iEMSs 2008: International Congress on Environmental Modelling and Software
Integrating Sciences and Information Technology for Environmental Assessment and Decision Making
4th Biennial Meeting of iEMSs, http://www.iemss.org/iemss2008/index.php?n=Main.Proceedings
M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli and G. Guariso (Eds.)
International Environmental Modelling and Software Society (iEMSs), 2008

Bridging the Gap between Geohydrologic Data and Distributed Hydrologic Modeling

G. Bhatt [a], **M. Kumar** [a] **and** **C.J. Duffy** [a]

[a] *Civil and Environmental Engineering, Pennsylvania State University, University Park, Pennsylvania, USA (gxb913/muk139/cxd11@psu.edu)*

Abstract: This paper outlines and demonstrates a strategy for coupling of integrated hydrologic model and Geographic Information System (GIS) to meet pre/post processing of data and visualization. Physically based fully distributed integrated hydrologic models seek to simulate hydrologic state variables and their interactions in space and time. The process requires interaction with a range of heterogeneous data layers such as topography, soils, hydrogeology, climate, and land use. Clearly, this requires a strategy for defining topology definitions, data gathering and development. Traditionally GIS has been used for data management, analysis and visualization. Integrated use and streamlineed development of sophisticated numerical models and commercial Geographic Information Systems (GISs) poses challenges inherited from proprietary data structures, rigidity in their data-models, non-dynamic data interaction with pluggable software components and platform dependence. Independent hydrologic modeling systems (HMSs), GISs and Decision Support Systems (DSSs) not only increase model setup and analysis time but they also result in data isolation, data integrity problems and broken data flows between models and the tools used to analyze their inputs and results. In this paper we present an open-source, extensible and pluggable architecture, platform independent "tightly-coupled" GIS interface to Penn State Integrated Hydrologic Model (PIHM) called PIHMgis. The tight-coupling between the GIS and the model is achieved by the development of PIHMgis shared-data model to promote minimum data redundancy and optimal retrievability [Kumar et al., 2008]. The procedural framework of PIHMgis is demonstrated through its application to Shaver's Creek Watershed located in Susquehanna River Basin in Pennsylvania.

Keywords: Geographic Information Systems (GIS); Hydrologic Model; Shavers Creek; Susquehanna River Basin.

1. INTRODUCTION

Physically based distributed hydrologic models simulate the spatio-temporal dynamics of the important hydrologic processes using spatially distributed watershed's physical properties and forcing fields [Feeze et al., 1969]. These models better represent natural heterogeneities [Entekhabi et al., 1989; Pitman et al., 1990] with the goal of enhancing our understanding and prediction of the spatio-temporal dynamics of hydrologic processes. Clearly, a key challenge in the development and use of distributed, physically based modeling frameworks is the large number of physical parameters that must be incorporated into the model. Geographic Information Systems (GISs) with their ability to handle both spatial and non-spatial data, and to perform data management and analysis operations have a strong potential to advance development and use of more complex modeling frameworks if used appropriately. A major deficiency of GIS that has been recognized is the lack of sophisticated analytical and modeling capabilities [Maidment, 1993; Wilson, 1996; Camara, 1999]. Likewise many, existing hydrologic models are not developed with data structures that facilitate close linkage to GISs and decision support systems (DSSs) [National Research Council, 139-63, 1999].

Prior efforts have implemented a range of different levels of coupling between a GIS and hydrologic models helping to elucidate the relative advantages and disadvantages of alternative coupling approaches in terms of representation of the watershed, watershed decomposition, sensitivity/uncertainty analysis, and parameter estimation as highlighted by Watkins et al. [1996]. Current GISs have limitations that impede coupling with hydrologic models [Abel et al., 1994; Kopp 1996]. Also since many of the advanced GISs are platform dependent, running mostly on Windows platform personal computers (PCs), they limit users from taking advantage of high performance computing architectures. Many commercial GIS framework suffer from closed data structures for GIS features, making it difficult to develop customized data manipulation/visualization tools that evolve with a modeler's/user's needs. Moreover, hydrologic models generally need other software support for pre- and post-processing tasks such as sensitivity analysis or decision support. The diverse needs of hydrologic research motivate the importance of developing coupled GIS and physical modeling systems able to incorporate more flexible tools and formats [Deckmyn et al., 1997].

