Interactive and animated visualization of highway air pollution

EPA grant # XA 97268501

Feng Qi

Department of Geology and Meteorology

Kean University

1000 Morris Ave., Union, NJ 07083, USA

Email: fqi@kean.edu

Francisco Artigas

Meadowlands Environmental Research Institute

New Jersey Meadowlands Commission

One DeKorte Park Plaza, Lyndhurst, New Jersey 07071, USA

Email: Francisco.Artigas@njmeadowlands.gov

Interactive and animated visualization of highway air pollution

Feng Qi
Department of Geology and Meteorology
Kean University
1000 Morris Ave., Union, NJ 07083, USA
Email: fqi@kean.edu

Francisco Artigas
Meadowlands Environmental Research Institute
New Jersey Meadowlands Commission
One DeKorte Park Plaza, Lyndhurst, New Jersey 07071, USA
Email: Francisco.Artigas@njmeadowlands.gov

Abstract

Background: Vast amounts of data are being collected by environmental protection agencies (i.e. the EPA in the U.S.) and through funded research projects to monitor and model environmental qualities at various scales. The mere presentation of this vast amount of data, however, makes little sense to the general public and even sometimes the project funding agencies. It would be desirable to reveal patterns among the data in terms of spatiotemporal variations of the environmental quality. Furthermore, if the human impacts on environmental quality can be revealed and presented in an interesting and easily understandable form, it would facilitate the communication of such information to the public, promote awareness of human impacts on the natural environment, and encourage protection activities.

As an example, air quality data has been collected along a transect from the heavily trafficked New Jersey Turnpike (NJTPK) in NJ, USA to monitor traffic-related vehicle emissions. Concentrations of a variety of Hazardous Air Pollutants (HAP) were measured routinely from September 2007 to September 2008. The HAPs measurements include particulate matters, trace metals and specific polycyclic aromatic hydrocarbons (PAHs). The monitoring stations are located close to a toll plaza on the turnpike through which detailed traffic information is available. Concurrent information on weather conditions was also recorded. Therefore, it would be possible to reveal patterns among the data in terms of spatiotemporal variations of air quality along the highway. Such spatiotemporal patterns may include seasonal cycles of air quality change and decay over distance from the highway, and relationships between air quality and traffic and/or weather conditions. We developed an interactive and animated visualization interface to facilitate knowledge discovery and present the findings to common audiences.

Objectives: The objective of this study is to develop and test a visualization tool for geographic knowledge discovery on the relationships between weather and traffic conditions

and spatiotemporal variations of air quality along the NJTPK. The specific goals include: 1) Design and test an interactive and animated visualization tool for displaying spatiotemporal patterns of highway air quality; 2) Examine and assess exploratory spatial data analysis (ESDA) techniques in terms of their suitability and efficiency in revealing the patterns and presenting the information to common audiences; and 3) Use this visualization tool to facilitate the extraction of relationships between highway air quality and weather and traffic conditions, and present the findings to the general public through a web-based interface to promote awareness of air pollution in relation to human factors, and encourage participation in protection activities.

Methodology:

Visualization We developed an animated and interactive interface, within which a number of exploratory data analysis (EDA) techniques were implemented and tested to facilitate the visual detection of spatiotemporal patterns of highway air pollution. These include temporal focusing, temporal re-expression, visual benchmarking, and cross-variable referencing, among others. Animation enables common users to visualize the change of air quality over space and time. With temporal brushing, the user is not only a passive viewer of the information, but can interact with the animation and learn actively. With temporal re-expression through multi-scale data aggregation, the cycles of air quality change at differential temporal scales can be revealed. With visual benchmarking, different air quality measures are correlated visually. With cross-variable referencing, the relationships between air quality and factors such as traffic and weather conditions can be direct visualized.

Knowledge Discovery Geographic knowledge discovery refers to the extraction of spatial relations or patterns not explicitly stored in spatial databases. Visualization plays a pivotal role in geographic knowledge discovery by utilizing the powerful human visual perception for pattern recognition. Our visualization tool helps to identify air quality anomalies, establish relationships between air quality and other variables, and facilitate general modeling of high way air pollution transmission. A common knowledge discovery procedure consists of four major steps: data preprocessing, pattern extraction, knowledge consolidation, and information presentation. Our visualization interface was incorporated into the knowledge discovery procedure with all four steps: it plays pivotal roles in both the data preprocessing and information presentation steps, and auxiliary roles in the pattern extraction and knowledge consolidation steps, where numerical algorithms are the key components.

