporal clusters of fire emergency services. The work targets on the development of efficient techniques to analyse the massive amount of geodata acquired by this large German fire department. For responsible decision makers it is important to be aware of the precise geographic patterns of specific emergency hotspots in space and in time. However, taken the amount of about 103,000 geocoded features into account, efficient methods of knowledge extraction are crucial to gain in-depth knowledge.

To facilitate knowledge extraction from large databases, data mining and exploratory spatial data analysis methods are applied (cf. Haining 2003, Andrienko et al. 2007). Each one of the geocoded features in the underlying database represents one particular case of emergency. In total more than 140 different cases are listed, ranging from accidents and fire emergencies to surgical- and internistic-related emergencies. The emergencies are all related to the German city of Cologne beginning on July, 1st 2007 and ending on June, 30th 2008.

To extract geospatial clusters of different emergency services, different methods of geostatistical analyses are conducted. Exemplarily for the cases of surgical and internistic-related emergencies, kernel-density-estimation (cf. McCullagh 2006, Smith et al. 2006), global indicators for spatial clustering (Nearest Neighbour Index, Moran's I, Getis Ord G), local statistics (Local Moran's I, Getist ord Gi*, Nearest Neighbour Hierarchical Clustering and K-Means) are used. For temporal analysis a chi-square-test is applied. The results are visualised by using primarily 2D techniques and also 3D visualisation techniques. To create first overview maps, however, the traditional pin-map approach can be used.

Figure 1 shows such a simple dot-based visualisation of surgical related emergencies and communicates therefore just a very rough pattern that represents the overall urban framework. To support decision makers with relevant information and to facilitate Knowledge Discovery, efficient methods and tools have to be designed to fit the needs of the potential user. Though depending on the study areas size and the map’s scale this kind of visual representation usually does not give much more insights than a rough impression of how events are distributed in general over the study area.

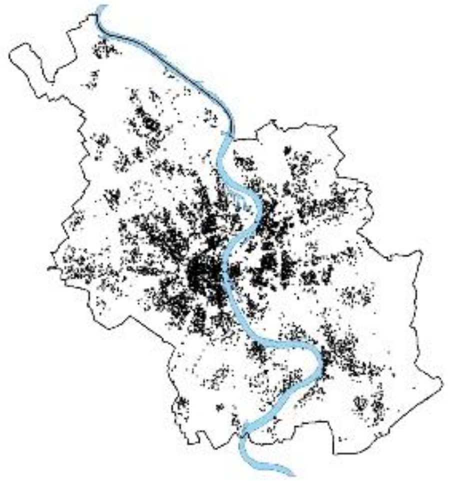

Fig. 1 Dot-based visualisation of surgical-related emergencies in the city of Cologne.

Kernel-density-estimation (KDE) is applied to extract geospatial clusters of emergency services. This method reveals first clusters. To verify these results on a global level, the Nearest Neighbour Index, Moran’s I and Getis Ord G are conducted. On that level all three indexes compare the observed distributions of surgery-related emergencies with a clustered distribution. This analysis confirms the overall cluster-tendency of this dataset. To visualise and proof whether this global observations apply on the local data level as well, Local Moran’s I and Getis Ord Gi* are calculated. The results of this analysis are shown in Figure 2.

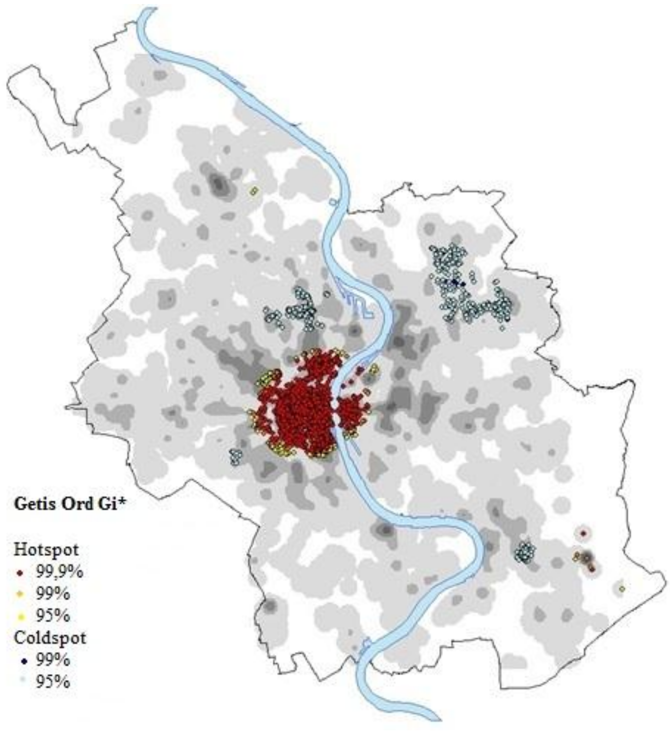

Fig. 2 Hot- and coldspots of surgery-related emergencies at different significance levels according to Getis Ord Gi statistics.*

Very high significant hotspots exist in the city centre of Cologne and in the southeast (airport). The decreasing significance level verifies the underlying KDE-results: the probability value (p) enlarges concentric. Coldspots are identified closed to private residential buildings, local parks and urban forests in the very northeast as well as northern the inner city. To verify these hot- and coldspots Nearest Neighbour Hierarchical Clustering (NNH) and K-Means Partitioning Clustering are applied. The resulting cluster-ellipses encircle the Gi*-hotspots (cf. Fig. 3).

