

Chapter 23

Data Manipulation of GIS for Modelling and Simulation in Resource Management

Jiguo Zhen

Chinese Visiting Scholar

William James and Fangju Wang

School of Engineering, University of Guelph
Guelph, Ontario, Canada N1G 2W1

GIS and modelling packages have become important tools for resource management and environmental planning at the watershed level. Integration of the two technologies evidently greatly improves data analysis and manipulation, and model turnaround times. High performance here depends heavily upon creating and organizing GIS data sets in a proper structure for transferring them to executing models in an efficient way. In this chapter, an effective method for handling gradient data, one of the major parameters for water resources modelling, is addressed. Experimental results are presented and analyzed.

Zhen, J., W. James and F. Wang. 1993. "Data Manipulation of GIS for Modelling and Simulation in Resource Management." *Journal of Water Management Modeling* R175-23. doi: 10.14796/JWMM.R175-23.

© CHI 1993 www.chijournal.org ISSN: 2292-6062 (Formerly in New Techniques for Modelling the Management of Stormwater Quality Impacts. ISBN: 0-87371-898-4)

23.1 Introduction: Models, GIS and Management

23.1.1 Modern Resources Management

Resources refers to any classes of objects, both those existing naturally and those constructed artificially, that are considered essential for human life and development. Natural resource components include soil, water, air, forests, animals, fisheries, minerals and so on; artificial resources can be regarded as all facilities, utilities, and engineered construction for human activity.

Ever since humans began exploring the resources around them, they began to learn how to manage them wisely, i.e. for their long-term or sustainable benefit. Resource management, including resource conservation, preservation, monitoring and exploration, has been and continues to be, a fundamentally important human activity. As the human population and modern technological development both explode uncontrollably, the complexity of resource management also increases, especially control of natural resources depletion and ecosystem deterioration. Accordingly, resource management has become more and more challenging, and, itself, technology-dependent.

Effective resource management is based on both the ever-accumulating knowledge base and on high-technology computer information systems. At the intersection of these systems, as described throughout this book, two powerful tools for resource management are being integrated: modelling packages and GIS's. Modelling packages are used to depict the structure of existing systems, to determine relationships among processes and systems, and to reveal the dynamic dependence of every kind of known process of the real world in a scientific way. They are especially useful for testing proposed resource management strategies, and to find the best among many alternatives. GIS's store, manipulate, analyze and display spatial data of real objects. They are especially efficient at managing large arrays of information that can be located in the environment.

23.1.2 Importance of Modelling to Resource Management

Modern resource management addresses its problems systematically and dynamically in a context of the whole environmental system. Complexities in both the human environment and in resource management today share three similar aspects: 1. more and more factors have to be taken into account, 2. larger and larger geographical areas must be studied, and 3. the study must encompass longer and longer timespans. Resource management problems have spatially-distributed parameters; all processes are multi-dimensional; most of the parameters are time-varying and non-linear; and more often than not, problems within large geographical regions are characterized by uncertainties. Large distributed-parameter systems behave quite differently from similar problems in small fields, the so-called point processes, and this introduces concomitant modelling difficulty (Vansteenkiste, 1978).

Basically, there are two categories of processes or factors considered in modelling: the natural and the artificial. Their spatial and temporal diversity are notably different, as are their effects on modelling.

Natural processes are characterized by high variability and unpredictability. For example, the runoff of a watershed is fundamentally dependent on: rainfall density, intensity and its spatial kinematics; evaporation and related processes of temperature, wind speed, and land cover type; soil and subsurface conductivity, storage and flow patterns; drainage pattern and related processes of topography; and so on. These natural factors vary spatially and temporally during a storm, a season, and over the decades. This aspect often demands spatial statistical analysis and probability inference for every important process, especially the rain falling on the watershed.

On the other hand, infrastructure and engineered facilities are also fundamentally important factors. They differ from place to place, as do the natural conditions, the human demands and their

dimensions, operational parameters, and their dynamic responses to the impacts of other factors. E.g., properties like curb length, pipe diameter, channel depth, linkage, shape, geometric and hydraulic parameters, although stable, have unique operational parameters. Changes of land use, population density and other demographic parameters significantly complicate the behaviour and modelling of these artificial resources.