Results: The visualization was tested by three user groups: one with expert researchers who are familiar with air quality measures and two with common audiences that include high-school and introductory level college students. Among the EDA techniques experimented, temporal re-expression was found to be highly efficient in revealing temporal patterns of air quality at different temporal scales. Cross-variable referencing was found to be an effective tool to reveal relationships among air quality measures and weather conditions, which dynamic benchmarking proved to be less effective in showing co-variation of multiple air quality measures.

The visualization proved to be an effective tool to facilitate geographic knowledge discovery. In the data preprocessing step, anomalies in the data was identified by visualizing the measured values of air quality variables through controlled animation. Relevant weather and traffic features were identified by visualizing the change of air quality in relation to the recorded weather and traffic conditions including wind speed, wind direction, humidity, traffic volume, etc. The relevant variables were used in the pattern extraction step for the establishment of statistical models. In the next step, the relationship model was then used to interpolate observed data records and results were visualized with the visualization tool through user interaction. Specifically, user may control the change of certain variables (e.g. wind speed, traffic volume, etc.) to perceive the predicted change in affected air quality measures. At last, the real data, together with the relationship model of spatiotemporal patterns is presented to common audiences with the visualization tool to promote awareness of air pollution in relation to human factors, and to encourage participation in protection activities.

Conclusion: This study aims to address the challenge of representing complex dynamic geographic processes to facilitate geographic knowledge discovery and the communication of scientific discoveries on such processes to the general public. The simple visualization developed in this study proves to be an effective tool for such purpose. The visualization tool for communicating information on highway air pollution to the project sponsors and the general public has apparent advantage over the traditional presentation methods with numbers and statistical results. Animation is an intuitive and appealing way to display complex geographic processes in a more appealing manner with a decreased level of abstractness. This tool also has advantages over other visualization methods such as charts, a static map, or a simple animated map in that it is interactive. Users are not restricted to the passive acceptance of information presented to them, but can interact with the tool to engage in active thinking and learning.

Introduction

"Protecting the environment is everyone's responsibility, and starts with understanding the issues."(EPA 2009) – As the Environmental Protection Agency (EPA) of the United States has put clear in its mission statement, encouraging public awareness and participation is among the most important approaches to achieving its simple mission "to protect human health and the environment"(EPA 2009). Numerous academic and community projects are being sponsored by the environmental agency each year to monitor and model environmental qualities at various scales. Vast amounts of data are being collected, often routinely, stored in databases, analyzed by scientists, and presented to project sponsors, policy makers, and sometimes the general public. A large portion of the collected environmental quality information is time-series data and is spatial in nature. To communicate information on the spatiotemporal pattern of environmental quality represented by such data, scientists have a long tradition of presenting results in statistical analysis reports or fitted numerical models. The potential audience for the results of such environmental projects, however, is no longer limited to scientists but now also includes the public, policymakers, the media, and a host of others. Reporting complex spatiotemporal environmental modeling results to policymakers and the public is challenging and sometimes daunting.

We argue that an effective communication to such audience should 1) help the audience identify patterns in the data, 2) engage the audience in active learning of such patterns, 3) encourage such learning experience, or in other words, make the learning experience enjoyable. In this sense, if the variation of certain environmental quality and human impacts on such quality can be revealed and presented in an interesting and easily understandable form, it would facilitate the communication of such information to the public, promote awareness of human impacts on the natural environment, and encourage protection activities.

As an example of such EPA-sponsored community environmental projects, year long air quality data has been collected along a transect from the heavily trafficked New Jersey Turnpike (NJTPK) in the United States. Concentrations of a variety of Hazardous Air Pollutants (HAP) were measured that include particulate matters, trace metals and specific polycyclic aromatic hydrocarbons (PAHs). The monitoring stations are located close to a toll plaza on the turnpike through which detailed traffic information is available. Concurrent information on weather conditions was also recorded. Therefore, it would be possible to reveal patterns among the data in terms of spatiotemporal variations of air quality along the highway. Such spatiotemporal patterns may include seasonal cycles of air quality change and decay over distance from the highway, and relationships between air quality and traffic and/or weather conditions.