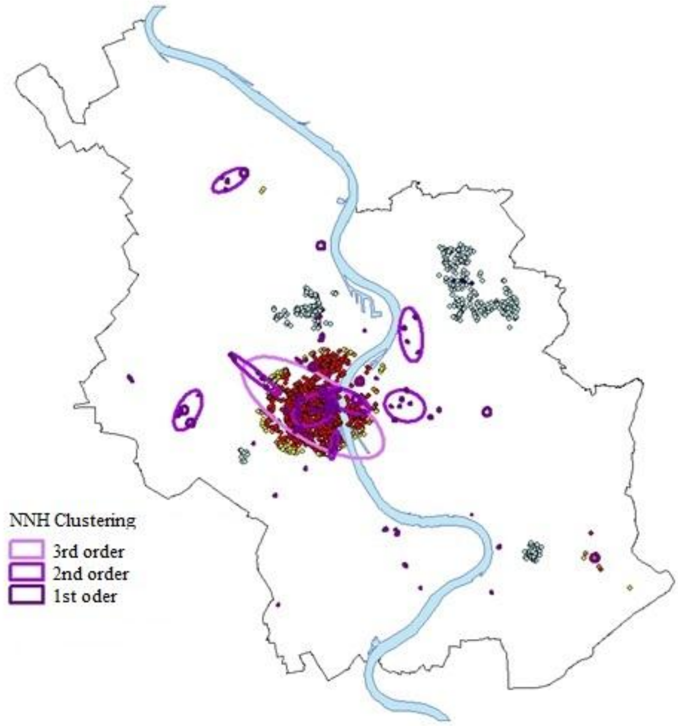

Fig. 3 Location of surgery-related emergency-clusters by using NNH .

In a next step certain analysis results are pipelined from the GIS back-end to a sophisticated 3D visualisation front-end (see Figure 4). Integrated into a geovirtual environment these visualisations are suitable for an intuitive communication of complex geospatial phenomena (cf. Wolff 2010). Furthermore different kinds of services are used to apply techniques for developing a GIS-based graphical user interface (GUI). This GUI is programmed by using the .NET framework and ERSI's ArcObject technology. By applying the ArcEngine SDK a standalone GIS-tool is developed that encapsulates the most important GIS functions. This tool allows, e.g., for selecting a specific emergency service (e.g. fire incidents or surgery services) from the database. Subsequently, geospatial statistics are calculated and visualised by different types of charts. Applying this tool the fire department is able to analyse extensive geodatasets without the need of professional GIS training.

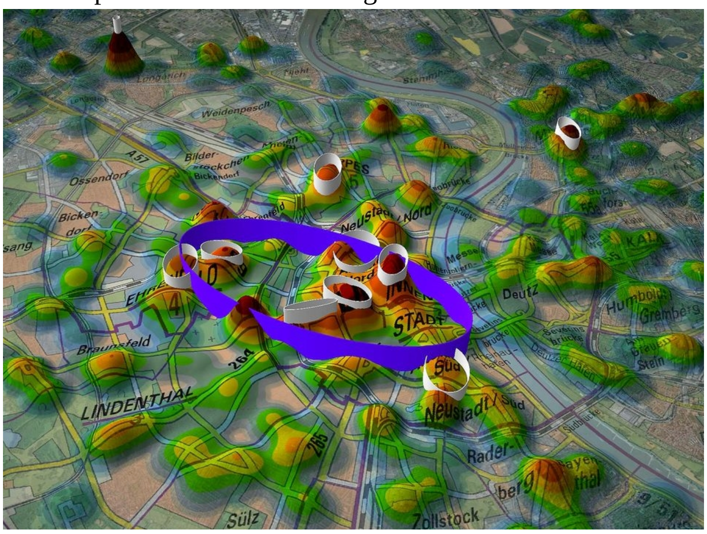

Fig. 4 Integration of NNH cluster boundaries into an interactive geovirtual environment.

