WEB-BASED EXPLORATORY DATA ANALYSIS (WEB-EDA): VISUALISATION MEETS SPATIAL ANALYSIS

Zhilin LI [1] , Menno-Jan KRAAK [2]

[1]Dept. of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University
Hong Kong, China
Tel: (+852) 2766 5960, Fax: (+852) 2330 2994
lszlli@polyu.edu.hk
[2]Division of Geoinformatics, Cartography and Visualization, ITC
PO Box 6, 7500 AA Enschede (postal), Hengelosestraat 99, 7514 AE Enschede (visiting)
The Netherlands
Tel: (+31 53 4874463), Fax: (+31) 53 4874335
kraak@itc.nl

Commission II, WG II/6

KEY WORDS: Visualization, Spatial analysis, Exploratory data analysis, Web-based exploratory data analysis, GIS

ABSTRACT:

This paper provides an overview of the development in visualisation and spatial analysis. It has been found that, from paper maps to web-based visualisation, the roles of maps have changed from “display and storage” to “display and exploration” and “storage and linking”. The variables for presentations have been expanded from traditional set of visual variables to fives set of variables, i.e. visual variable, screen variables, dynamic variable, exploratory acts and web variables. It has also been found that spatial analysis has also undergone similar transition. The concept of space has been expanded from Euclidean space to other spaces such as social space, perceptual space, topological space, scale space and so on. The analysis methods have also been expanded from traditional statistics to spatial statistics, and to a broader concept including modelling, computation and optimisation methods. However, visualisation and spatial analysis are moving together to form web-based exploratory data analysis (Web-EDA), which focuses on active collaboration between the visual and numeric/statistical analysis. (Indeed, the emerging of Web-EDA also justifies the grouping of these two topics into a working group).

1. INTRODUCTION

Traditionally, the ISPRS Commission II concentrates on the systems for the acquisition of image-based spatial data. No doubt, geographical data acquisition is an important topic because it is often reported that about 80% of the budget of a geo-information project is spent on data. Looking back at the development of data acquisition methodology, the following trends may be recognized (Li and Chen, 2001):

- From single sensor to multi-sensor integration;
- From static mode to dynamic mode;
- From multi-stage process to real-time;
- From manual operation to full automation; and
- From measurement system to monitoring system.

On the other hand, more and more attention is now paid to value-added products while many organisations and companies are still concentrating on spatial database development and hardcopy maps, as pointed out by Prof. Chen Jun, the President of Commission II, in his proposal of Commission II working group structure. He further pointed out that, the need for efficient processing and presentation of geospatial data in a value added form has been fully recognised at the ISPRS Congress held at Amsterdam in 2000, and a recommendation has been made on the development and validation of end-to-end processing systems for specific applications, making use of a range of imaging systems, a range of components from the spatial information sciences and paying particular attention to techniques for the delivery and presentation of geospatial information. As a result, the Commission II has changed its name into "Systems for Spatial Data Processing, Analysis and Presentation" and working groups on analysis and presentation have been established. And this work group, i.e. WG II/6 is dealing with the two important topics of this Commission, i.e. Analysis and Visualisation Systems.

At this inter-congress symposium, it seems pertinent to produce an overview of these topics in this working group. In this paper, the development of visualisation and spatial analysis are reviewed, the state of the art reported and an outlook presented.

2. FROM PAPER MAP TO WEB-BASED VISUALISATION

It is well known that mapping sciences have undergone a dramatic change in the last two decades. The environment for map production has changed considerably. Traditionally, maps were made on paper. On the making process, cartographers design the maps and end users make use of the maps. However, in digital environment, the situation has been changed dramatically. Users of geographical information systems (GIS) and/or computer assisted cartographic systems (CAC) can produce customised products without restriction. Therefore, the distinction between map designers and users diminishes (Su and Li, 1995).

