CHINESE GEOGRAPHICAL SCIENCE
Volume 5, Number 3, pp. 257-264, 1995
Science Press, Beijing, China

A MULTI-WINDOW GIS APPROACH FOR WATER QUALITY DATA ANALYSIS

Lin Hui
(*Department of Geography, The Chinese University of Hong Kong, Shatin, New Territory, Hong Kong*)
Xia Fuxiang
(*NCGIA and Great Lakes Program, State University of New York at Buffalo, Buffalo, NY 14260, USA*)

ABSTRACT: For many water quality studies, a data analyst, or modeler, may need to know the spatio-temproal patterns of the data sets and their relationships in both pre-processing and post-processing. Geographic Information System (GIS) can provide an exploratory spatio-temporal data analysis environment with its capability of spatio-temproal data query, logical and arithmetical operations, visualization, and the integrated statistical functions. This paper will introduce the basic concept and the framework of a GIS exploratory data analysis module (GISEDA) with an emphasis on a multi-window approach for the spatio-temporal pattern analysis of water quality data. A water quality database of GEO-WAMS, a GIS-based watershed modeling support system, has been used to test the prototype of GISEDA.
KEY WORDS: Geographic Information System, multi-window approach, water quality analysis

I. INTRODUCTION

Geographic Information System (GIS) has become a popular topic in watershed analysis after the studies to couple it with the related models[1-6]. A view of GIS-based modeling support system (GBMSS) has also been introduced into watershed analysis. The GBMSS is deemed "a software system that facilitates the job of a water quality modeler in accomplishing the various data-interactive tasks necessary to develop and apply a site-specific, problem-specific, process-oriented mathematical model"[6]. One of the major reasons to integrate GIS with the watershed models is the spatial context of the data from field observation. GIS can provide an efficient spatial data management and a visualization environment. With a GBMSS, modelers will have a clearer image of the quality of the site-specific data and their changing patterns. Therefore, the uncertainty of a modeling process may be reduced when a modeler manipulates data, such as system geometry and initial conditions of modeled constituents.

In addition to the spatial/temporal data management and modeling process, a GBMSS should be able to provide the modeler with a handy data analysis capability . The capability includes viewing the spatial and temporal patterns of the data and examining the statistical characteristics of the data and the relationship between the data sets, before the modeling process. As Hartwig and Dearing said"The more one knows about the data, the more effectively data can be used to develop, test, and refine theory"(1979). Recently, improving GIS data analysis capability with exploratory methods has gained much attention[7-12].

The concept of exploratory data analysis is not new. Tukey defined the exploratory data analysis as "detective work—numerical detective work—or counting detective work—or graphical detective work"[11]. However, the rising GIS has shown its capabilities for spatial/temporal data query, logical and arithmetical operations, visualization environment, and the integrated statistical functions. With those capabilities, the exploratory data analysis can be extended to a more efficient interactive method for the examination of the spatial/temporal patterns, the relationships of the concerned data sets, and the data quality.

Goodchild, Haining and Wise (1992) bring together a summary of three core techniques for exploratory analysis of GIS data: Brunsdon's cartographical and statistical methods; Haining's methods based on the conventional and spatial statistic theories; and Openshaw's methods which given an emphasis on the pattern analysiss. To examine the interrelationship of spatial phenomena and the change pattern of the spatial relationship, a spatiotemporal intersection (STIN)model is proposed[14-16].

The study discussed in this paper focused on the time variable when developing an exploratory data analysis module GISEDA in a GBMSS-Geographically-based Watershed Analysis and Management System (GEO-WAMS).

II. REASONS FOR A MULTI-WINDOW APPROACH

The analysis of water quality data involves multidimensional data processing. The term dimension used here means the spatial or temporal measurement. Usually, water quality data are collected at particular time of day, date location (x and y or longitude and latitude), and depth. For many temporal modeling processes, the date may need to be transferred to DAY to show modeling results, such as the water quality in a harbour 10 days after a oil spill within that water body. At least, data sets in four dimensions should be handled if a space-time reference system is used to organize the water quality data.

