COMPUTER ASSISTED MAPPING AS PART OF A STUDY ON ACID RAIN

Donald L. Batkins, Ph. D.
Assistant Professor of Geography
and
Director of the UMD Cartography Laboratory
University of Minnesota, Duluth
Duluth, Minnesota 55812
and
Gregory Gustafson
Computer Specialist
University of Minnesota, Duluth
Duluth, Minnesota 55812

ABSTRACT

Recently the University of Minnesota, Duluth and the Environmental Protection Agency have begun a study titled "Inventory and Documentation of the Susceptibility of Aquatic Resources to Damage by Airborne Pollutants." The airborne pollutant of most concern is acid precipitation and the object of the inventory and documentation is to develop a modeling methodolgy to discern the susceptibility of aquatic resources in various environments to damage from airborne pollutants. The study site is the Upper Midwest Lake States, with data collection focused on Northeastern Minnesota, Northern Wisconsin and Upper Michigan. The data collected in the inventory are being placed in a large computerized data base. Part of the analysis and documentation calls for the use of computer assisted mapping.

This paper is a progress report on how computer assisted mapping and data base management is being used for this project. There will be a discussion of how several off-the-shelf computer mapping programs (e.g. GIMMS, EPPL, and SURFACE II) are being utilized. The paper documents an attempt to use the World Data Bank II as a GBF for this project. Specialized software developed for the project is also described.

INTRODUCTION

Various units of the College of Letters and Science at the University of Minnesota, Duluth (UMD) are cooperating in a study titled "Inventory and Documentation of the Susceptibility of Aquatic Resources to Damage by Airborne Pollutants." This study is part of a larger "Airborne Pollutant Impacts Research/Assessment Program" managed by the Enironmental Protection Agency's Environmental Research Laboratory-Duluth (ERL-D)*. The UMD study involves many different

* Although the research described in this paper has been funded wholly or in part by the United States Environmental Protection Agency through contract or grant CR809332-01-00 to The University of Minnesota, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the view of the Agency and no official endorsement should be inferred.

components, but a common theme is the collection and analysis of data coverng a broad geographical area. Although the collection of data is not a topic for this paper, the computerized data handling, analysis, and display of it, within it's geographical context, is our concern.

The UMD study encompasses several levels of geography. The highest (or macro) level is the land area around the three upper Great Lakes (Huron, Michigan, and Superior). This includes sections of the States of Michigan, Wisconsin, and Minnesota, and part of the Canadian Province of Ontario. The next (or micro) level is the select regions within these four political units. Each of these micro levels in turn is made up of a number of sample sites with each site correspondings to a watershed for a sample lake (see Figure 1).

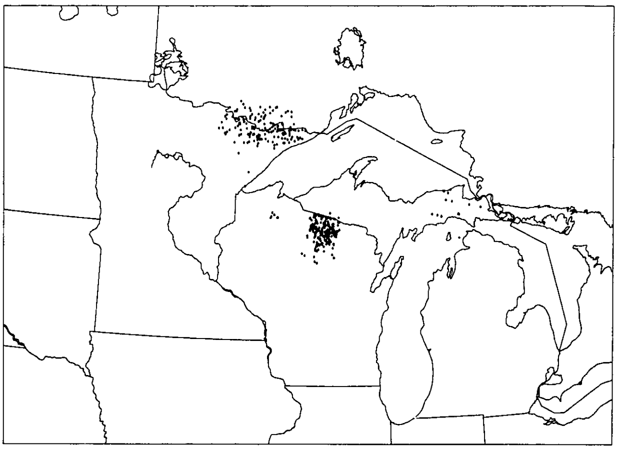

Figure 1: Study Area With Current Sample Sites

Of special concern for this paper are the micro level sample sites within the State of Minnesota. These sites are mostly located in the Superor National Forest. These sites are special because they involve the use of the Minnesota State Geographic Information System.

The project follows a traditional Geochemistry research methodolgy which uses computers for supplemental analysis and display capablities. This research philosophy does not require that a major computerized geographical information system be developed or utilized to coordinate data handling, analysis, and display. The traditional approach simply incorporates into the project those available computer programs which can perform as many of the data handling, analysis, and display tasks as possible. While making no value judgement as to whether this research methodolgy is best, the paper will show the workability of the methodolgy selected for this study, insofar as it concerns computer applications.