After moving from paper map (hardcopy) to digital (softcopy), then Internet-based visualisation, the functions of maps have also been changed. In analogue form, the map serves for two major roles, i.e. data display (visual communication) and data storage, (although others may also want to add another two: i.e. spatial index and data analysis tool). However, in digital form, data storage and data display have been separated. One could store the data in a database and display the map in a screen or plotted out in analogue form. At the stage of cartographic visualisation, visual analysis is emphasised. Some researchers even try to raise visual analysis to a level equivalent to numeric analysis. Other terms such as visual thinking and visual exploration are also in use (Jiang, 1986; Kraak, 1998; Kraak and MacEachren, 1999). For Internet-based visualisation, data storage is extended, i.e. hyperlink to other sources are available. Such changes are summarised in Table 1 and Figure 2.

To make maps possess such roles, some kind of visual variables need to be employed. It can also be noted that different sets of variables are used at different stages of the development. Table 2 lists the sets of variables in use at such different stages. At digital map stage, theoretically, those screen variables may have been in use although it is widely used at a later stage. At visualisation, dynamic variables have been introduced and exploratory acts are widely used for visual analysis. At a stage of web-based visualisation, more exploratory acts are in use, particularly hyperlink and plug-in. These sets of variables include:

- Visual variables: Size, shape, orientation, hue, saturation, intensity, position, texture, arrangement (Bertin, 1983; Robinson et al., 1995; Jiang, 1996);
- Dynamic variables: duration, rate of change, order (DiBiase et al, 1992);
- Exploration acts: Drag, pan, zoom, click, blink, highlight (Jiang, 1996), mouse over,
- Screen variables: Blur, focus, transparency (Kraak and Brown, 2001); and
- Web variables: Browse, plug in.

Stage of Development	**Major Roles of Maps**			
Paper Maps	Visual communication & data storage			
Digital maps	Visual communication		Data storage	
Cartographic Visualisation	Communication	Exploration	Data storage	
Web-based Visualisation	Communication	Exploration	Data storage	Data linking

Table 1. Change of map roles in different environments

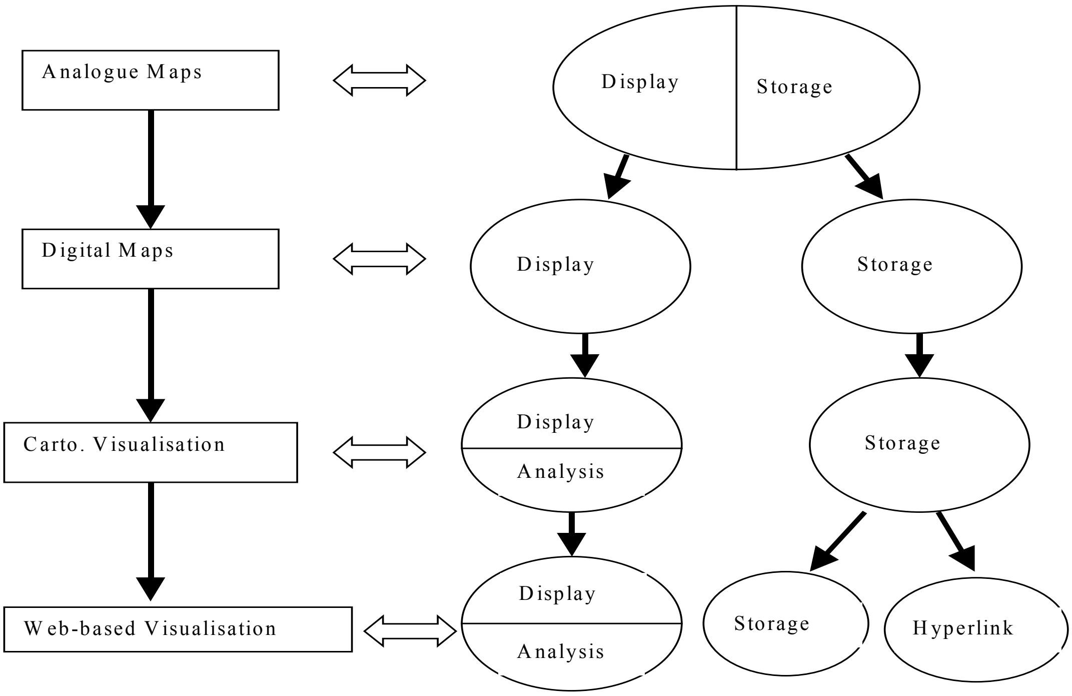
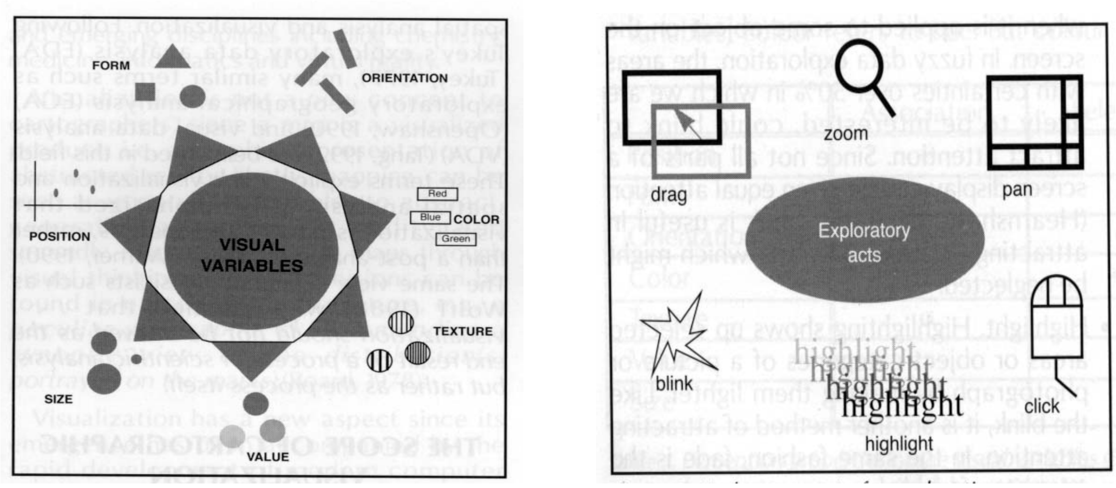