To bring to users a better view on those multi-dimension data and data processig results, a multi-window GIS approach can be very helpful for the following analyses:

1) comparing temporal patterns of different attributes such as DO and temperature for the same time period(they have different measure units);

2) interactively checking spatial correlations by liking the statistical data window with the map window;

3) visualizing the 4-D (space-time) data by projecting 4-D data sets into several 3-D data sets in different windows, such as x-y-t window, y-z-t window, or x-z-t window.

With the multi-window approach, users may also compare the pre or post processing data sets.

III. FRAMEWORK OF GISEDA

The GISEDA is a multi-window exploratory data analysis software system in a GIS environment. It consists of seven major components: user interface; spatial/temporal database; data-tool management interface; statistical analysis toolkit; pattern analysis toolkit; graphical analysis toolkit; and multi-window manager. Fig. 1 presents the conceptual framework of GISEDA.

1) User interface—a hierarchical screen menu system for modelers to select the data sets, analytical tools, variables (space, time and parameters), and a multi-window for visually examining the spatial/temporal patterns and other qualities of the data sets.

2) Spatial/temporal database—a data management system which allows the modeler to input, store, analyze, retrieve and display the data according to their spatial locations and temporal identities.

3) Data-tool management interface—a group of programs for interpreting the modeler's commands and the selections on data sets and tools.

4) Statistical analysis toolkit—a group of programs for statistical computation and display, which include scatter plot, histogram, autocorrelation and association, and other tools.

5) Pattern analysis toolkit—a group of programs for generating and dis-

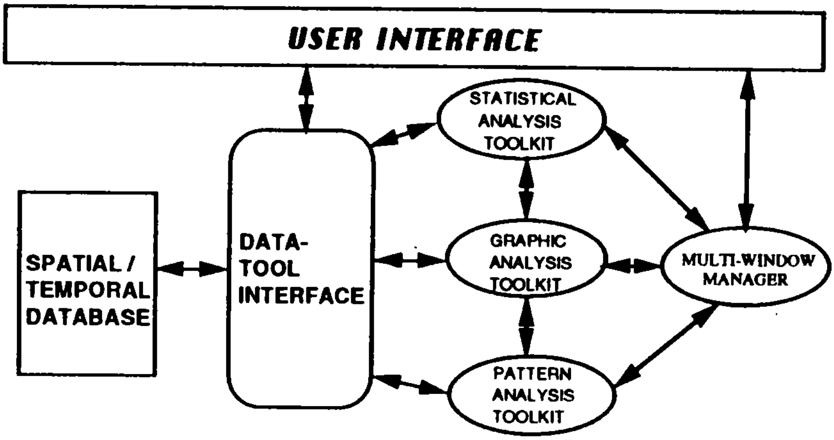

Fig. 1 Major components of GISEDA

playing spatial/temporal profiles and boundary shape, temporal pattern of single point or a group of points, comparing the temporal pattern and spatio-temproal pattern, and examining the spatiotemporal intersection.

6) Graphical analysis toolkit—a group of programs for selecting the graphical data of interest and linking the data sets graphically displayed in different windows, such as highlighting.

7) Multi-window manager—a group of programs for window selection and maintainance.

With the intergration of those components, a normal procedure of GISEDA operation is presented in Fig. 2. The major processes in GISEDA operation are:

1) Select data set—selecting the related data set such as the location data and attribute data to be analyzed, and the background map as the geo-reference.

2) Select tools—selecting tools such as scatter plot, temporal trend, and spatiotemporal intersection (STIN).

3) Select variables—selecting the variables to form the reference system in analysis such as the temperature and the depth of the water.

4) Select window—selecting the window(s) for the display or graphical analysis purpose.

5) Data processing—data manipulation by using the selected tools and data sets.