In recent years, advanced modelling packages dealing with process-oriented issues have been developed by interdisciplinary teams. For example, WEPP, the USDA Water Erosion Prediction Project models, which are based on fundamentals of stochastic weather generation, infiltration theory, hydrology, soil physics, plant science, hydraulic and erosion mechanics (Nicks et al. 1989), account for many factors. It follows that many factors with very different types of variables are required in most advanced model packages. For example, in WEPP, for determining the conditions of snowmelt and frozen soil, no less than 60 variables and parameters are addressed (Young, 1989).

Such complex resource management requires huge amounts of modelling effort. In most cases, the number of factors modelled is based on the amplitude of the processes, the sensitivities of the factors, and the spatial resolution required by the study. The larger the study area, the more factors that are involved; the greater the accuracy of spatial and time resolution and the increased sensitivity demanded, the greater the modelling effort.

Numerical weather prediction, very important in storm water management, is a good example. Evidently, if five factors are to be taken into account (horizontal grid points, vertical levels, number of equations, operations per equations, time steps) for a 48-hour forecast in an area measuring 5 degrees longitude by 5 degrees latitude, then, even with very rough spatial resolution (horizontal 250 km, vertical 2 km), the number of operations can reach as many as 2.8 billion (Koermer, 1988).

To cope with this, more sophisticated and professional modelling packages have been developed in the past two decades

(see for example James, 1992). These packages have improved contemporary resource management in three ways:

1. Understanding the relationship and interactions among factors in resource management can be achieved better than ever before, since sophisticated models account not only for spatial diversity of the facts, but account also for their dynamic interactions in different phases or processes. For example, SWMM is a comprehensive water quality and quantity simulation model developed primarily for urban areas. Single-event and continuous simulations can be performed for almost all components of the rainfall, runoff and quality cycles of a mixed rural, urban and industrial water and sewershed (Huber et al., 1991).
2. Modelling is especially useful for accurate prediction. For example, in stormwater resources management, the accuracy of prediction of storm water flows is critical to human security and investment. Disasters are caused by floods every year (eg. China in 1991). In many cases, the costs of these disasters have been largely reduced by accurate prediction. More precise and reliable predictions could reduce the misery even further.
3. Findings and results of modelling can be used to determine fundamental parameters for systematic engineering designs for every kind of infrastructure in watersheds. For example, modelling the Universal Soil Loss Equation has brought about many suitable tillage designs to control soil erosion in many countries.

23.1.3 Roles of GIS in Supporting Modelling and Simulation

GIS's are computer-based systems that are used to store and

manipulate geographically-referenced data or information. A GIS software package consists of a set of programming tools and a database management system (DBMS). It provides utilities for data input, storage, retrieval and output. GIS's make the production and analysis of resources information more efficient for planning and provide a natural resources potentiality assessment (Stanley, 1989).

GIS's support modelling that requires abundant sources of information, including spatial entities and their relationships, described as a neighbouring and networking representation of the real world. Basically, GIS's are very effective for handling data sets of spatially-distributed entities and their attributes, such as land cover types, soil and subsurface hydraulic conductivities, and so on.

GIS's can be used to create special data sets which usually cannot be obtained by conventional means. Quite often, most natural factors with their model variables are spatially dependent on each other. Unfortunately, a large proportion of the input data required in modelling, such as the rainfall density and local rate of rain, evaporation, temperature, and wind speed with their spatial distribution, are hardly measurable sufficiently accurately, or are even unobtainable in practice. This is why conditions have to be assumed in modelling, even though they are sometimes by no means close to *real world* estimates. For example, climatic default values are most often used when observed or real-time data is not available (Koermer, 1988).

Obviously, results would be more credible and perhaps more accurate if field observations are used. Coverage can be aided by using the computerized spatial data interpolation techniques of GIS's, which are based on limited control points. These GIS interpolation techniques are adapted from established manual methods used in many academic fields, especially in geography and meteorology.