The computer programs currently used and those under consideration for future use are reviewed first. The project is currently using, or planning to use, the following existing computer programs or computerized data bases.

1. MLMIS: The Minnesota Land Management Information System, a grid cell based data analysis and display system.

2. World Data Bank II: A geographical data base for North America, which includes definitions of rivers, lakes, and regional boundaries.

3. Surface II: A generalized program for handling spatially distributed data, produces isoline maps and block diagrams.

4. GIMMS: The Geographical Information Mapping and Manipulation System, a high quality polygon based mapping system.

5. Other Software: Examples would include SIR, The Scientific Information Retrieval System, and SPSS, Statistical Package for the Social Sciences.

As part of this review, the use of these programs within the context of the study will be explained. Information on specialized software developed for this project will also be presented.

THE DATA FOR THE STUDY

The data includes a number of variables from a wide range of sources. This data all focuses on the relationship between airborne pollutants and aquatic resources. For example it includes:

1. The chemical make-up of the lake water from the study sites;

2. The chemical make-up of precipitation, both snow and rain;

3. The analysis of the biological aspects of study sites (zooplankton, phytoplankton, fish, etc.);

4. Miscellaneous geographical data, which includes vegetation, surface area of watershed, geology, soils, etc.

The common thread connecting all of this data together is its geographical location. Because of this both the analysis and the display of the data is conducive to computerized geographical mapping and manipulation programs.

THE SOFTWARE AND HARDWARE CONFIGURATION

DATA BASE AND STATISTICAL PROGRAMS

Because of the research methodolgy used, a large number of computer programs are being used. Since the study is ongoing, not all of the decisions about program usage have been made. However, what follows is a review of what is currently being done and what is likely to happen in the future.

THE MLMIS SYSTEM

The MLMIS system is a sophisticated, grid oriented, geographic information system which was developed by units of the University of Minnesota and Minnesota's planning agency (Hsu, 1976). The system dates from the late 1960's but it has undergone substantial improvement over the years. MLMIS allows the user considerable control over the data for manipulation purposes. Some examples of possible data manipulation include:

1. Bigtab: a routine which allows the user to create a new map with a data level for each combination of two variables per grid cell.

2. Flow: will create a new map that is based on logical paths through multiple variables.

3. Score: permits assignment of values relative to the characteristics of each variable and the importance of each variable for the Bigtab process.

4. Edge: analysis where two variable are adjacent to each other.

MLMIS outputs the results of its manipulation in either map or tabular form and uses the Environmental Planning and Programming Language (EPPL). This study utilizes two versions of EPPL. EPPL5, the older of the two, runs on UMD'S computer and will be used for most of the work. A newer version, EPPL6, runs on the state's Department of Energy, Planning & Developments Prime computer system, will be used when its added capacities are necessary. The data file structure is similar for both versions thus facilitating this operating procedure.

In addition to having a very powerful language in EPPL, MLMIS also maintains a number of geographical data bases for the state of Minnesota. This study plans to use these data bases to obtain the geographical data for sample sites within the State of Minnesota. Moreover, much of the data collected for the State is planned to be put into MLMIS formatted data bases which will allow for analysis and display by EPPL (see Figure 2). Some of the data maintained by MLMIS and of interest to this study include:

1. Geology: from geology maps of the state updated with more current field data;

2. Vegetation: such as forest and grassland cover;

3. Soils: from the soil conservation maps of the state;

4. Watersheds: defined for areas of 5 1/2 square miles.

EPPL maps can be output on a wide range of devices, including a Dicomed image recorder, a color CRT, a Benson dot plotter, a Tri-Color printer, and a standard line printer. At UMD only the line printer option is available since the other devices are not currently on site. This precipitated some experimentation with other types of output devices. For instance, experiments have been successfully conducted using an Apple computer-controlled IDS 560G printer and a Zeta pen plotter to print or plot maps produced by EPPL. An experiment is planned using a recently purchased Decwriter 3, with graphics option. Should the Decwriter 3 experiment prove successful, and there is no reason to doubt that it will not, it will probably be used to print EPPL maps for reports etc.