6) Display—representing the analysis results such as the graphics of scatter plot or the iso-surface after STIN operation.

7) Interactive graphic analysis—linkinng the original data sets in the main

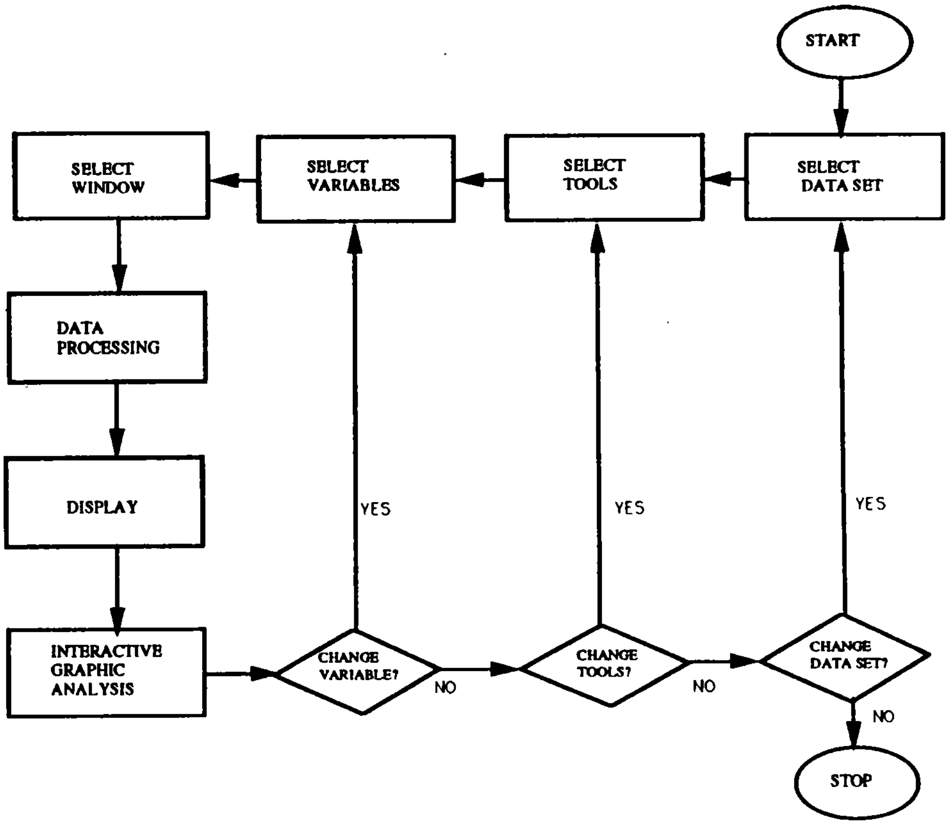

Fig. 2 Flowchart of GISEDA operation

window and the analysis graph in a selected analysis window and identifying the subset(s) of the data concerned. When the analysis results are presented in the analysis window, the related data sets in main window can be high-lighted or classified.

IV. PROTOTYPE STUDY

The GISEDA prototype study is a part of the GEO-WAMS project. This project has been guided by the concept of the coupling of GIS with water quality models. This study focus on the analysis of the water quality data for the Buffalo River of Western New York. The time series data of individual observation sites were brought into a multi-window analysis environment.

As a GBMSS, GEO-WAMS consists of five major components: user interface; spatial temporal database; data-model management interface; process models; and toolkit utilities(Depinto, et al., 1993). Avector GIS ARC/INFO is used to integrate the watershed analysis models, such as WASP4. The computer platform used in this study is Sun Sparcstation 10. This system facilitates the use of multiple windows when performing exploratory data analysis.

The Buffalo River, located in western New York, empties into Lake Erie. The study area is the main stream(lower reach)of the Buffalo River. This area has been affected by heavy industrial development, a high number of combined sewer outfalls, and circulation of non-contact cooling water from lake Erie(Depinto, et al. 1993).

