DEVELOPMENT OF A NETWORK BASED HYDROLOGICAL ANALYSIS SYSTEM

Assela Pathirana, Srikantha Herath and Katumi Musiake

University of Tokyo

ABSTRACT: This paper discusses a distributed hydrological simulation system, for streamlining the process of hydrological analysis. Some of the reasons preventing the adaptation of the distributed hydrological models by a wider circle of practicing hydrologists are the complexity in terms of operation and data requirement, the data management problems created by the sheer immensity of the information requirement and the huge computing power needs. The system described, a solution implemented to make hydrological analysis accessible to a wider audience, consists of a hydrological database designed based on the relational database theory, integrated to a GIS and a number of distributed hydrological simulation models and graphical interfaces accessible on WWW. The database stores hydrological time series data and two-dimensional spatial data of various resolutions. A detailed meta-data system is implemented to aid the use and the management of data. A system implemented using proprietary technologies is opened to the end user equipped with general-purpose software like web browsers, which use standardized protocols. Thus, a broader user base can benefit from advanced hydrological analysis tools without the complication of a host of unfamiliar and expensive technologies and operating environments.

INTRODUCTION

Hydrological simulation of watersheds has been evolving rapidly from simple conceptual representations to physically based, distributed models that can describe in detail, the conditions of a watershed in various phases of hydrological cycle (Herath, 1994, Jha, 1997, Yang, 1998). This increasing complexity has added many data management problems into the modelling domain. Information required by the models as well as the results of the models has become complex. Under these conditions the traditional methods for managing simulations has no longer been adequate and new methods for better organization of simulation process were explored. Integrated approach or organizing the simulation models and other required resources like GIS systems into a single work-platform is a result of this need (Dangermond and Schutzberg, 1998).

A case for integrated systems

Most of the time spent by the hydrologist in a simulation exercise can no doubt, be attributed to the setting up of models, and formulation of results rather than on the creative part of the exercise that mostly need human involvement. This drawback is a result of the lack of a unified methodology for treating many types of hydrological analysis in a generic manner, which can evolve as newer hydrological models are introduced. Simulation models are becoming fast. The actual time taken for simulation has become a small fraction of what it has been a decade ago. Hence, there are a number of new research possibilities, which were not feasible in past due to restrictions imposed by the available limited computing power. However, these heavy analyses almost always demand better data management concepts too, due to the huge information involvement, both at the setting up and interpretation ends. Present day distributed hydrological models are complicated in operation. A large amount of time has to be spent in order to understand the inner workings of a model, before starting to use it. This fact has severely restricted these models being popular among practicing hydrologists and other related parties outside the immediate research circles.

The objective of the system integration is to make the simulations fast to learn, easy to use and productive. An ideal integrated hydrological analysis system must demand only the understanding of modeling principals, but not the mastering of the inner workings of the model from the new user. Since distributed hydrological models generally are heavy applications, resource sharing by network systems is advantageous.

The integrated system approach is not new. Especially after the popularity of distributed hydrological models and Geographic Information Systems (GIS) there have been many attempts to interface these elements to produces integrated systems (Jakowski, 1995, Watkins, et al., 1996). The available literature ranges from pre and post processor design for simulation models, to full integration of DBMS and GIS with simulation models as integrated decision support systems (Yoon and Padmanabaran, 1994). Recent view of hydrological modelling suggests that a complete hydrological model is one that is fully integrated with GIS (Johnson, 1989).

The hydrological information management system described here was designed to examine the feasibility of extending the capabilities of such integrated system, to a network based multi-user environment. In order to make the system available for a wider circle of users, it was intended to be interfaced on the World Wide Web (WWW)

First the conceptual framework envisioned for designing the system is discussed. Then the details of the design of the system are presented. Next, with an introduction of the specific technologies used in the project, the system implementation is presented.

THE CONCEPTUAL FRAMEWORK

Figure 1 shows the typical cycle of procedures a hydrologist follows through, in order to use simulation models in her problem solving process. As shown, there are number of areas where total automation is possible while there are other stages, which can be partially automated in order to utilize best possible properties from the two parties involved in the problem solving process, namely, man and the computer. Thus it must be possible to hand the most repetitive and mechanical work to the computer while the human counterpart can concentrate on the actual problem at hand.