However, no temporal trends are revealed so far. The question is therefore, whether these spatial clusters are related to temporal trends and clusters, too. A chi-square goodness-of-fit test is applied to detect such temporal information. Concerning the significance level of $p<0.05$ the hypothesis of equal distribution can be refused for certain weekly- or daily-based periods. Figure 5 shows surgical emergencies differed by days of the week (Friday to Sunday). The bar chart shows that more than fifty percent of all emergencies occur from Friday to Sunday. The expected equal distribution is represented by a straight line. A KDE-grid visualises the spatial distribution according to the seven time frames (days of the week). Obviously the temporal information is reproduced in space: there are permanent spatial clusters in the northern and southern city centre of Cologne and also densities increase from Friday to Sunday.

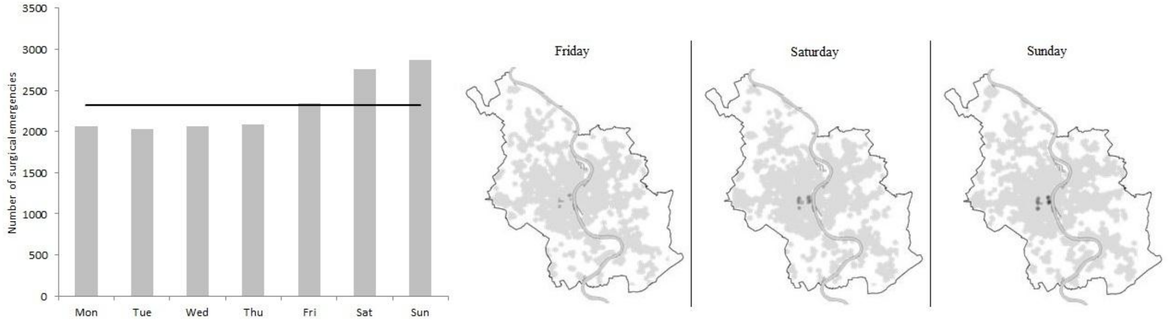

Fig. 5 Temporal and spatial distribution of surgery-related emergencies (days of the week).

RESULTS

Spatio-temporal knowledge concerning specific emergencies is obtained by using the presented methods. For validation different explorative and visualisation-techniques are combined to extract spatial and temporal distribution patterns of emergency services of the city's fire department of Cologne at different spatial scales and different precision. Nearly every space of the city of Cologne was scene of emergencies between July, 1st 2007 and June, 30th 2008. Spatial hot-clusters were identified: in northern and southern inner city, at confusing traffic areas, residential areas in the districts of Meschenich and Muelheim, nearby nursing homes and at outdoor sports and leisure facilities.

CONCLUSION AND FUTURE PLANS

This paper elaborates the potentials of exploratory data analysis for extracting spatio-temporal knowledge from large datasets by the example of a major German fire department. It is shown how surgical emergency services cluster in space and time. Some of the described functions are still under investigation and not yet fully implemented in the graphical user interface. In a next step this graphical user interface will be further developed. Based on the Microsoft .Net framework an application will be developed that allows the user to select a specific emergency case from the database. Afterwards, bar charts are generated automatically. These charts depict the temporal distribution of that emergency case during a specific time period (year/month/week/day). Extensive geostatistical analysis is supported by using ArcObjects (etc. KDE, Getis Ord). To integrate specific analysis methods, functions of the open source statistic-package R will be integrated. For instance, this allows the user to choose between different kernels for kernel-density-estimation analysis. Nevertheless, the presented framework facilitates efficient knowledge extraction from large datasets and provides interfaces for graphic 3D geovisualisations. In summary the integrated methods, workflows and approaches can be characterised as a promising toolkit for civil security analysis purposes.

ACKNOWLEDGEMENTS

Funding of this study by the German Federal Ministry of Education and Research (BMBF) within the framework of the InnoProfile research group '3D Geoinformation' (www.3dgi.de) is gratefully acknowledged. The authors also like to thank the fire department of the city of Cologne for providing extensive emergency datasets. Furthermore the authors give a special thank to the municipality of the city of Cologne for providing geospatial base data. Finally, the authors thank Autodesk, Inc. for supplying the LandXplorer system.

REFERENCES

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M.J., MacEachren, A. and Wrobel, S. (2007). Geovisual Analytics for Spatial Decision Support: Setting the Research Agenda. In: International Journal of Geographical Information Science, 21 (8), Taylor & Francis, pp. 839-857

Haining, R. P. (2003) Spatial Data Analysis: Theory and Practice, Cambridge University Press, 432p.

McCullagh, M. J (2006). Detecting Hotspots in Time and Space. ISG06

Smith, de M.J., Goodchild, M. F., Longley, P.A. (2006) Geospatial Analysis, Troubador Publishing

Wolff, M. and Asche, H. (2010). Towards 3D Tactical Intelligence Assessments for Crime Scene Analysis. In Taniar, D., Gervasi, O., Murgante, B., Pardede, E., Apduhan, B.O., eds, Computational Science and Its Applications - ICCSA 2010. Fukuoka/Japan, March 23-26 2010. Springer Lecture Notes in Computer Science (LNCS), Volume 6016, Proceedings, Part I. Berlin. pp. 346-360.