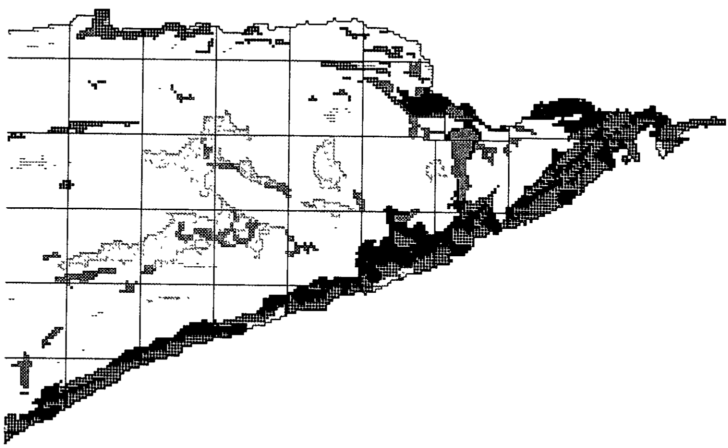

Figure 2: An EPPL Generated Map of Surficial (1-5 feet) Texture for Part of the Minnesota Sample area

The Zeta option, because of the size capabilities and the fact it is possible to create maps in color, can be used to produce poster sized EPPL maps for display. This option is worth going over in more detail because it illusrates how easy it is to interface EPPL to miscellaneous output devices. EPPL allows the user the option of saving a raster formated data file of the map. It is quite easy to take that data and convert it into a formal map. In the case of the Zeta map, all that needs to be done is write a program which plotted a color shaded square at each row and column location. The Z value at that location determines the color, based on a simple classification scheme designated by the user. Since row and column output assumes a regular rectangle, the plotted squares retain the correct geographical shape.

This map data file option also makes it easy to use the output data from EPPL as input to other programs such as SURFACE II.

SURFACE II

Surface II is basically a very sophiscated isoline mapping computer program (Sampson, 1975). The program inputs standard X-Y-Z data, it then grids and displays that data. The user has considerable control over the gridding and display functions, and Surface II can output on either a line printer or a pen plotter (see Figure 3).

It is planned to use the Surface II program to display select data in either isoline maps or block diagrams. Much of the data collected lends itself to isoline mapping. For example, the chemical make-up of precipitation, both snow and rain, could be isoline mapped. This type of mapping would clearly show patterns of concentration for select airborne pollutants.

In addition to mapping raw data, some experiments have been conducted showing the possibility of using Surface II to display output from MLMIS. While the experiments were successful it was not clear that this method of reprocessing MLMIS output by Surface II will be used for the project.

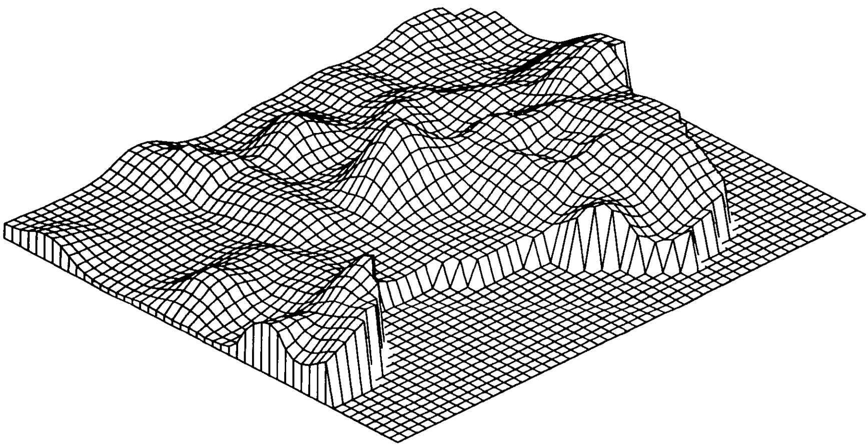

Figure 3: EPPL Data Displayed in a Block Diagram Created by SURFACE II

GIMMS

GIMMS is a general purpose, user oriented, integrated, geographic processing system (Waugh and Taylor, 1976). This means that GIMMS can be used to map any geographical area so long as it is defined (i.e. as a GBF) in a way compatible to the program. GIMMS accepts a wide variety of GBF's including those in the old SYMAP format. The preferred GBF format is labeled line segment strings with which the program can do a considerable amount of polygon checking. GIMMS is extremely user friendly since most of the commands use a very flexible input system called General Parameter Input System (GPIS), which contains over 100 commands and at least another 200 sub-commands (Waugh, 1977). The mapping quality of GIMMS is exceedingly high and the system produces both choropleth and point symbol maps.