The system was planned to have the following components: a) A hydrological database of temporal and spatial data: this includes data structures to store data as well as a suit of data management; query and retrieval tools oriented for the hydrological data manipulations; b) A linked GIS system for extended spatial data handling capabilities; c) Incorporation of distributed hydrological models; d) An intuitive and widely available user interface.
The required design and the operation of the system are illustrated in the Figure 2.

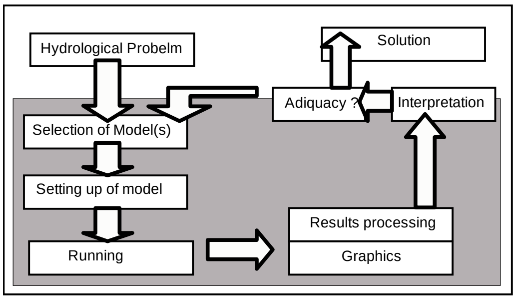


Figure 1 : the process of hydrological analysis

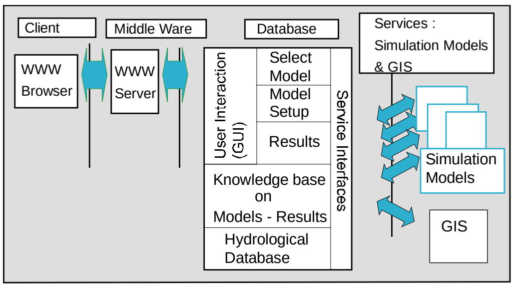


Figure 2: The macro scale design of IISHA

A unique feature of this design framework is that, a database server acts as the central component of the system. All the transactions with the users are managed by this system. Scheduling of tasks and utilization of the available services in order to fulfill those requests are also managed by the database. Having such a 'single entry point' to the system, simplifies many problems related to the multi-user environment, the *stateless* nature of the WWW transactions and the security issues of the Internet.

A logical hierarchy

Considering the diverse nature of the hydrological data structures, perhaps the best repository for hydrological data would be of an object based design. However, a system based on the relational theory (see the next section) was selected due to the fact that many mature database systems based on relational theory were available.

Relations defined in a database system are very much appropriate to organize information in manageable and efficient way. But, at the same time these relations become too rigid to represent the diverse relations that are useful to effectively group hydrological data components, in a unique way that is intuitive and rational for the hydrologist. For example: various geophysical datasets concerning a particular watershed should be linked in a hierarchy under the watershed. Similarly, soil properties describing to a soil distribution dataset should be linked to that soil data set.

A simple way to overcome this problem within the relational framework was proposed and is explained in the system design.

THE RELATIONAL MODEL

In 1970, Edgar F. Codd published a paper (Codd, 1970) that is considered to be the most influential single document on database technology. The concept Codd proposed is known as the `relational model' and in contrast to the previous models; it is based on mathematical principals, which are associated to the set theory and Boolean logic.

Among the major strengths of relational theory are a) it avoids repeating of information b) information does not become inconsistent. Consider a small company database where the following details of the employees have to be stored: name, salary, department belonging to, location of the department. In a relational database that information in organized as shown in figure 3. Though four employs belong to the 'Sales' department, the details of the sales department are recorded only once. The EMPNO in table employee and the DEPTNO in table departments are known as primary keys. Primary key is the column or set of columns, which can identify a row of a database uniquely. The DEPTNO column in table employee is known as a foreign key of that table referring to table DEPARTMENTS. By pointing to a primary key in another table, foreign keys provide relations among tables.

Department

DEPTNO	DNAME	LOC
10	Accounting	New York
20	Research	Dallas
30	Sales	Chicago
40	Operations	Boston

Employee

EMPNO	ENAME	SAL	DEPTNO
7369	Smith	300	20
7499	Allen	1600	30
7521	Ward	1250	30
7566	Jones	2975	20
7654	Martin	1250	30
7698	Blake	2850	30

Figure 3: Example Relations

The databases implementing the relational model are known as relational databases. The Structured Query Language was developed to manipulate both data and data structures in relational databases. For a technical discussion on relational theory and databases the reader is referred to Rolland (1992).

SOFTWARE

During the last two decades the database and GIS programs has attained a certain level of maturity. There were many good commercial as well as free systems in this category for use. However, this was not true in the same degree, about the technologies required to disseminate a interactive analysis system on the World Wide Web and to link the various functional systems together. In the entire process of development of the present system one of the most laborious parts of the entire project was trying-out various software solutions and selecting an appropriate set.