Visualization has been proved to be an effective tool for presenting results of spatiotemporal analyses of urban air pollution (Koussoulakou 1994). Human visual perception offers excellent pattern recognition capabilities that facilitate knowledge construction and the detection of spatiotemporal relations (Gahegan et al. 2001). Geo-visualization is an effective tool to explore geographic data and communicate geographic information (MacEachren and Monmonier 1992). Our study employs some of the most straight forward tools in this tool kit for the purpose of communicating information to the most common audiences (i.e., school students, non-expert project sponsors, etc.) and hopes to facilitate geographic knowledge discovery not only by expert statisticians or 'data miners', but also these common audiences.

Visualization

The visualization interface we developed is both animated and interactive. A number of exploratory data analysis (EDA) techniques were employed to facilitate the visual detection of spatiotemporal patterns of highway air pollution by common audience. These include temporal focusing (MacEachren et al. 1997), temporal re-expression through multi-scale data aggregation (Dibiase et al. 1992, Harrower 2007), and visual benchmarking (Harrower 2002), among others. In terms of symbolization, simple proportional circles were used to depict intuitively 'air bubbles' that represent concentrations of air quality along a transect from the highway. The air bubbles use bright saturated colors for an alerting effect of harmful pollutants. Meteorological variables and traffic measures all use simple linear symbols that change synchronously not to overwhelm viewers' perceptual space.

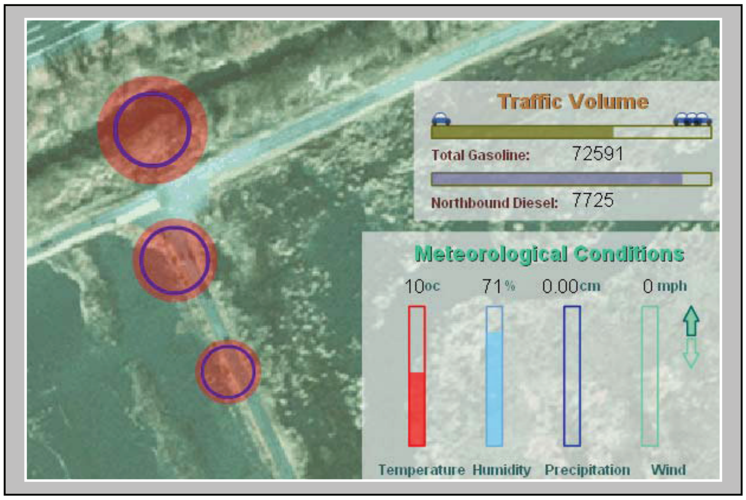

Figure 1. Symbolization of the air quality, traffic, and meteorology variables.

Animation

A number of strategies have been studied and suggested to visualize geographic time-series data (Monmonier 1990), among which cartographic animation is the most straight forward approach. Cartographic animation uses time to represent time and avoids data transformation and abstraction from temporal to any other format. This allows audience to directly observe the dynamics of air quality change. In our interface, the pollutant bubbles change size along the linear time line while whether and traffic bars change length simultaneously (Figure 1).

User Interaction

Animation is a natural strategy to present time-series data. However, our visual memory is short-term and simple animation is not very useful for displaying large amount of data, especially when complex associations existing among the data (Harrower 2007). Therefore, user interaction is necessary to avoid overloading users and limiting their ability to learn patterns and relationships. Our interface provides basic play controls and a draggable timeline. With such controls and temporal focusing, the user is not only a passive viewer of the information, but can interact with the animation and learn actively (Figure 2).

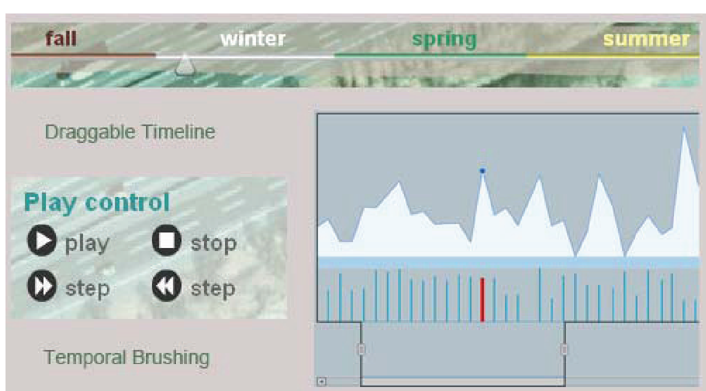

Figure 2. User controls of the visualization environment

Visual Benchmarking

With fixed visual benchmarking, the air quality level at any recorded time spot can be compared visually with health standards to give the viewer a direct alert on how high air quality is affecting human health. We experimented with dynamic benchmarking in our interface to facilitate the visualization of co-variation of multiple pollutants. As illustrated in Figure 3, concentrations of more than more pollutant can be displayed at the same time, with one represented as solid symbol and others outlines (benchmarks).