GIS computerized map interpolation can increase the efficiency of thematic map layer generation. Because of their

efficacy, digital thematic map layers become the main source of the GIS database, and have very positively impacted spatial analysis for resource management. By efficient data manipulation, GIS's manipulate large amounts of spatial data, some of which may be critical in modelling. It is paramount that GIS selection be based on the availability of utilities for thematic map layer generation; many inexpensive GIS's promoted for water resources engineering lack these capabilities.

Another inherent utility (of fully-fledged GIS's) for dynamic and interactive modelling is high-performance data-acquisition, specifically: map layer conversion, digital data communication and incorporation of remote-sensing data. As shown elsewhere in this book, these utilities save database management resources and time.

There seems to be no doubt that the GIS features described here can significantly improve accuracy and enable more feasibility modelling in the limited time period usually available for design or planning studies.

Natural factors, and many artificial factors, always change temporally. For example, rate of rain and precipitation density, temperature, and wind speed and direction; they all change stochastically throughout a storm. In order to predict as accurately as possible their effects on relevant processes such as pollutographs, an appropriate, usually shorter, time-step is chosen for input data files. The choice of timestep is based on the intrinsic speed of the process being modelled.

Modelling is practical and reliable only when the required data, such as precipitation, temperature, and wind speed, are made available. Where modelling also requires other spatial data, such as hydraulic conveyance, and hydrological parameters of land use, that change during the simulation, only GIS data manipulation techniques can efficiently provide the large variety of input data-sets in a reasonably short time.

GIS's enhance both the accuracy and the operational speed of modelling, because they make it easier to rearrange the spatial

scope, the model integration time-step, the simulation parameters for every subwatershed and sometimes even the extension of the study to cover peripheral watersheds. With functions for processing remote-sensing datasets, manipulation of imaginary and spatially consistent data, many more sensitivity analyses, validation tests, design optimizations and error analyses can be performed.

Many GIS's support modelling for decision-making on environmental issues. The work of the European Community and the United Nations Environment Program provides an example of a GIS application in resources management: the CORINE program. Its objective is to provide a comprehensive, integrated spatial database of environment data relevant to European policy-making. The ARC/INFO GIS from the Environment Research Institute (ESRI) was chosen as a development platform for the CORINE system.

There has been a rapid increase in the number of GIS-based predictive modelling packages for: surface water management modelling (this book); evaluating and protection-planning of groundwater resource pollution (Barry, 1990, Baker, 1990); land resource management at the urban, regional, state, and national levels of administration (Nielsen,1990); and so on.

23.1.4 Efficiency Issues in GIS Supported Modelling and Simulation

Efficiency of resource management modelling depends largely on GIS data supply. Usually, GIS data manipulation for this purpose includes two objectives: sufficient and useful data capture, and high-speed generation and tailoring of data sets. A GIS should be evaluated for three criteria: 1. the desirability, 2. the sufficiency of output data quality, and 3. the operating speed.

As indicated above, increasingly sophisticated modelling must employ more and more variables (and the number is already

large), smaller time steps, and complete more iterations and more design-runs in a limited time period. These demand high-speed GIS data manipulation. Besides hardware factors, high GIS operating data manipulation speed is related to: the numbers, the precision requirements, and the manipulation of the attributes of the variables in the input data file. As well as the GIS operating speed, modelling speed relies on the manipulation skills evident in the user's GIS command file.

The model variables have various relationships and different fits to the spatial entities present in the GIS's databases. The complexity of the relationship and the closeness of fit of model variables to the GIS entities may be divided into four classes:

1. physical dimensions such as area and length;
2. variables directly-related and plainly-manifested, such as soil type and soil pH values;
3. directly-related but requiring arithmetical computations, such as slope and slope aspect; and
4. stochastically-dependent such as rainfall intensity and distribution patterns.

The higher the class of variables, the more important is the consideration of GIS data manipulation efficiency. In most cases, the recently-developed GIS's support the first-three types of variables well, but of course with different efficiencies. For the fourth class, GIS's usually include spatial-data generators to suit their demands. These generators could reside in either a GIS or modelling package (Flanagan, 1991).

Once sufficient data has been generated, high-speed dataset tailoring is indicated. For a heavy modelling workload for complex watershed systems, data manipulation skills in both GIS command tools and other software is necessary.