The database

A site license for Oracle 7.3 Universal Database server was already available and thus, this was used as the database. Oracle at this version was implementing the full relational model and had 'entry-level' compliance with the ISO/IEC 9075:1992 "Database Language SQL" standard. There is oracle's own 'procedural extension' to SQL, known as PL/SQL, which is intended for in-database programming. The pre-compilers were available to mix database access with popular programming languages like FORTRAN.

Oracle had supplied the Oracle Web Server - a special web server that provides a programming platform to enable CGI (common gateway interface) programming with PL/SQL.

The GIS

The already available Arc/Info GIS version 7 was used as the GIS. While providing spatial data manipulations both in raster (GRID) and vector format, it has a macro programming language called AML (Arc Macro Language) for automation and extension of the system.

SYSTEM DESIGN

A modular design was adopted in order to provide flexibility for future extensions to the system. It was envisioned that the system to have six distinct components, namely, a (traditional) Hydrological Database (HDBASE), a hydrological simulation interface (HIS) – to provide data infrastructure for running hydrological simulations, a spatial operations subsystem (SPSS) - to manage spatial data manipulations, External processing environment (EOS) – the collective set of external links between various components like GIS, database and models, a Central task scheduling system (CSS), a User interaction sub system (UISS).

Due to the limitation of space only the components HDBASE and HSI are explained in detail. Refer to Herath (1998) and Pathirana (1998) for further details on Design and Implementation of various components of the system.

The Hydrological Database

The design of the hydrological database consists of twofold tasks. The first is the design of data structures to store the actual data. The second and most challenging is a meta-data structure. Meta-data may be defined and 'descriptive data on actual data'. Due the complicated and heterogeneous nature of hydrological data, the design of a good meta-data structure is of paramount importance to the proper use of a database.

The meta-data involved in a hydrological database can be classified into three broad categories:

- <u>Data Related</u>: Including
  Type information like parameter (rainfall, elevation, etc), units,
  Basic statistics
  Details on gathering like instrument used, field conditions
  Temporal/Spatial resolutions and controls
- <u>Location Related</u>: Including
  Geographical reference (geo-coordinates)
  Hydrological reference (Watershed information)
  Political reference: (Country, province, etc.)

- Management Related:
  Storage (location, media, access)
  Maintenance (Persons in-charge at host site and at the data source)
  security levels

The metadata structure was organized in 22 relational database tables. This final model was based on a series of discussions with a panel of hydrological researchers about the appropriate nature of the metadata structure. The main table (named DATA_INFO) contains an entry (a row) for each set of data. This table has parameters which are unique to each set of data or those which cannot be effectively and foreign keys to the other tables. DATA_ID (an integer) is the primary key.

The actual data records were related to the DATA_ID in the DATA_INFO table. Figure 4 shows this relationship for time-series data. All the time series data is stored in a single database table having three columns namely, DATA_ID, TIME and VALUE. The composite primary key is (DATA_ID, TIME). The DATA_ID in this table is related to the DATA_ID in DATA_INFO table. The performance of the time series data table, which contains millions of records, is maintained by using indexing and partitioning methods (Singh et al., 1997).

Hierarchical organization of data

The implementation of hierarchical relationships on the existing relational model was achieved simply by introduction of an extra table names CHILD_LINK in relation with DATA_INFO (main) table. Consider the following example: The DATA_INFO table contains information on a catchments X, rain gauges R1, R2 and R3 where R1 and R2 are related to X because they are inside the catchment X. Figure 5 illustrates (with simplified database tables) how this relationship is realized in the database. The DATA_ID of X (D1) appears twice in CHILD_LINK's PARENT_DATA_ID, with D2and D3in CHILD_DATA_ID. D2 and D3 are the DATA_IDs of R1 and R2 respectively.

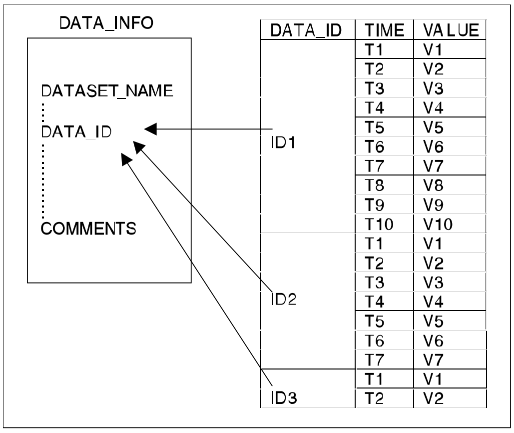


Figure 4: Linking data to meta-data

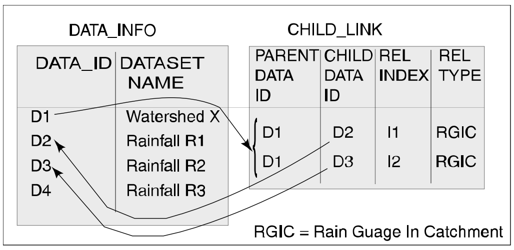


Figure 5: Linking data objects in a hierarchy

Presently the system contains temporal and spatial data for a number of countries in Asia including Japan, Sri Lanka, Thailand and Nepal.