In this paper, we demonstrate an integration methodology for an open source GIS framework and an integrated hydrologic model that enables users to take advantage of object oriented programming (OOP) to provide direct access to the GIS data structure, to better support efficient query and data transfer between the hydrologic model and GIS [Kumar et al. 2008]. The data structure has been designed to be flexible for modification and customization of the model or GIS, rich enough to represent complex user defined spatial relations and extensible to add more software tools as the need be. The "tightly-coupled" integrated GIS interface to Penn State Integrated Hydrologic Model (PIHM) has been created in the Open Source Quantum GIS [www.qgis.org]. The software framework used to create the tightly coupled PIHMgis system is generic and can be used in other model applications. Beyond describing the software framework for PIHMgis, this paper also demonstrates the importance and use of the framework for representing, modeling, visualizing and analysis data to Shaver's Creek Watershed in Susquehanna River Basin in Pennsylvania as case study.

2. INTEGRATION METHODOLOGY

2. 1 Introduction

Goodchild [1992], Nyerges [1993] and Sui et al. [1999] have discussed software integration strategies for GIS frameworks and models that range from loosely coupled to fully integrated systems. As discussed in Table 1, loose coupling, where a distinct GIS and model system exchange information using files, may be prone to data inconsistency, information loss and redundancy, leading to increased model setup time and post-processing. On the other extreme embedded coupling, where the model itself is developed in the GIS framework leads to a large and complex source code structure, which leaves the code inertial to change results it in being closed and isolated. Tight coupling preserves identity of GIS and hydrologic model behind the shared user interface and allows data exchange using shared data and method base.

Table 1. Different levels of integration between a GIS and a hydrologic model

Coupling Level → / Characteristics ↓	Loose	Tight Integration	Embedding
Shared User Interface	×	√	√
Shared data and method base	×	√	√
Intra-simulation Model Modification	×	×	√
Intra-simulation Query and Control	×	×	√

In this study we follow a tight coupling methodology [Kumar et al., 2008] based on the classification listed in Table 1. Tight coupling has the advantage of (1) preserving the advantage of independent development of various tool boxes as is the case for loose coupling and (2) the shared memory access to GIS data and model data linked through a carefully designed object oriented programming strategy for both the GIS and hydrologic model mimics many of the advantages of an embedded coupling system. As listed in Table 1, one of the major pre-requisites for tight coupling between GIS and a hydrologic model is to have a shared data and model structure base. Developing a shared data model for a GIS and a hydrologic model requires a careful consideration of both software systems and identification of connection points between them.

2.2 GIS Framework

The architecture of the GIS data model determines the ease of coupling a GIS with a hydrologic model. Generally a data model for a GIS includes constructs for spatial data, topological data and attribute data [Nyerges, 1987]. Data structures and associated descriptive constructs used in the data-management subsystem of GIS can lead to efficient data storage, editing and retrieval, and definition of new customized feature object representations within a GIS and integrated hydrologic model. This implies that a data model with its data, rules and relationships base can be a suitable basis for supporting GIS applications as well as hydrologic modeling. The data structure of the integrated hydrologic model will be then determined by the type and properties of the data models used in GIS. One of the pre-requisites for a tightly coupled integration, based on a shared data model will be access to the GIS architecture. Open source access to a GIS's architecture facilitates the development and use of GIS classes and methods while also providing the interface and linkages necessary for tight coupling. In this study, Quantum GIS (QGIS) is used as the base GIS system which is tightly coupled with PIHM. QGIS is open-source GIS and has been developed in C++, C and Qt (http://trolltech.com/), which makes it attractive as a base framework to develop a model interface.