Temporal Re-expression

Temporal re-expression is a type of cartographic transformation applied to temporal data (Harrower 2007). It is used to display the same data set using different temporal units. For example, daily alcohol-related incident data was animated over an hourly, weekly, or monthly temporal unit to present spatial patterns at different time scales (Liu and Qi 2003). In our visualization interface, temporal re-expression through multi-scale data aggregation was employed on weekly, monthly, and seasonal units. User clicks the corresponding tabs (Figure 3) to view and learn the weekly and seasonal cycles of air quality change. The minimum and maximum values of the chosen air pollutant can be also displayed as visual benchmarks.

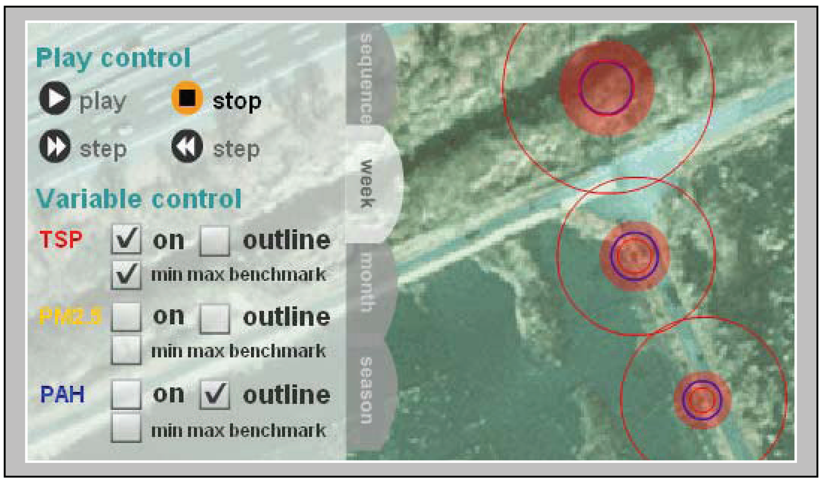

Figure 3. Temporal reexpression tabs and visual benchmarking in the visualization interface.

Knowledge Discovery

Geographic knowledge discovery refers to the extraction of spatial relations or patterns not explicitly stored in spatial databases. Visualization plays a pivotal role in geographic knowledge discovery by utilizing the powerful human visual perception for pattern recognition (Gahegan et al. 2001, Wachowicz 2001). Our study aims to encourage knowledge discovery through the visualization environment, especially knowledge discovery by common audiences (not expert statisticians or data miners).

A usual knowledge discovery procedure consists of four major steps: data preprocessing, pattern extraction, knowledge consolidation, and information presentation. Data preprocessing focuses on the removal of outliers and feature selection (Qi 2004). In our experience with testing the visualization tool with environmental scientists, anomalies in the data were quickly identified when measured values of air quality variables were visualized through controlled animations. Such anomalies include false readings of PAH values during days when the monitoring station malfunctioned, miscalculated traffic counts in some days, among others. Relevant weather features were identified by controlled visualization of air quality change in relation to the recorded weather parameters (wind speed, wind direction, humidity, etc.) During user testing, environmental scientists easily identified wind direction's important roles in determining the existence of a spatial gradient for some air pollutants.

The pattern extraction step is the establishment of relationships between the selected features and air quality variables. In order to obtain quantitative relationship models, statistical models were fitted to the data. In the next step of knowledge discovery, the extracted relationship models were implemented in the visualization environment where users can manipulate weather and traffic variables to see responses from the air quality measures. This was enabled by adding slider bars to the linear symbols that represent the control variables (Figure 4). An important contribution was made by the visualization tool during this step when users found that traffic volume reversely influenced the TSP values. This led to re-examination of the traffic count records and re-establishment of predictive models.

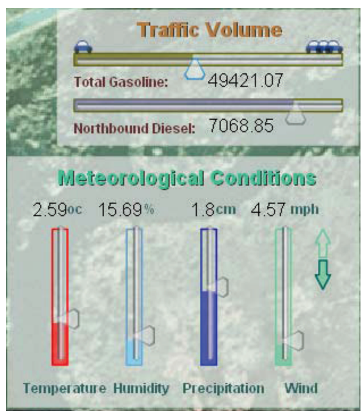

Figure 4. Sliders to visualize predictive model results.