The hydrological Simulation Interface (HSI)

The HSI is the component, which links simulation models to the system. When a user sets-up a hydrological model to run, the database-structure of HSI records the information needed to run the model. These information includes, pointers to the spatial and temporal datasets to be used, starting and ending times of simulation, at what points or what intervals to log output, etc. Later, depending on the control of the CSS, the actual processing of the simulation job is started. During the running of the model, data is streamed both from HDBASE (the original hydrological data) and the local storage of HSI (the data specific to the particular run of the model. e.g. duration of simulation) to the model as required and the model output is stored in the HSI. The interaction with the GIS is also carried out as needed. As the model provides output, the data-structures of the UISS are also updated, to enable the user to check the progress of the simulation process.

THE SYSTEM IMPLEMENTATION

The implementation of the database resident part of the system was trivial once a proper design is arrived at. At this stage the most challenging components was the linking of the various system components together (EOE). Two aspects of these components are discussed.

Modification of the simulation models

As the case with the traditional computing in civil engineering, all the models available for use were coded in FORTRAN. These models were written keeping the ad-hot flat-file based data system in mind. Thus, the input output to the models is through data files. One simplest way to use this models was to have system interface to write the input files needed, then call the model to run on that data and read-in the output of the model. However, in order to have a tighter control over the simulation process and to avoid the delay in reporting the progress of the model to the user (some of these models take execution times of order of hours or days to complete), slightly different approach was adopted in coupling.

The model code was changed (using pre-compilers mentioned under software base) to do all the input-output tasks with the database directly. Then the model code was compiled into a shared library (a type of code which another program can load an run when needed). A small server program was produced to 'listen' to database requests and to load and execute the appropriate models.

The internal structure of the EOS was also implemented as a flexible network based system. Many computers can participate in the service providing services to run single or multiple instances of one or many models. Thus, the EOS together with the central scheduling system provides IISHA with a simulation service, which utilizes the available system resources to the best possible extent to cater for the user requests.

THE SYSTEM OPERATION

The user interface of IISHA is accessed through a common web browser. Thus, anybody with a web browser and an Internet connection is a potential user of the system. The interface pages are generated on-the-fly, based on the nature of the analysis being performed and the nature of the models used.

Data Search

With the help of the meta-data structure, extensive search capabilities are provided with the system. User can build her own search criteria using Boolean logic and query the database. The search results have direct pointers to the data and the results are interactively displayed for examining, use in models or downloading (see figures 6 and 7).

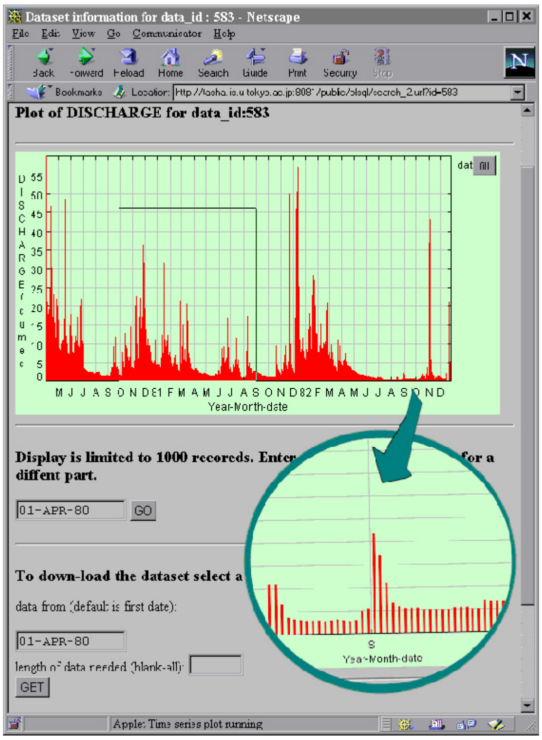
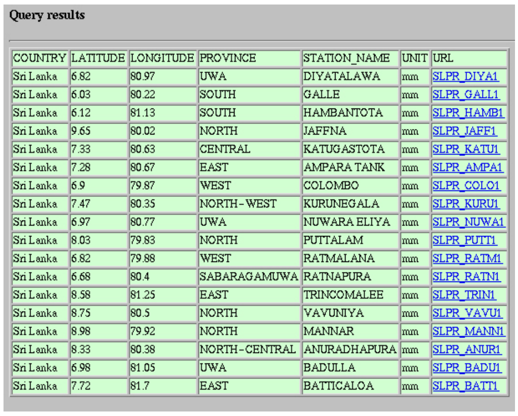