Figure 1. Roles that maps played in different stages of development

Map form	**Variables in use**				
Paper Maps	Visual variables				
Digital maps	Visual variables	Screen variables			
Cartographic visualisation	Visual variables	Screen variables	Dynamic variables	Exploration acts	
Web-based visualisation	Visual variables	Screen variables	Dynamic variables	Exploration acts	Web variables

Table 2. Variables in use at different stage of development



(a) Visual variables for communication (b) Exploratory acts for analysis

Figure 2. Variables for cartographic visualisation (Jiang, 1996)

Indeed, with web as a medium, the traditional cartographic variables as proposed by Bertin (1983) have been extended popularly. Web design software enables the full use of new variables such as blur, focus, and transparency. Blur gives symbols a fuzzy appearance and can for instance be applied to visualise uncertainty, while focus will introduce blinking symbols to attract attention. Transparency can be seen as a kind of fogginess, by which part of map content is obscured or faded in favour of other information. For example it can be applied to subdue the background in a map in order to enhance the main theme in the foreground (for instance a drape of geology over terrain features). Both transparency and shading/shadow can be used to simulate a three-dimensional look. In a three-dimensional "landscape" environment it can also be used as a depth cue. The use of shadow and shading increase the sense of depth. Shading is commonly used to increase the contrast between "figure" and "ground" or, for instance in relief maps to create a three-dimensional terrain impression. Shadow, also known as cast or drop shadow, can applied to give the symbols a three-dimensional look. In web maps this three-dimensional feel of the symbols invites user to click on them to activate a hyperlink or mouse-over effects. The visual effect of shadow is casting a shadow of the symbol on to the background.

Apart from the browsing, another particular interesting feature of web-based environment is the plug-in function. Virtual reality technique could be integrated into web-based visualisation, via JAVA, JAVA-Script and/or VRML.

3. FROM STATISTICAL ANALYSIS TO GENERALISED SPATIAL ANALYSIS

"Broadly speaking, spatial analysis can be defined as the formal quantitative study of phenomena that manifest themselves in space. This implies an attention to location, area, distance and interaction." (Anselin et al., 1993).