Water quality data used in this study were collected by researchers at Buffalo State College in 1991 as a part of the project supported by U. S. Environmental Protection Agency (EPA). The profiles were measured at six sites at different depth. The profiles include temperature, conductivity, dissolved oxygen (DO), PH, light transmission, and fluorescence. Those profiles were measured between late April and early of November in 1991 on a weekly basis, for four dimensions (longitude, latitude, depth, and time dimensions). Since the spatial reference system used in ARC/INFO is a two dimensional system, the depth data are recorded in the INFO file which includes other parameter data. These data are managed through INFO(relational)database management system as described in Fig. 3. The site ID is used to link these data to the related spatial positions which are stored in other topological data file.

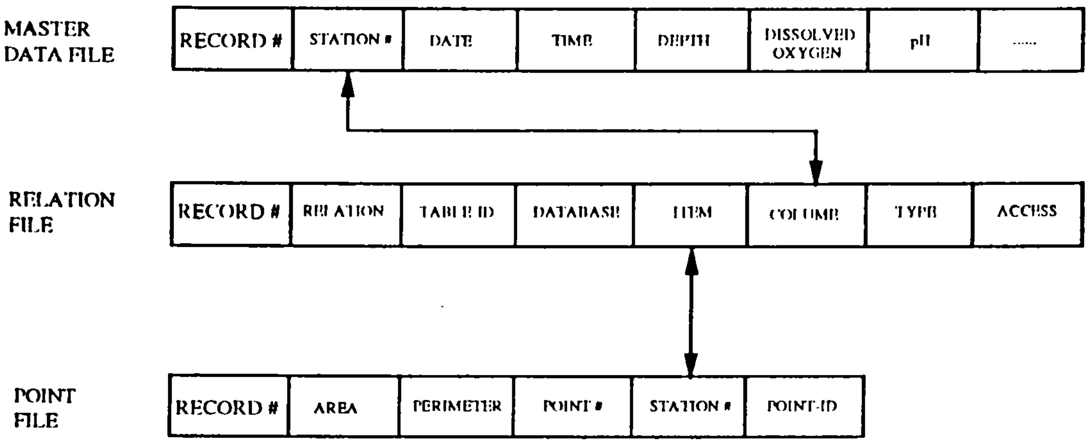

Fig. 3 Link between point coverage and master water quality data file

The conventional exploratory data analysis tools used in this study include scatter plot, histogram, and trend identification. The scatter plot is used for examining the relationship between two variables, such as temperature and dissolved oxygen. The histogram can be used by modelers to visualize the shape of the distribution of the data. The trend identification is to graphically express the data change as a function of space or time, such as the change in pH values over the distance from a river mouth.

Since the data used were collected along the Buffalo River, a profile analysis method can be used to check the spatial characteristics (i.e. depth and distance from river mouth) and temporal characteristics(i.e. the temporal scale and the periods). The time series graphs can also be displayed and compared according to modelers selection of the data item, the time period, the observation site, and the depth.

To explore the spatial and temporal patterns of the 4D data, a multiple window method is used in GISED. Through the user interface, modelers can select the variables (space, time, parameters) and windows for the spatial/temporal pattern analysis and the further interactive graphical analysis.

The multi-window manager contributes to exploratory data analysis through several advantages: 1) comparison for the results of different analysis tools or the results from same tool but with different parameters; 2) visualization of the higher dimensional images by displaying the image into different windows with lower dimension numbers.

The user interface, pattern analysis toolkit, graphic analysis toolkit, multi-window manager and the data-tool management interface are programs majorly written in ARC Macro Language (AML) supported by ARC/INFO GIS system.

V. SUMMARY AND FUTURE STUDIES

The study of the GISEDA prototype discussed in this paper has shown the potential of GIS exploratory data analysis method in water quality data analysis and modeling. A liked multiwindow user interface can provide users a flexible data analysis and visualization environment. Exploring the spatial and temporal patterns of the point data sets can help modelers to understand the data quality and their characteristics. Therefore, modelers may have better idea of how to select the proper parameters and adjust the initial conditions and other related variables before running their models.