Figure 6: A time series data set.

Simulations

Users are guided through all aspects of setting up of hydrological models like selecting the model, selecting the data, providing the necessary model parameters, running the model and visualizing of data, with interactive and intuitive interface pages. Java applets are used to add interactive abilities to the results visualization. Figure 8 shows an example of a simple model output – the water balance results. The model used was the grid based hydrological model described in Herath (1994).

Whenever a dataset is displayed, the hierarchical relationships under that dataset are also presented. In hydrological analysis, these related datasets generally contain most of the information required. For



Query results

COUNTRY	LATITUDE	LONGITUDE	PROVINCE	STATION_NAME	UNIT	URL
Sri Lanka	6.82	80.97	UWA	DIYATALAWA	mm	SLPR_DIYA1
Sri Lanka	6.03	80.22	SOUTH	GALLE	mm	SLPR_GALL1
Sri Lanka	6.12	81.13	SOUTH	HAMBANTOTA	mm	SLPR_HAMB1
Sri Lanka	9.65	80.02	NORTH	JAFFNA	mm	SLPR_JAFF1
Sri Lanka	7.33	80.63	CENTRAL	KATUGASTOTA	mm	SLPR_KATU1
Sri Lanka	7.28	80.67	EAST	AMPARA TANK	mm	SLPR_AMPA1
Sri Lanka	6.9	79.87	WEST	COLOMBO	mm	SLPR_COLO1
Sri Lanka	7.47	80.35	NORTH-WEST	KURUNEGALA	mm	SLPR_KURU1
Sri Lanka	6.97	80.77	UWA	NUWARA ELIYA	mm	SLPR_NUWA1
Sri Lanka	8.03	79.83	NORTH	PUTTALAM	mm	SLPR_PUTT1
Sri Lanka	6.82	79.88	WEST	RATMALANA	mm	SLPR_RATM1
Sri Lanka	6.68	80.4	SABARAGAMUWA	RATNAPURA	mm	SLPR_RATN1
Sri Lanka	8.58	81.25	EAST	TRINCOMALEE	mm	SLPR_TRIN1
Sri Lanka	8.75	80.5	NORTH	VAVUNIYA	mm	SLPR_VAVU1
Sri Lanka	8.98	79.92	NORTH	MANNAR	mm	SLPR_MANN1
Sri Lanka	8.33	80.38	NORTH-CENTRAL	ANURADHAPURA	mm	SLPR_ANUR1
Sri Lanka	6.98	81.05	UWA	BADULLA	mm	SLPR_BADU1
Sri Lanka	7.72	81.7	EAST	BATTICALOA	mm	SLPR_BATT1

Figure 7: A query results page. Search constraints were `PARAMETER = RAINFALL AND COUNTRY = SRI LANKA`

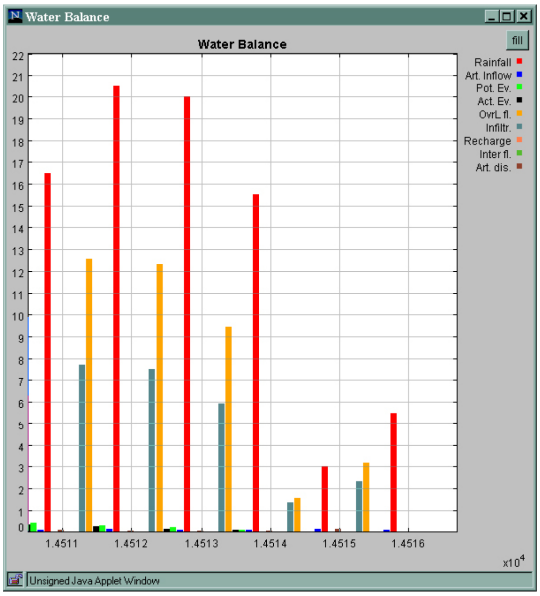


Figure 8: Part of a model output.

example for a long term water resources simulation of a particular watershed, various spatial data layers like, river networks, soil distribution, water usage data and temporal datasets like rainfall records, sewerage flow patterns are needed. Due to the design of the system, these data are automatically listed with the base dataset representing the catchment. Thus, the time spent for isolating these data is greatly reduced. As the system is used for analysis these hierarchical relationships are continuously updated to reflect the needs of many types of hydrological analysis.