In a conventional approach, spatial or spatially-related datasets of variables such as area, length, elevation, slope, aspect and other aspects of entities are manipulated manually and manually input into the data file. The large number of variables leads to great data volumes and is error-prone. Evidently, manual approaches are inefficient for sophisticated modelling.

GIS's support modelling not only by means of their DBMS, but also through their other, handy tools. These may facilitate modelling efficiency at four levels:

1. First, spatial and attribute data can be stored and interrelated in a GIS database for convenient retrieval. They can be tailored easily to suit specific needs (Smith, 1987). For example, lengths of streams, channels, areas of land, and cover types can be related, coded and retrieved as model variables.
2. Second, useful data from other types of databases can be implemented into a GIS database either systematically or on a necessity basis. This includes important data conversion between GIS and non-GIS systems (eg. from AUTOCAD to ARC/INFO) and among different GIS databases (for example, from ARC/ INFO to GRASS).
3. The third level is represented by the capability to create new data sets containing information needed for modelling. This involves generating new thematic layers or combinations of different thematic layers, such as climatic, physiographic, land use or soils, and their component layers. The former involves problem-addressed map interpolation while the latter is map overlaying. Isotope maps are usually created by automatic interpolation and then stored digitally as a thematic map layer in a GIS database. With functions for triangulation, delineation (Jenson and Domingue, 1988) and interpolation, GIS's can save space and time.

4. The fourth level is integration of the above techniques with other software or data manipulation tools. This mainly involves macros programming of a GIS or in other software workspace. It features well-developed interface software written to overcome obstacles or practical problems that GIS's alone could handle with difficulty. There is a conspicuous recent trend to improve GIS's interface for practical analyses (Smith, 1987).

23.2 Problems Related to Efficiency

Although GIS data manipulation has greatly improved resource management modelling (Nielsen et al, 1990; Barry; 1990), there are still gaps between GIS's and specific modelling objectives. In the following, two such gaps are addressed.

23.2.1 Different Data Formats Between GIS's and Modelling Packages

Usually, a GIS specifies a standardized data format for its database, called a tabulated or plat file. On the other hand, the formats of input files for various models are all quite different. A few lines from a standard SWMM input data file provides an example:

file: TRANS14.DAT

SW 1 8 9

...

A1 ' STEVEN'S AVE., NOVEMBER 28 '

	NDT	NINPUT	NNYN	NNPE	NOUTS	NPRINT	NPOLL	NITER	IDATEZ	METRIC	INTPRT
B1	60	0	2	4	4	0	4	4	731128	0	10

...

* SEWER ELEMENT DATA

	NOE	NUE(1)	NUE(2)	NUE(3)	NNTYPE	DIST	GEOM1	SLOPE	ROUGH	GEOM2	BARREL	GEOM3
E1	1	151	157	0	21	0	3.0	0	0	0		

...

Note that sewer variables such as DIST, SLOPE, ROUGH can be the attribute items of the same objects in a GIS database. The DIST values for every sewer will be available from a digitized sewer-map layer formed in a GIS; the ROUGH values can be added to each record with relevant data from another source, such as a facilities management system, or sewer inventory system. It is also a normal and simple data manipulation to extract them from a GIS output data file. However, if there are many subwatersheds, sewersheds or catchments in a watershed, data manipulation will be tedious.

23.2.2 Gradient Calculation

Gradient Calculation, Modelling and Simulation

An additional problem is calculation of gradients. Slopes and gradients are commonly used in topography, hydrology and related engineering fields. Gradients are also widespread in many academic disciplines such as meteorology, biosphere modelling, mass/energy transfer, and even in regional economic development strategy. Basically, a gradient refers to the partial differentiation of any arithmetic values of an attribute over the distance between spatial points. It represents the spatial rate of change of an attribute between objects of the same kind in spatial analysis.

Since resource management modelling mostly deals with process analyses in spatial systems, gradient variables are used to reflect potential change rate, to depict dynamic characteristics of components of a system, precisely as the SWMM variable SLOPE above. For calculating gradients of simple or a non-spatial system, it is unnecessary to assume GIS technology. However, for modelling, especially for hydraulics research in large geophysical areas, which usually need large amounts of gradient values and parameters, gradient calculation has to be carried out by resorting to GIS data manipulation.