At last, the information presentation step aims at presenting the relationship model, together with both the real data and predicted spatiotemporal patterns to common audiences to promote awareness of air pollution in relation to human factors, and to encourage participation in protection activities. This visualization interface was presented to two student groups-one consists of high school students attending a summer science workshop and the other are college students in an introduction to earth science class. After being introduced to the tool and freely playing with it for 10 minutes, students in the test groups were asked questions that include selected air pollutants, the factors that they thought affected the concentrations of these pollutants, and the spatial variation of air quality in the vicinity of a highway.

Participants agreed on the easiness of this tool and how it is more interesting than graphs and the statistical models that were also presented to them as comparisons. In terms of learning time and correctness of answers to the questions asked, the two student groups did not show significant differences and both correctly answered satisfactory number of questions. One interesting finding was that the younger student group kept being engaged in exploring the data with the visualization tool for a longer time than the older group, who, after answering the questions given to them, often stopped playing with the tool and switched to other things (such as web browsing).

In summary, our visualization tool helped geographic knowledge discovery in the conventional sense. The visualization interface was incorporated into the knowledge discovery procedure with four steps. It played pivotal roles in both the data preprocessing and information presentation steps, and auxiliary roles in the pattern extraction and knowledge consolidation steps, where numerical algorithms were the key components. Furthermore, the visualization tool encouraged knowledge discovery by non-experts, in our study, student groups in a way to help them learn spatial variation of air pollutants in the vicinity of highways and also the influences of traffic and whether conditions on air pollution. It attracted especially common audiences with less knowledge in the subject matter (the high school student group).

Discussion

It has long been recognized that visualization could play important roles in knowledge discovery (Andrienko and Andrienko 1999, MacEachren et al. 2001, Peuquet and Kraak 2002, Lin et al. 2005). Our intention for designing this simple animated visualization interface was to encourage visual learning by common audiences and allow for participatory knowledge discovery. Our user testing with environmental scientists have indicated that the animated and interactive tool aided data preprocessing and knowledge consolidation and is especially useful for result presentation. Preliminary user testing with school students indicated that common audience is able to learn air quality patterns and human-environment relationships using the interface and the learning process is more interesting and engaging than other numerical or graphical methods.

Hallisey (2005) pointed out that any animated map is designed to convey information its maker intended to present. Our study showed, however, that the animated interface also conveyed new information that the designer didn't have any knowledge of and prompted the discovery of new information. The designer of this interface had no previous knowledge on air pollution, its measures, patterns, and causes, etc. The animated map helped identify outliers in the data and relationships among variables.

Animation has long been proven effective to display processes, clarify trends, and very often to provide insight into spatial relations (Dibiase et al., 1992, Harrower 2007). Harrower (2007) also discussed how animation can be used to represent non-temporal data. Our study further showed that animation is especially useful for outlier detection, attributing to the visual power of human perception to detect vital changes.

Statistician John Tukey (1977) used to point out: "We have not looked at out results until we have displayed them effectively." Our visualization tool turned out to be especially useful in the knowledge consolidation step during knowledge discovery – the reverse relationships between traffic and air pollution from the statistical models were easily detected and further efforts were taken to correct the data and models.

Our visualization interface, however, has showed limited effectiveness in terms of examining co-variation between air pollutants. Even with the visual benchmarking methods employed, it was difficult for viewers to perceive such changes. Only when the animation series was played step by step, were such methods useful. As reported by Ahn (2007), animated map is less useful for comparison and association tasks compared to other visualization methods such as temporal coordinate plot and 3D cube. However, thinking of our objective as to use it for public communication, other scientific visualization methods (temporal coordinate plot, 3D cube, etc.) could be less effective as they demand a much long learning curve for common users.