2.3 Hydrologic Modeling Framework

The hydrologic processes incorporated in the model require data coverage sets of physical properties and system states at time t to predict system states the results at $t + \Delta t$. Δt is adaptively determined depending on the time scales of the interacting processes at each time t. In this study, we present PIHM [Qu and Duffy, 2007] tightly coupled with QGIS. PIHM is a physically-based, distributed hydrologic model that uses a finite volume formulation for the governing physical equations and constitutive relationships interacting on and across the unit elements of the decomposed domain. The governing physical equations generally represented by partial differential equations (PDEs) are discretized in space using the 'method of lines' [Leveque, 2002] approach to reduce them to ODEs. Figure 1 shows a typical "kernel" defined for triangular and linear prismatic elements along with the interacting physical processes to be coupled in the model. The kernel is designed to capture dynamics of multiple processes while maintaining the conservation of mass at all cells, as guaranteed by the finite volume formulation [Leveque, 2002].

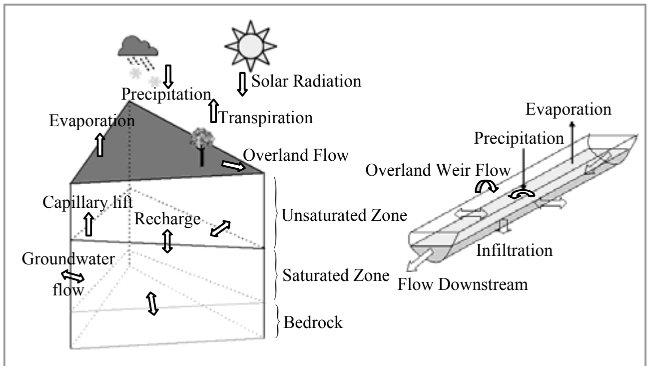

Figure 1. Interacting hydrologic processes on each prismatic element (left) and on each linear river element (right) [modified from Qu and Duffy, 2007]

The PIHMgis framework developed in this study supports the organization, development and assimilation of the extremely large set of spatial and temporal data for each model cell and its neighbors. With the shared data model, relationships and schemas between GIS and hydrologic model, tight coupling leads to an integrated system where GIS is simply another option to generate addition state and output variables in the model and to provide additional management, analysis and visualization options while the hydrologic model becomes one of the analytical functions of the GIS.

3. PIHMgis INTERFACE

PIHMgis is an integrated and extensible GIS system with data management, data analysis, data modelling, unstructured mesh generation and distributed PIHM modelling capabilities. The underlying philosophy of this integrated system is a shared geo-data model between GIS and PIHM thus making it possible to handle the complexity of the different data models, representation structures and model simulations. PIHMgis has been developed using basic QGIS source code. The Graphical User Interface (GUI) component have been designed in Qt [http://trolltech.com/products/qt], which is a standard framework for high-performance, cross-platform graphical widget toolkit development while the algorithms for several modules and the hydrologic model PIHM have been implemented in C and C++.

PIHMgis interface is procedural and interactive. Figure 2 shows a snapshot of PIHMgis interface. More snapshots are available at http://www.pihm.psu.edu/ >> PIHMgis >> Documents >> Users Guide. "Help" guides the user in selecting control parameters, the underlying algorithm through each PIHMgis module. Modularity is achieved via the plugin architecture which provides a mechanism for third parties to extend the QGIS core application.

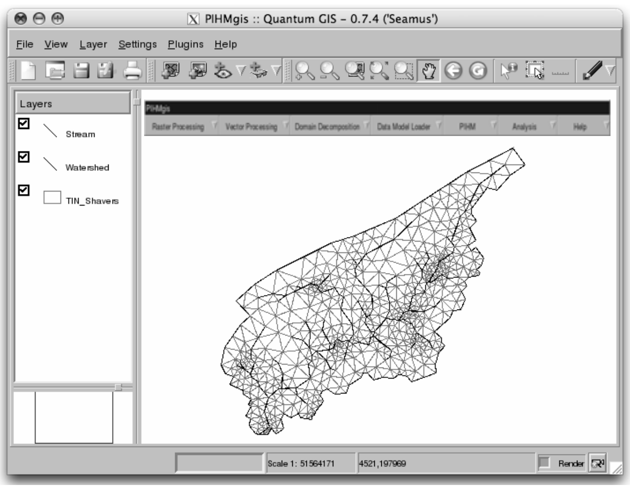

Figure 2. PIHMgis Interface. The data frame shows the unstructured domain decomposition for Shaver's Creek Watershed.