Conventional Methods of Calculating Gradient Values in GIS and their Limitations

To calculate slope values in GIS's, relevant elevation data of any entity for an area has to be in a digitized format. There are two methods of preparing slope map layers in GIS's. One is to employ a GIS with a three-dimensional elevation data model and its triangulation and interpolation tools. Another is to digitize a slope map whenever available. The first method seems simple, but the powerful functions and tools required prove demanding for some GIS's, and usually limited three-dimensional elevation data are available. The second is also difficult because slope maps are not readily available.

It seems to be possible to calculate slopes of entities when a digital slope map layer is available, if the GIS employed has at least a two-dimension-coordinate system and even if it does not contain triangulation functions. Without a slope map layer, slope calculation of any kind of an entity is halted.

Furthermore, it is difficult to calculate slope values between points solely, or along line-features with direction (along a profile) options, even though the slope, or an isotope map layer, is already digitized. It is unlikely that slopes of polygon-featured entities in a watershed will be calculated; overlaying these two map layers, both polygon-featured, is an overlay operation with two different entities: the overlaying map layer is polygon-featured or a slope or isotope map layer, and the overlaid map layer is a point- or line-featured entity.

This above method for calculating slope values is not always favourable and applicable for modelling. Gradients used as parameters for modelling mostly refer to an attribute of the entities with line-feature or point-features. When this kind of map layer is overlaid, the lines joining two nodes will be intersected by the boundary lines of the polygons of the isotope map layer. Therefore one line is split into many segments with dependent attribute values in the GIS database. The number of

segments created from one line is equal to the number of boundary lines crossing it. Accordingly, the output file of the outcome map layer will expand.

These data explosions raise many problems for both modelling and the GIS itself. It is difficult to tailor very large files and compile useful values from them into the input data file for modelling, not only because it is time consuming, but also because of the possible errors inherent in such tailoring. Also, enormous data redundancy occupies too much space in a GIS database. Therefore this method of slope calculation by no means enhances the efficiency of modelling. Modelling needs useful data sets, not data redundancy.

For example, we conducted an experiment; a map layer resulting from overlaying the following two layers: a pipe network map of 25 pipes, and a slope isotope map layer. The resulting output file contains the slope values of each line on this map layer and has 71 records about different lines, split by the overlaying. The operation time to compute such an overlay is 22% more than that of our special procedure developed for the same purpose. Besides this, much more time will be needed for reorganizing these records and tailoring this file for practical modelling.

23.3 Integration of Data Manipulation Techniques for Modelling and Simulation

As recorded in this book, there is much strong activity in the integration of spatial data and modelling capabilities (Grayman et al, 1982). Note, however, that only with powerful GIS data manipulation can the above problems be solved. The proposed new method is based on integration of data manipulation techniques as follows:

23.3.1 Exploration of GIS Manipulation Capabilities

Sophisticated GIS's feature flexibility by providing commands, routines, and modules at different levels, and independent macro programming functions. These greatly enhance data manipulation for problems such as slope calculation. For our example, we used ARC/INFO with only a two-dimensional coordinate system. The basic procedure could be as follows:

1. Prepare the elevation, line-feature or point-feature map layers in digital format either by digitizing a contour map in the ADS module or by data conversion from other data format (Most GIS have data conversion facilities).
2. Build a database of the elevation map layer by using conventional commands assigning elevation values for every cell or polygon of the elevation map layer.
3. Overlay the elevation map layer onto the line-feature or other prepared layers.
4. Extract all the cell or polygon ID records and their relevant average elevations to form a relationship file.
5. Append this file to the line-feature or point-feature map layer by twice issuing the relate command.
6. Calculate the slope values of a line or between points by the calculation abilities of the GIS.
 - 7a. Extract and compile these values into the modelling and simulation input file, or
 - 7b. Combine it with a subwatershed or catchment boundary map layer to generate all slope-related input files. (The macro

programming language SML of ARC/INFO can facilitate most of the above procedure provided the objective area and entities of the line-feature are precisely allocated).