Spatial analysis has been a popular topic in geography since the 1980s. Real spatial analysis is viewed as deeply rooted in spatial statistics. Quantitative geographers devise statistical methods for analysis of geographical phenomena. In the early 1990s, the US National Center for Geographic Information and Analysis (NCGIA) has also had an initiative entitled "GIS and Statistical Analysis" (Rogerson and Fotheringham, 1994; Fotheringham and Rogerson, 1994), compared to the initiative on "Visualisation of Spatial Data". However, there is a contrast between the results of these two initiatives. The advance in visualisation is very fast but the progress in spatial analysis is relatively slow. Possibly this is because the contribution to visualisation comes from various communities such as geography, photogrammetry and remote sensing, GIS, computer science and cartography, while the contribution to spatial analysis was mainly from geographers and GISers. Ironically, the contribution to spatial analysis from statistical community was limited. This can be evidenced by the story of Stan Openshaw's geographical analysis machine (GAM).

The first version of his geographical analysis machine (GAM/1) was developed to analyse child leukaemia data in Northern England (Openshaw et al., 1987). The machine successfully spotted the suspected Sellafield cluster easily and it also found an even stronger major new cluster: the Gateshead cancer cluster in 1986. This is possibly the only instance of a major cancer cluster being found by computer analysis, as claimed by Openshaw (1998) himself. However, GAM/1 was severely criticised by some statisticians (Alexander and Boyle, 1996), although it was praised by many geographers as a major development in useful spatial analysis. These critics have been approved being wrong at a later event. The International Agency for Research on Cancer (IARC) commissioned a study in 1989-91 of several clustering methods, with many developed by critics of the original GAM. Openshaw reported: "Some 50 synthetic cancer data sets were created for which the degree of clustering and locations of clusters were known but kept secret. The data were given to the participants who performed their analyses without any knowledge of the correct results. *Nine different methods* were investigated. It was anticipated that the statistical methods preferred by the critics of GAM would work best but, much to the surprise of its original critics, the GAM was shown to be the best or equivalent best means of *testing for presence of clustering* and for *finding the locations of clusters*".

Indeed, Openshaw (1998) severely criticises current situation in spatial analysis: "The spatial analysis situation in the late 1990s is little improved on that of 10 years previously! ……. Most existing spatial analysis methods are either not useful or usable in GIS environments. The problem is deeply rooted and needs urgent and dramatic action. So throw away many of the old statistical books that assume *random sampling* of *non-spatial data*. Junk most of the quantitative geographic methods that assume hardly any data or use inappropriate and unsafe inferential significance tests". Openshaw (1998) directs attentions to "the more heavily duty statistical methods that rely on mathematical proofs which are at the expense of empirical performance and geographical common sense".

Indeed, Openshaw is not alone in criticising the slow progress in spatial analysis. Marble is another noticeable critic. In December 1998, a "Workshop on Status and Trends in Spatial Analysis" was held in Santa Barbara. The workshop brought together some of the most prominent spatial analytic researchers in the world. A special issue resulted from this workshop was published in *Journal of Geographical Systems* as the first issue of Year 2000 (Getis, 2000). At that workshop, Marble (see Marble, 2000) emphasises that some of those *modelling* functions should also be considered as spatial analysis to broaden the concept. The example he mentioned is the transportation optimisation models."

Another authoritative researcher in the area -- Anselin (2000), also suggest that spatial analysis needs to go beyond dealing with Euclidean space and physical geographical locations to include location in "social" and "perceptual" space (social distance, economic distance).

As a result of the above discussions, the development of spatial analysis can be summarised into Table 3. It is meant that the traditional statistics assumes random sampling of non-spatial data. Here, non-spatial data means that the locations of the samples are not considered. In other words, the two characteristics associated with spatial data, i.e. spatial dependency and heterogeneity, are not considered in such traditional statistics. The "other spaces" means the "social" and perceptual" spaces, as suggested by Anselin (2000), as well as topological space, scale space, and so on. The "other spatial operations" means some of the optimisation methods, modelling function and computation methods, such as networking models.