Architectural framework of PIHMgis shown in Figure 3 outlines the functionalities provided by the framework. Tight coupling shares the user interface between the GIS and the modelling framework. Direction of the arrow in shows the possible data flow within the framework. All the modules of PIHMgis have been organized in a procedural structure. Procedural framework of PIHMgis has been categorized into six processing stages. Raster Processing modules facilitate stream

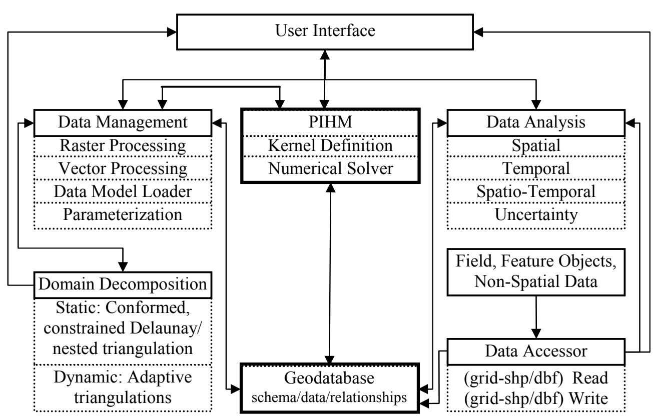

Figure 3. Architectural Framework of PIHMgis showing functionalities and conceptualization; direction of arrow shows the possible data flow within the

definition and watershed delineation. The Vector Processing aids users in defining watershed properties using nodes, polygons and polylines which eventually serve as domain constraints. The domain constraints are used to generate constrained Delaunay triangulations with certain mesh quality criteria. Before solving the finite volume based system of ODEs using RunPIHM module, the model parameters associated with soil, land cover as well as forcing and boundary conditions are assigned to each triangular and stream element in automated fashion in Data Model Loader modules. Finally, statistical and other kind of spatial and temporal data analysis and visualization can be performed to the model output using Analysis modules.

4. PIHMgis APPLICATION: CASE STUDY

PIHMgis takes advantage of the fact that modern geohydrologic datasets are stored and distributed in the form of a geodatabase [Arctur and Zeiler, 2004]. PIHMgis facilitates easy and accurate data development leading to easy model setup, model run, analysis and visualization. To demonstrate the procedural framework a case study application of PIHMgis to Shavers Creek Watershed located in Susquehanna River Basin is discussed in this section.

4.1 Raster Processing

Raster Processing facilitates stream definition and watershed delineation from the Digital Elevation Model (DEM) of the modelling domain. It is executed in a procedural framework involving computation of: (1) Fill Pits Grid; (2) Flow Grid; (3) Flow Accumulation Grid; (4) Stream Grid; (5) Link Grid; and (6) Catchment Grid. Figure 4A shows the 30 meter DEM of the Shavers Creek and Figure 4B shows the catchment polygon and stream polyline feature obtained after Raster Processing modules. A threshold of 2000 grids was applied to Flow Grid computed using d8 algorithm [Tarboton, 1991] for stream definition.