23.3.2 Combination of Different GIS and GIS-Related Software

Problems such as no suitable functions within one GIS to deal with slope or gradient calculations could be overcome by combining with another GIS package which does contain the necessary functions. For example, GRASS (Geographic Resources Analysis Supporting System) has routines to deal with SLOPE and ASPECT calculations, but works in raster data structure, which does not match the ARC/INFO format. However GRASS also has handy functions for converting raster data to vector data and vice versa, and to convert data files from and to ARC/INFO, AUTOCAD and other GIS-related software packages. So, when using GRASS to calculate slope or other gradient values of polygon-features, such as soil type or vegetation map layers, first convert them into an appropriate input data file that will greatly benefit the modelling.

23.3.3 Harnessing the Operation System Commands and Shell Utilities in Input Data File Tailoring

Most GIS's take advantage of the DOS commands or shell utilities (eg. C shell) to improve data manipulation. There are many skills regarding this aspect that can be developed and verified to support GIS data manipulation for modelling. For example, the pipe command (>) of operating systems for PC machines can be combined with a GIS command for listing map layer names, or for processing them in some other convenient way. Moreover, it is especially useful to use batch processing

techniques when many data files have to be processed or tailored.

23.4 Conclusions

1. It is necessary to integrate techniques of GIS for modelling and simulation for resource management since it will greatly enhance efficiency.
2. Through data manipulation in a GIS, data sets of area of subwatersheds, and length of linear features in a watershed or regions can be obtained automatically rather than manually for dynamic modelling and simulation of resources.
3. Data sets of slope or gradient for terrain analysis can easily be computed if the GIS employed has TIN functions and three-dimensional elevation data. It is especially suitable for area-feature map-layer analysis even when the data manipulation is tedious. However, for line-features, it is better to use a module as described herein, since there are no such limitations on the TIN functions and the module is more flexible, time-saving and easier to handle.
4. The key step in this method is to join items or relate items; the items to be related must be consistent with those in the point-coverage or linear-feature coverage. Otherwise, the results may be totally wrong. To make sure of this, it is necessary to understand thoroughly the topology of a GIS coverage and the output data structure.
5. Further interface programs should be developed in order to improve the integration of GIS and modelling/simulation files.

23.5 References

Barry M. Evans and Wayne L. M. (1990). *A GIS-based approach to evaluating regional groundwater pollution potential with DRSTIC*. J. of Soil and Water Conservation, 45(2):242-245.

Carol Pringle Baker and Ernest C. Panciera.(1990). *A Geographic Information System for Groundwater Protection Planning*. J. soil and Water Conservation, 45(2):246-248

Flanagan, D.C., Ferris, J.E., and Lane L.J. (1991). *WEPP Hillslope Profile Erosion Model*. NSERL Report No. 7. National Soil Erosion Research Laboratory, USDA-Agriculture Research Service.

Huber W., EPA Storm Water Management Model, SWMM v4.05. (1991). Department of Environmental Engineering Sciences, University of Florida.

James, W. 1992. "Introduction to the SWMM Environment." *Journal of Water Management Modeling* R175-01. doi: 10.14796/JWMM.R175-01.

Jenson, S. K. and Domingue, J. O. (1988). *Extracting topographic structure from digital elevation data for geographic system analysis*. Photogrammetric Engineering and Remote Sensing, 54(11):1593-1600.

Koermer J. P. (1988). *Atmospheric modelling---a meteorologist's perspective in Modelling of the Atmosphere*. SPIE Volume, 928:2-7. Editor: Laurence S. Rothman. Orlando, Florida.

- Nielson G. A., Caprio J. M., Mcdaniel P.A. (1990).
MAPS: A GIS for land resource Management in Montana.
Soil and Water Conservation, 45(4): 450-453.
- Smith, T., D. Peuquet, S. Menon, and P. Agarwal. (1987).
KBGIS-II: A Knowledge-Based Geographic Information System. International Journal of Geographical Information Systems. 1(2):149-172
- Terence R. Smith.(1987). *Requirement and Principles for the Implementation of Large-scale Geographic Information Systems.* Int. J. Geographical Information Systems, 1(1): 13-31.
- Vansteenkiste.(1978). *Introduction: modelling, identification and control in environmental systems.* Proceedings of the IFIP Working Conference on Modelling and Simulation of Land, Air and Water resources Systems:vii--ix.