Future studies of GISEDA include the development of a 3D modeling and the continued development of the existing toolkits with more flexible operations and better graphical output environment. One of the 3D modeling tasks is to have a 3D segmentation module. This is a key issue for the water quality modeling procedure, such as the 3D finite segment approach used in WASP4 modeling system to solve time-dependent concentrations of state variables.

VI. ACKNOWLEDGEMENTS

This paper is a part of research project "GEO-WAMS: Geographically Based Watershed Analysis and Modeling System", funded by US Environmental Protection Agency, ERL-Duluth. Support for the Great Lakes Program by US EPA is gratefully acknowledgements.

REFERENCES

[1] Sasowsky, K. C. and T. W. Gardner, 1991, Watershed configuration and geographic information system parameterization for SPUR model hydrologic simulations. Water Resources Bulletin. 1:7—18.

[2] Maidment, D. R. 1991. GIS and hydrologic modelling. Proceedings of First International Conference/ Workshop on Integrating Geographic Information Systems and Environmental Modeling. Boulder, Coloredo, USA.

[3] Guan, W. , 1993. Integrating Water Quality Modeling with Geographic Information System—Application to Lake Sidney Lanier, Unpublished Ph. D. Dissertation, University of Georgia, 249pp.

[4] Mertz, T. 1993. GIS targets agricultural nonpoint pollution. GIS World, 4:41—46

[5] Calkins, H. , F. F. Xia, and W. Guan, 1993. GIS based 3-D segmentation for water quality modeling. Proceedings of GIS/LIS'93, Minneapolis, 1:92—101.

[6] Depinto, J. , H. Lin, W. Guan, J. F. Atkinson, P. J. Densham, H. Calkins, and P. W. Rodgers, 1994. Development of GEO-WAMS; an approach to integration of GIS and Great Lakes watershed modeling system. Special GIS issue of Microcomputer and Civil Engineering, (accepted).

[7] Goodchild, M. F,, 1987. A spatial analytical perspective of geographical information system. International Journal of Geographical Information Systems. 4:327—334.

[8] Openshaw, S. A. Cross and M. Charlton. 1990. Building a prototype geographical correlates exploration machine. International Journal of Geographical Information Systems. 3:297—311.

[9] Rowlingson, B. S. , R. Flowerdew and A. Gtrell. 1991. Statistical Spatial Analysis in a Geographical Information Systems Framework. Research Report No. 23, ESRCRRL, Lancaster University, UK.

[10] MacDougall, E. B. 1991, A prototype interface for exploratory analysis of geographic data. Proceedings of the Eleventh annual ESRI User Conference, Palm Springs, California, USA.

[11] Goodchild, M. F. , R. Haining and S. Wise, 1992. Integrating GIS and Spatial Data Analysis: Problems and Possibilities. (Manuscript).

[12] Ding, Y. and A. S. Fotheringham, 1992. The integration of spatial analysis. Computers, the Environment and Urban Systems, 16:3—19

[13] Tukey, J. W. 1977, Exploratory Data Analysis. Addison-Wesley Publishing Company, 688pp.

[14] Lin, H. and H. W. Calkins, 1991. A. rationale for spatiotemporal intersection. Technical Papers of 1991 ACSM-ASPRS Annual Convention. Baltimore. 2:204—213

[15] Lin, H. and D. M. Mark, 1991. Spatiotemporal intersection (STIN)operation and volumetric modeling. Proceedings of GIS/LIS'91, Atlanta, 2:982—991.

[16] Lin, H. 1992. Reasoning and Modeling of Spatiotemporal Intersection, unpublished Ph. D. dissertation, State University of New York at Buffalo, Buffalo, USA, 157pp.