4. WEB-BASED EXPLORATORY DATA ANALYSIS (WEB-EDA): VISUALISATION MEETS SPATIAL ANALYSIS

It can be noticeable that both visualisation and the spatial analysis are simultaneously moving into Web-based exploratory data analysis (Web-EDA), which essentially consists of three a new topics as follows:

- Exploratory data analysis;
- Integration with GIS technology; and
- Web-based development.

Exploratory data analysis (EDA) focuses on active collaboration between the visual and the numeric (Anselin, 2000). Data mining relies on both visual and numeric techniques in search for anomalies, outliers and patterns has been regarded as an example (Goodchild, 2000).

With the integration with GIS and other technology, a new term "Geovisualisation" has become popular in the visualisation community. Indeed, Geovisualisation integrates approaches from scientific visualisation, cartography, image analysis, information visualisation, exploratory data analysis (EDA), and geographic information systems (GIS) to provide theory, methods, and tools for visual exploration, analysis, synthesis, and presentation of geospatial data (MacEachren and Kraak, 2001). Primary themes addressed in Geovisualisation are representation of geospatial information, integration of visual with computational methods of knowledge construction, interface design for Geovisualisation environments, and cognitive/usability aspects of Geovisualisation.

Similarly, a new term "GeoComputation" has also become popular in spatial analysis. GeoComputation is defined as the bringing together of spatial databases, high performance computers, and AI-computational intelligence technologies to solve practical problems (Openshaw, 1998). GeoComputation emphasises new computationally intelligent technologies for GIS-based spatial analysis.

For the development of GIS-based spatial analysis methodology, Openshaw (1998) has also established a set of GISability criteria as follows:

- Relevant to GIS data environments?
- Sensible given the nature of GIS data?
- Reflect likely end-user needs?
- Compatible with the GIS style?
- Capable of being used by real users?
- Adds value to the GIS investment?
- Promises tangible and significant short-term or instant benefits?
- Safe and reliable?
- Capable of being understood by non-experts?
- Easily applied?

It has been expected that there will be an explosion in the use of the web to broaden public access to GIS data and systems as means of publicity, awareness raising, propaganda, manipulation, and conversion of public attitudes. Opensahw in 1998 predicts "Meanwhile developments in Java and heterogeneous GIS will undoubtedly create new opportunities for distributed spatial analysis machines on the Internet in the near future". We may soon send standard spatial analysis tasks to a particular web site for computation and receive the results back sometimes later. The web site could be in Beijing, Tokyo, London, New York or somewhere else. Virtual spatial analysis machines will soon come to exist, accessible over the Internet, each of which offers and performs only one type of widely applicable analysis.

Stage of Development	Methods for use		Spaces	
Statistical Analysis	Statistics		None	
Spatial Statistical Analysis	Spatial statistics		Euclidean Space	
Spatial Analysis	Spatial statistics		Euclidean Space	Other spaces
Generalised Spatial Analysis	Spatial statistics	Other spatial operations	Euclidean Space	Other spaces

Table 3. Spatial analysis at different stage of development

Eventually, a new business called "spatial analysis service" may be formed similar to the so-called "location-based service". Similarly, visualisation has also moved onto the web. The significant achievement is marked by the publication of two books on web-based visualisation. One is entitled Web Cartography (Kraak and Brown, 2001) and the other is entitled Mapping Cyberspace (Dodge and Kitchin, 2000). As has been discussed previously in Section 3, new variables have been introduced and new technology employed.

5. CONCLUSIONS

Visualisation and spatial analysis are moving together to form a new discipline called exploratory data analysis. The trend is that such EDA will be carried in GIS environment and on web. Indeed, web-based exploratory data analysis (Web-EDA) will play a central role in the future, which enable users to perform distributed RMSE service. Here the "RMSE" is coined by (O'kelly, 2000), meaning reliable answers to meaningful questions about significant problems, as evidenced by spatial data. And the Web-EDA methods should be automated, user-friendly, robust, efficient, and effective (AUREE)! (This trend also justifies the existence of this working group -- Visualisation and Spatial Analysis Systems).

As has been discussed previously, concepts of spaces should be extended to Euclidean space to social, perceptual, topological, scale, thematic space and so on. It might be the time to include temporal space as well.