Conclusion

This study aims to address the challenge of representing complex dynamic geographic processes, facilitate geographic knowledge discovery and the communication of scientific discoveries on such processes to the general public. The simple visualization developed in this study proves to be an effective tool for such purpose. Animation is an intuitive way to display complex geographic processes in a more appealing manner with a decreased level of abstractness. This tool has advantages over other visualization methods such as charts, static map, or a simple animated map in that it is interactive. Users are not restricted to the passive acceptance of information presented to them, but can interact with the tool to engage in active thinking and learning. Such interaction makes the presentation more interesting and engaged. Among the EDA techniques experimented, temporal re-expression was found to be highly efficient in revealing temporal patterns of air quality at different temporal scales. Dynamic benchmarking proved to be less effective in showing co-variation of multiple air quality measures. This interface also has advantages over other sophisticated visualization methods in terms of its easiness to use, given that our objective is to present information to common audiences. With these advantages, the proposed visualization tool may serve as a prototype of generic methods to effectively present the results of scientific studies.

References

Ahn, J. S., 2007, Exploratory space-time analysis: Geo-visual analytics approach based on geovisualization and knowledge discovery, Ph.D. thesis, Seoul National University, Korea.

Andrienko G. and Andrienko N., 1999, Knowledge-based visualization to support spatial data mining. In Advances in Intelligent Data Analysis, Proccedings of the 3rd International Symposium, IDA-99 Amsterdam, The Netherlands.

Dibiase, D., MacEachren, A. M., Krygier, J. B., and Reeves, C., 1992, Animation and the role of map design in scientific visualization. Cartography and Geographic Information Systems, 19(4): 201-214.

Dorling, D., 1992, Stretching space and splicing time: from cartographic animation to interactive visualization. Cartography and Geographic Information Systems 19(4):

215-227, 267-270.
EPA, 2009, What we do, United States Environmental Protection Agency, URL: http://www.epa.gov/epahome/whatwedo.htm
Gahegan, M., Wachowicz, M., Harrower, M., and Rhyne, T-M., 2001, The integration of geographic visualization with knowledge discovery in databases and geocomputation. Cartography and Geographic Information Science 28(1): 29-44.
Hallisey, E. J., 2005, Cartographic visualization: An assessment and epistemological review, The Professional Geographer, 57(3): 350-364.
Harrower, M., 2002, Visual Benchmarks: Representing Geographic Change with Map Animation. PhD Dissertation, Pennsylvania State University.
Harrower, M., 2007, Cartographic Animation, International Encyclopedia of Human Geography. R. Kitchin and N. Thrift (eds). Elsevier Press.
Koussoulakou, A., 1994. Spatial-temporal analysis of urban air pollution. In Visualization in Modern Cartography. MacEachren, A. M. and D.R.F. Taylor (ed). London: Pergamon, pp. 243-269.
Lin, J., Keogh, E., and Lonardi, S., 2005, Visualizing and discovering non-trivial patterns in large time series databases. Information Visualization, 4(2): 61-82.
Liu, J. and Qi, F., 2003, Alcohol-related incidents at UW–Madison. Animated Map, available http://www.kean.edu/~fqi/alcohol.html
MacEachren, A. M. and Monmonier, M., 1992., Geographic Visualization: Introduction. Cartography and Geographic Information Systems 19(4): 197-200.
MacEachren, A. M., Polsky, C., Haug, D., Brown, D., Boscoe, F., Beedasy, J., Pickle, L., and Marrara, M., 1997, Visualizing spatial relationships among health, environmental, and demographic statistics: interface design issues. 18th International Cartographic Conference Stockholm, June 23-27, pp. 880-887.
MacEachren, A., Wachowicz, M., Edsall, R., Haug, D., and Masters, R., 2001, Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods. International Journal of Geographical Information Science, 13(4): 311-334.
Monmonier, M, 1990, Strategies for the visualization of geographic time-series data. Cartographica 27(1): 30-45.
Monmonier, M., 1989, Geographic brushing: Enhancing exploratory analysis of the scatterplot matrix. Geographical Analysis 21(1): 81-84.
Peuquet D. and Kraak, M. J., 2002, Geobrowsing: creative thinking and knowledge discovery using geographic visualization. Information Visualization, 1(1): 80-91.
Qi, F., 2004, Knowledge discovery from ‘area-class’ resource maps: data preprocessing for noise reduction. Transactions in Geographical Information Science, 8 (3), 297-308.
Tukey, J. W., 1977, Exploratory data analysis. Reading, Mass.: Addison-Wesley.
Wachowicz, M., 2001, An approach for developing a knowledge construction process based on the integration of Gvis & KDD methods. In Geographic Data Mining & Knowledge Discovery Edited by:Miller HJ, Jiawei H. Taylor & Francis, pp. 239-259.