DISCUSSION AND SYNTHESIS

A conceptual framework for a Database central, WWW interfaced hydrological analysis system has been proposed. By the implementation of the system, the workability of the proposal was verified. These kind of systems while on one hand makes it easy for the hydrologist who is new to a particular hydrological model to rapidly become acquainted and to master the use of the model, on the other hand makes the advanced hydrological analysis tools available for a much wider user base.

In the implementation of the system, many commercially available software systems, which may be considerably expensive, have been used. However, there are many freely available SQL databases, GIS systems and WWW interfacing techniques, which may be able to do these tasks. The authors have examined at least some of these alternatives and do not perceive many problems of using these for the implementation of similar systems.

As mentioned before, the attempts of hydrological system integration in the past did not address the network implementation aspects adequately, and thus they had a limited impact on the practicing hydrologist. However, during the last few years, after the implementation of the system described there have been several projects geared towards this goal in many places of the world (e.g. Simonovic 1998). During these few years the technologies involved in integration of databases and GIS with other programs has become considerably matured. As a result of the steady growth of the Internet, better methodologies have been introduced for interfacing of such interactive systems on the WWW.

Though the technologies involved in the system have evolved a long way, the authors believe that, the theoretical framework has not been outdated. Based on the same principals, today it is possible to implement much versatile and useful decision support systems with even lesser effort

REFERENCES:

Codd, E.F. (1970) "A relational model of data for large shared databanks", Communications of the Association for Computing Machinery, 13(6):377-387, http://www.acm.org/classics/nov95/

Dangermond, J. and Schutzberg, A. (1998), "Engineering geographic information systems and databases: A new frontier", Journal of Computing in Civil Engineering, ASCE

Herath, S. and Musiake, K. (1994), "Analysis and modeling of urban hydrological cycle", Proc: International Conference on Urban Engineering in Asian Cities in the 21st century.

Herath, S. and Pathirana, A. (1998), Database Centered Simulation Using the WWW, Proc: International Symposium on Information Technology Tools for Natural Disaster Risk Management, International Center for Disaster-Mitigation Engineering, Tokyo

Jakowski, P. (1995), "Integrating geographic information systems and multiple criteria decision-making systems", International journal of GIS.

Jha, R. (1997), "Development of Distributed Hydrological Model and Incorporating Surface Flow Interactions and Its Application to Large Tropical Catchments", PhD Thesis, University of Tokyo

Johnson, L.E. (1989) "Maphyd – a digital map based hydrologic modeling system", Photog. Engrg. and Remote Sensing.

Norman, Ronald J. (1988) "Object –Oriented Systems Analysis and Design", Prentice Hall, Upper Saddle River, New Jersey

Pathirana, A. (1998), "Database Centered Hydrological Simulation on the WWW", Masters thesis, University of Tokyo.

Rolland, F.D. (1992), "Relational database management with ORACLE", Addison-Wesley

Singh, L., Leigh, K. and Zafian, J., (1997) Oracle 7.3 Developer's Guide, Sams Publishing

Simonovic, S.P., (1998) "Decision Support System for Flood Mangement in The Red River Basin", Report prepared for the International Joint Commission, Red River Task Force, Winnipeg, http://www.ijc.org/boards/rrb/red_dss.pdf

Watkins, D.W., Mckinney, D.C., Maidment, D.R. and Lin, M.D. (1996), "Use of geographic information systems in ground water flow modelling", Journal of water resource planning and management, ASCE.

Yang, D. (1998), "Distributed hydrologic Model Using Hillslope Discretization Base on Catchment Area Function: Development and Applications", PhD thesis, University of Tokyo, 1998

Yoon, Y and Padmanabharan, G. (1994), "A decision support system for non-point source pollution management using a distributed model-GIS-DBMS linkage", CiCE, 94. Proc. The 1st congress held in conjunction with A/E/C systems, ASCE.