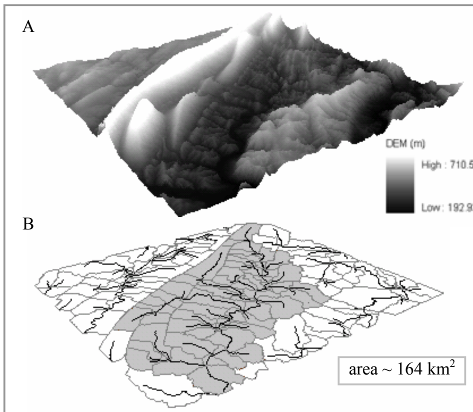

Figure 4. A shows DEM of Shaver's Creek. **B** Catchment and stream feature obtained after Raster Processing

4.2 Vector Processing

Geohydrologic features such as soils, land cover and other physiographic coverages can be used as constraining layers for the purpose of decomposition or mesh generation of modelling domain in addition to features generated using Raster Processing modulated by the modelling purpose. However, Stream polylines and catchment polygons obtained using Raster Processing retains the signature of the grid used, which of course depends on the DEM resolution. Vector Processing modules address issues specifically pertaining to modelling exercise as it allows development of a GIS layer which contains all the information of preferentially simplified constraining layers enabling efficient and quality

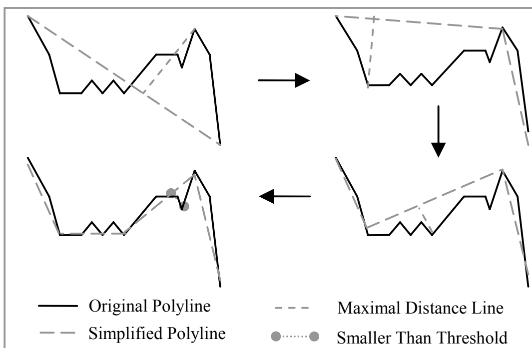

Figure 5. Intermediate steps involved in a polyline simplification

domain decomposition for modeling.

Features of the type polygon or polyline contain fluctuations or extraneous bends. Preferential simplification is a crucial module, part of Vector processing which simplifies the feature by eliminating nodes responsible for those fluctuations and still preserving the essential shape of the feature using simplification algorithm [Douglas and Peucker, 1973] as shown in Figure 5.

4.3 Domain Decomposition

PIHM uses vertical projection of triangular irregular mesh to form a local control volume which facilitates better representation of terrain [Kumar et al., 2008]. TRIANGLE [Shewchuk, 1996] has been integrated to decompose the domain into a high-quality, constrained, conforming Delaunay triangulation. TRIANGLE uses Ruppert [1995] and Chew [1993] algorithms for triangulation to generate non-skinny triangles and enforces the user selected quality constraint to the constraining layer prepared after Vector Processing modules.

In this study only the external boundary of the watershed is considered. A simplification tolerance of 200 meters was applied to both the watershed and stream feature. The decomposed unstructured mesh for the modelling domain is shown in Figure 2, where a 23 degree minimum-angle quality constraint was used.

4.4 Data Model Loader

The shared geo-data model contains all the topological and relational information needed to represent the modelling domain as well as the geohydrologic data needed for model parameterization. The Data Model Loader modules enrich the geodatabase defined by the classes and relationship of the shared geo data base. Several algorithms have been incorporated in Data Model Loader modules to facilitate topology assignment and model parameterization related to each triangular element and river segment.

An element is defined by the collection of three nodes in relation to the decomposed domain. Each element is assigned with a representative parameter value corresponding to each geohydrologic data layer along with nodes and neighbour information as shows in Figure 6A. Where as, a river segment is identified as one of the edge of an element, therefore defined by two nodes. Topology for channel segments is defined by From Node, To Node, Downstream segment, Left triangular element, Right triangular element [Figure 6B].

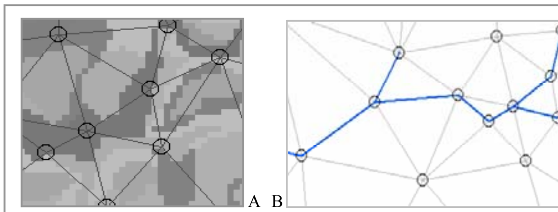

Figure 6. A. Data parameterization of each layer for each representative element. **B.** Topology for each stream segment using nodes and elements