GIMMS will be used to display some of the point data (i.e. field collected sample data). The advantage of GIMMS is that it is capable of producing multi-variate symbols, which of course allow for the display of related data from a select sample site. It is planned to convert WDB II data to GIMMS format so it could be used as background for the point symbol maps.

WORLD DATA BANK II

The World Data Bank II (WDB II) is a very large GBF for the whole world which was developed by the United States Central Intelligence Agency. The GBF consists of five reels of magentic tape, one reel of which defines North America. The WDB II format for North America consists of four files, one each for:

1. coastlines, islands, and lakes;

2. rivers;

3. international boundaries;

4. internal boundaries.

Each of the four files contains a series of line segments defining the geographaical data. Each line segment in turn has a rank. For example, the coast, islands and lakes file has 12 ranks:

1. Coast, islands, and lakes that appear on all maps;

2. Additional major islands and lakes

3. Intermediate islands and lakes;

4. Minor islands and lakes;

5. Intermittent major lakes;

6. Intermittent minor lakes;

7. Reefs;

8. Salt pans-major

9. Salt pans-minor;

10. Ice shelves-major;

11. Ice shelves-minor;

12. Glaciers.

For the most part line rank is analogous to map detail. Thus by selecting certain line ranks, it would be possible to create maps of varying accuracy.

Since the WDB II contained data for the whole North American continent it was necessary to strip out those line segments defining the study region. At the same time the files organization was mofified, though not its basic line segment structure.

The data for the Great Lakes study area retains its basic line segment structure but it has been reformated into a standard FORTAN direct access file. The new file has a record length of 100 words and contains some 80,000 plus /minus records. This new file allows random access and this principle was incorpoated into some special software under development for the study.

<u>SOME SPECIALLY DEVELOPED SOFTWARE</u>
Because of the different data formats required by the various software used in the study, several small data conversion programs had to be developed. These programs are relatively simple and require no special reference here. However one new piece of software is worth mention. As it is currently configured, this new software manipulates and displays data from WDB II. At this writting, the program is not yet completed but is far enough along for inclusion here.

The World Data Bank Manager computer program is interactive and utilizes a special file, containing pointers, to access the WDB II direct access data file. Basically, the program querys the user on the area to be mapped, level of detail to be shown, scale necessary, and other related information (see Figure 4).

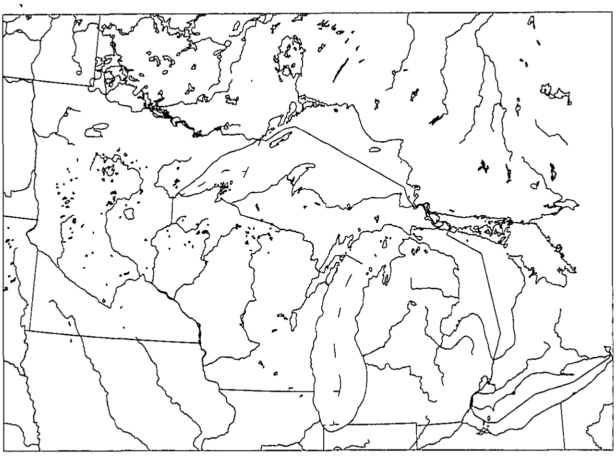

Figure 4: A Sample of Output from the Reformated WDB II

Using the input constraints, the pointer file is searched for the information needed to read the data file and map it. This direct access and pointer file logic has some advantages. For example, windowing is speeded up since the pointer file contains information on the location of each line segment string. Thus only those line segment strings which might fall inside the window are read and processed. This of course reduces cost since fewer computer resources are used. In addition the direct access uses computer resources more efficiently.