For future directions in these areas, some research agenda have been developed already. For spatial analysis, a comprehensive agenda has been set by Openshaw and Fischer (1995). For visualisation, MacEachren and Kraak (2001) have addressed a number of issues.

REFERENCES

Alexander, F. E. and Boyle, P., 1996. *Methods for investigating localised Clustering of Disease*, IARC Scientific Publications No 135, Lyon, France.

Anselin, L., 2000. GIS, spatial econometrics and social science research. *Journal of Geographical Systems,* 2, pp.11-15.

Anselin, L., Dodson, R. and Hudak, S., 1993. Linking GIS and Spatial Data Analysis in Practice. *Geographical Systems*, 1(1), pp. 3-23.

Bertin, J., 1983. *Semiology of Graphics*, The University of Wisconsin Press. Wisconsin.

CHEN, J., 2000. *Proposed Commission II Working Group Structure*. Report to ISPRS Council. 7pp.

DiBiase, D. W., MacEachren, A. M., Krygier, J. B. and Reeves, C., 1992. Animation and the role of map design in scientific visualisation. *Cartography and Geographic Information Systems*, 19, pp.201-214 & 265-266.

Dodge, M. and Kitchin, R., 2000. Mapping Cyberspace. *Routledge*, 296pp.

Fotheringham, A. S. and Rogerson, P. A. (eds.), 1994. *Spatial Analysis and GIS*, Taylor & Francis, London. 281pp.

Getis, A., Spatial analysis and GIS: An introduction. *Journal of Geographical Systems,* 2, pp. 1-3.

Goodchild, M., 2000. The current status of GIS and spatial analysis. *Journal of Geographical Systems,* 2, pp. 5-10.

Jiang, B., 1996. Cartographic visualization: Analytical and communication tools. *Cartography*, 25(2), pp. 1-11.

Kraak, M.J., 1998. The cartographic visualization process: from presentation to exploration. *Cartographic Journal,* 35(1), pp.11-16.

Kraak, M.J. and Brown, A., 2001. *Web Cartography: Developments and Prospectus*. Taylor & Francis, London. 231pp.

Kraak, M.J. & A.M. MacEachren (1999) Visaulization for exploration of spatial data. *International Journal of Geographical Information Sciences,* 13 (4), pp. 285-287.

LI, Zhilin and CHEN,Y.Q., 2001. Current Trends in Geographical Data Acquisition, In: Chen, Y.Q. and Lee, Y.C. (eds.), *Geographical data acquisition*, Springer-Verlag, Vien, pp.251-160.

MacEachren, A. M. and M. J. Kraak, 2001. Research challenges in geovisualization. *Cartography and Geographic Information Systems*, 28(1), pp.3-12.

Marble, D.F., 2000. Integration of spatial analysis and GIS. *Journal of Geographical Systems,* 2, pp. 31-35.

O'kelly, M., 2000. GIS and educational and institutional challenges. *Journal of Geographical Systems,* 2, pp.23-29.

Openshaw,S.,1998. Some trends and future perspectives for spatial analysis in GIS. *Geographic Information Science*, 4(1-2), pp. 5-13.

Openshaw, S. and Fischer, MM., 1995. A framework for research on spatial analysis relevant to geo-statistical information systems in Europe, *Geographical Systems*, 2, pp.325-337.

Openshaw, S., Charlton, M., Wymer, C. and Craft, A.W., 1987. A mark I geographical analysis machine for the automated analysis of point data sets, *International Journal of Geographical Information Systems*, 1, pp.335-358.

Robinson, A.H., Morrison, J.L., Muehrcke, P. C., Kimerling, A. J. and Guptill, S. C., 1995. *Elements of Cartography* (6th edition), John Wiley & Sons, Inc. 674pp.

Rogerson, P. A. and Fotheringham, A. S., 1994. GIS and spatial analysis: Introduction and Overview. In: Fotheringham, A. S. and Rogerson, P. A. (eds*.), Spatial Analysis and GIS*, Taylor & Francis, London, pp.1-10.

SU, B. and LI, Z.L., 1995. Expert system for automated map design and production: A review and some considerations. *Cartography*, 24(1), pp.33-42.