4.5 Run PIHM

RunPIHM module embraces the PIHM and facilitates its execution right from the GIS framework. PIHM uses semi-discrete finite volume approach to reduce the governing PDEs into ODEs. The local system of ODEs defined on the each unit element and linear stream segments are assembled over the entire modelling domain forming a global system of ODEs. A state-of-art stiff-ODE solver SUNDIALS [http://www.llnl.gov/CASC/sundials/] is used to solve the global system of ODEs. RunPIHM module directly interacts with the geodatabase previously enriched by the Data Model Loader modules to retrieve all the topologic and geohydrologic model parameters. As simulation progress all the spatio-temporal model simulated data feeds back the geodatabase in the Network Common Data Form (NetCDF) format.

4.6 Analysis

PIHMgis modules discussed in section 4.1 to 4.5 provides easy data development, efficient model setup, and model execution. PIHMgis also provides modules to meet specific need for analysis and visualization of model simulated data in addition to basic GIS functionalities of QGIS. RunPIHM provides several optional parameters for the purpose of model calibration. However in this paper no calibration has been performed as part of model simulation. That is, data used in the simulation can be considered as a-priori information from independent sources. The Time Series module allows visualization of time series of model simulated parameter [described in Figure 1]. Figure 7 shows a time series plot of saturation averaged over the whole domain. Spatial Plot module allows creation of spatial maps as time series doesn't provide any information regarding spatial distribution of any simulated parameter. Figure 8 shows the spatial distribution for the annual average soil saturation. Since the motivation behind analysis of simulated results may vary widely depending on modelling interest, it is necessary that PIHMgis have extensible and pluggable architecture which allows easy addition of customized analysis and visualization modules.

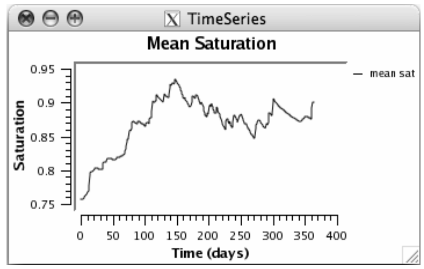

Figure 7. Spatial mean saturation over the entire domain for 1yr.

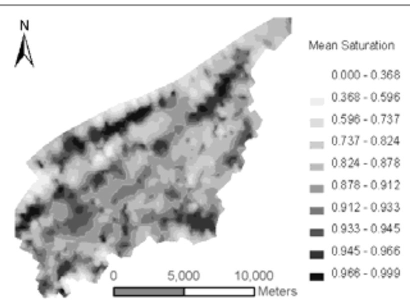

Figure 8. Mean annual saturation over the entire domain.

5. CONCLUSIONS

Isolated and independent hydrologic models and pre-processing (input data preparation) and post-processing (analysis and visualization), leads to increased model setup time and errors due to broken data flow. PIHMgis uses a tightly-coupled GIS framework which is based on shared-geo-data model to bridge hydrologic model and geohydrologic data (GIS framework). It offers a strategy for integration of modelling, analysis and visualization of complex multidimensional geohydrologic and land surface information.

Open source development of PIHMgis provides transparency, free access, modification to the source code. PIHMgis source code documentation is available at http://www.pihm.psu.edu/pihmgis_documents.html. The tight coupling strategy leaves the frameworks extensible and allows independent development. Moreover, the procedural framework of PIHMgis provides ease of use and preserves independence of each module at the same time.

REFERENCES

Abel, D.J., and Kilby, P.J., The systems integration problem, International Journal of Geographical Information Systems, 8(1):1-12, 1994.

Arctur, D., and Zeiler, M., Designing Geodatabases: Case Studies in GIS Data Modeling, ESRI Press, 2004.

Camara, G., Souza, R., Monteiro, A., Paiva, J., and Garrido, J., Handling complexity in GIS interface design, In Brazilian Symposium in Geoinformatics, Geoinfo, 1999.