OTHER RELATED SOFTWARE

While the mapping software is the major focus of this paper, it is necessary to touch on two other areas of interest. The first is the data based management software and the problems revolving around it. A second area worth mention is the statistical packages in use or under consideration.

Currently the field data is being stored and handled in the SIR data base format (Robinson, 1980). However, while that was and still is the basic plan, some problems have developed with this decision. The first is the computer skill level of some of the project's staff. Unforunately, some of the staff has an insufficient understanding of computer data handling processess. Also, the version of SIR used to make the project cost estimates was a preliminary one and was subsquently withdrawn. Do to the inefficienties of the withdrawn version of SIR II, the cost estimates were very high and this raised the question of whether to change to a more cost effective data based management system. Cost, therefore, has become an issue in the debate as to whether to continue the use of SIR.

Several possiblities have been discussed. One solution would be to upgrade staff personnel so they would be truely conversant with SIR. This might take some time but it would mean that the data formatting would remain constant. Another possiblitiy would be to purchase or lease another data based management system. Finally, a proposal has surfaced to develop a unique data based management system for the study. At this time, no decision has been reached.

Recently some questions have arisen concerning the statistical packages. The three possible choices are SPSS, BMDP, and MINITAB and these all have been used for various tasks. The reason for using all three is that not all of the packages offer a complete selection of statistical procedures. Thus it is not possible to standardize on a single package. However, the question that remains to be answered, is if a select statistic (e.g. regression) generated for a set sample of data is the same regardless of the statistical package used. it. This question has not yet been resolved. Until it is, the only solution is to select a package for a certain task and always use it. Thus the statistics would always be comparable for very data set for all time periods.

CONCLUSION

While some of the problems outlined in the last section of the paper would not have occured if a comprehensive geographic information and analysis system had been used, other problems would probably have surfaced. Both the data base management and the statistical package problems seem to have resulted more from planning deficiences than any inherent unworkability. By definition, the use of a large number of independent programs, with no existing interfaces, does require much more planning than if one single system is used. Clearly problems still exist with the data base management and statistical programs but the problems are not insurmountable.

The mapping segment of the project has generated less problems overall. Part of the reason for this is that fewer staff were involved and their knowledge of the mapping programs was quite high. Thus it was easier to interface programs or to pick the right system for a given task.

While the project is not far enough advanced to say that no serious mapping problems will surface, it can be safely stated that none have come up so far. The methodolgy of using a number of mapping programs for a study such as this clearly causes some logistical problems but it appears a workable solution for this project. However, since no testing was or will be carried out using a large integrated geographic information and analysis program, this report cannot definatively say which research methodogly is best.

One thing shown by this paper is that a wide range of mapping programs can be used in a single comprehensive analysis. Thus while it might be desirable to use a large geographic information system, it is not always necessary.

REFERENCES

Batkins, Donald L. 1982. "Micros and Mainframes: Some Ideas on Their Combined use for Computer Assisted Mapping," paper presented at Harvard Computer Graphics Week.

Dixon, W. J. and Brown M.B. (eds) 1979. BMDP-79 Biomedical Computer Program, P-Series. Berkely: University of California Press.

Hie, Norman H. et al 1979. SPSS: Statistical Package for the Social Sciences. New York: McGraw-Hill.

Hinkley, Elizabeth ed. 1978. MINITAB II, Reference Manual. Minneapolis: University of Minnesota, Univeristy Computer Center.

Hsu, M. L., Kozer, K., Orning, G. W., and Streed, P. G. 1975. "Computer Aplications in Land-use Mapping and the Minnesota Land Management Information System," in John C. Davis and Michael J. McCullagh eds. Display and Analysis of Spatial Data.

Robinson, Barry N. et al 1980. SIR: Scientific Information Retrieval User's Manual, Version 2. Evanston: SIR, Inc.

Sampson, Robert J. 1975. Surface II Graphics System. Lawerence: Kansas Geological Survey.

Waugh, Thomas C. 1977. "A Parameter Driven Lanuage for use in Application Systems." Proceedings of Advanced Symposium on Topogical Data Structures, Harvard University.

Waugh, Thomas C. and Taylor, D. R. F. 1976. "GIMMS: An Example of an Operational System for Computer Cartography." Canadian Cartographer 13: 156-66.