Chew, P.L., Guaranteed-Quality Mesh Generation for Curved Surfaces, Proceedings of the Ninth Annual Symposium on Computational Geometry (San Diego, California), pages 274-280, Association for Computing Machinery, 1993.
David, D., Peucker, T., Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, The Canadian Cartographer 10(2), 112-122, 1973.
Deckmyn, J., Paniconi, C., Giacomelli, A., and Kleinfeldt, S., A Modular Approach to Integrated Environmental Modeling Systems Incorporating GIS and Visualization Functionality, CRS4-TECH-REP-96/26, 1997.
Entekhabi, D. and Eagleson, P.S., Land surface hydrology parameterization for atmospheric General Circulation models including subgrid scale spatial variability, Journal of Climate, 2(8), 816-831, 1989.
Freeze, R.A., and Harlan, R.L., Blueprint for a physically-based digitally simulated, hydrologic response model, J. Hydrol., 9, 237– 258, 1969.
Goodchild, M.F., Geographical information science. International Journal of Geographical Information Systems 6(1): 31-45, 1992.
Kopp, S.M., Linking GIS and hydrological models: where have we been, where are we going? In: HydroGIS 96: Application of Geographic Information Systems in Hydrology and Water Resources, 13-21. IAHS Publ. no. 235, 1996.
Kumar M., Bhatt, G. and Duffy, C.J., An Object Oriented Shared Data Model Coupling. International journal of Geographic Information System, In preparation, 2008.
Kumar M., Bhatt, G. and Duffy, C.J., An efficient domain decomposition framework for accurate representation of geodata in distributed hydrologic models, In preparation, 2008.
Leveque, R.J., Finite Volume methods for hyperbolic problems. Cambridge University Press, 2002.
Madsen, N.K., The method of lines for the numerical solution of partial differential equations (abstract), ACM SIGNUM Newsletter, v.10 n.4, p.5-7, 1975.
Maidment, D.R., GIS and hydrological modeling. In M. F. Goodchild, B. Parks, & L. Steyaert, (Eds.), Environmental modeling with GIS (pp. 147–167). New York: Oxford University Press, 1993.
Nyerges, T.L., GIS Research Issues Identified During a Cartographic Standards Process, Proceedings, of International Symposium on Geographic Information Systems, 1987.
Nyerges, T.L., Understanding the scope of GIS: Its relationship to environmental modeling. In Environmental Modeling with GIS, edited by Goodchild, M.F., Parks, B.O. and Steyaert, T. (New York: Oxford University Press), pp. 75-93, 1993.
Pitman, A.J., Henderson-Sellers, A., and Yang, Z.L., Sensitivity of regional climates to localized precipitation in global models, Nature, 346, 734-737, 1990.
Qu, Y., Duffy, C.J., A semidiscrete finite volume formulation for multiprocess watershed simulation, Water Resour. Res., 43, W08419, doi:10.1029/2006WR005752, 2007.
Ruppert, J., A Delaunay Refinement Algorithm for Quality 2-Dimensional Mesh Generation, Journal of Algorithms 18(3):548-585, 1995.
Shewchuk, J.R., Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator, in Applied Computational Geometry: Towards Geometric Engineering, volume 1148 of Lecture Notes in Computer Science, pages 203-222, Springer-Verlag (Berlin), 1996.
Sui, D.J. and Maggio, R.C., Integrating GIS with hydrological modelling: practices, problems and prospects, Computers, Environment and Urban Systems, 23, pages 33-51, 1999.
Tarboton, D.G., Bras, R.L., Rodriguez-Iturbe, I., On the Extraction of Channel Networks from Digital Elevation Data, Hydrologic Processes, 5(1): 81-100, 1991.
Watkins, D. W., McKinney, D. C., Maidment, D. R., and Lin, M., Use of geographic information systems in ground-water flow modeling, J. Water Resour. Plng. and Mgmt., ASCE, 122(2), 88–96, 1996
Wilson, J.P., Inskeep, W.P., Wraith, J.M., and Snyder, R.D., GIS-based solute transport modeling applications: Scale effects of soil and climate data input. Journal of Environmental Quality 25: 445-53, 1996.